

Nodos de origen, proceso y resultado  
de IBM SPSS Modeler 15



*Nota:* Antes de utilizar esta información y el producto que admite, lea la información general en Avisos el p. 526.

Esta edición se aplica a IBM SPSS Modeler 15 y a todas las versiones y modificaciones posteriores hasta que se indique lo contrario en nuevas ediciones.

Capturas de pantalla de productos de Adobe reimpresas con permiso de Adobe Systems Incorporated.

Capturas de pantalla de productos de Microsoft reimpresas con permiso de Microsoft Corporation.

Materiales bajo licencia: Propiedad de IBM

© **Copyright IBM Corporation 1994, 2012.**

Derechos restringidos para los usuarios del gobierno de Estados Unidos: Uso, duplicación o revelación restringidos por GSA ADP Schedule Contract con IBM Corp.

---

# Prefacio

IBM® SPSS® Modeler es el conjunto de programas de minería de datos de IBM Corp. orientado a las empresas. SPSS Modeler ayuda a las organizaciones a mejorar la relación con sus clientes y los ciudadanos a través de la comprensión profunda de los datos. Las organizaciones utilizan la comprensión que les ofrece SPSS Modeler para retener a los clientes más rentables, identificar las oportunidades de venta cruzada, atraer a nuevos clientes, detectar el fraude, reducir el riesgo y mejorar la prestación de servicios del gobierno.

La interfaz visual de SPSS Modeler invita a la pericia empresarial específica de los usuarios, lo que deriva en modelos predictivos más eficaces y la reducción del tiempo necesario para encontrar soluciones. SPSS Modeler ofrece muchas técnicas de modelado tales como pronósticos, clasificaciones, segmentación y algoritmos de detección de asociaciones. Una vez que se crean los modelos, IBM® SPSS® Modeler Solution Publisher permite su distribución en toda la empresa a los encargados de tomar las decisiones o a una base de datos.

## ***Acerca de IBM Business Analytics***

El software IBM Business Analytics ofrece información completa, coherente y precisa en la que los órganos de toma de decisiones confían para mejorar el rendimiento comercial. Un conjunto integral de [inteligencia empresarial](#), [análisis predictivo](#), [rendimiento comercial y gestión de estrategias](#), así como de [aplicaciones de análisis](#) le ofrece una información clara, inmediata e interactiva del rendimiento actual y la capacidad para predecir resultados futuros. En combinación con extensas soluciones sectoriales, prácticas probadas y servicios profesionales, las organizaciones de cualquier tamaño pueden conseguir el máximo de productividad y alcanzar mejores resultados.

Como parte de esta familia, el software de análisis predictivo de IBM SPSS ayuda a las organizaciones a predecir eventos futuros y actuar proactivamente según esa información para lograr mejores resultados comerciales. Los clientes comerciales, gubernamentales y académicos de todo el mundo confían en la tecnología de IBM SPSS como ventaja ante la competencia para atraer, retener y hacer crecer los clientes, reduciendo al mismo tiempo el fraude y reduciendo el riesgo. Al incorporar el software de IBM SPSS en sus operaciones diarias, las organizaciones se convierten en empresas predictivas, capaces de dirigir y automatizar decisiones para alcanzar los objetivos comerciales y lograr una ventaja considerable sobre la competencia. Para obtener más información o contactar con un representante, visite <http://www.ibm.com/spss>.

## ***Asistencia técnica***

La asistencia técnica está disponible para el mantenimiento de los clientes. Los clientes podrán ponerse en contacto con el servicio de asistencia técnica si desean recibir ayuda sobre la utilización de los productos de IBM Corp. o sobre la instalación en los entornos de hardware admitidos. Para ponerse en contacto con el servicio de asistencia, visite el IBM Corp. sitio Web en <http://www.ibm.com/support>. Prepárese para identificarse, identificar a su organización y su acuerdo de asistencia al solicitar asistencia.

---

# Contenido

## **1 Acerca de IBM SPSS Modeler 1**

Productos IBM SPSS Modeler . . . . .	1
IBM SPSS Modeler . . . . .	1
IBM SPSS Modeler Server . . . . .	2
IBM SPSS Modeler Administration Console . . . . .	2
IBM SPSS Modeler Batch . . . . .	2
IBM SPSS Modeler Solution Publisher . . . . .	3
IBM SPSS Modeler Server Adaptadores para IBM SPSS Collaboration and Deployment Services . . . . .	3
Ediciones de IBM SPSS Modeler . . . . .	3
Documentación de IBM SPSS Modeler . . . . .	4
Documentación de SPSS Modeler Professional . . . . .	4
Documentación de SPSS Modeler Premium . . . . .	5
Ejemplos de aplicaciones . . . . .	6
Carpeta Demos . . . . .	7

## **2 Nodos de fuente 8**

Resumen . . . . .	8
Nodo Enterprise View . . . . .	9
Opciones de configuración del nodo Enterprise View . . . . .	10
Conexiones de Enterprise View . . . . .	12
Selección de la DPD . . . . .	13
Selección de la tabla . . . . .	14
Nodo de fuente Base de datos . . . . .	15
Configuración de las opciones de nodos Base de datos . . . . .	16
Adición de conexión a la base de datos . . . . .	18
Especificación de valores predefinidos para una conexión de base de datos . . . . .	19
Selección de una tabla de la base de datos . . . . .	22
Realización de consultas a la base de datos . . . . .	23
Nodo Archivo variable . . . . .	25
Opciones de configuración para el nodo Archivo variable . . . . .	27
Nodo Archivo fijo . . . . .	29
Configuración de opciones para el nodo Archivo fijo . . . . .	29
Configuración de almacenamiento de campos y formato . . . . .	32
Nodo Data Collection . . . . .	35
Opciones de archivo de importación de Data Collection . . . . .	36
Propiedades de metadatos de importación de IBM SPSS Data Collection . . . . .	40
Cadena de conexión a la base de datos . . . . .	41

Propiedades avanzadas . . . . .	42
Importación de conjuntos de respuestas múltiples . . . . .	42
Notas de importación de columna de IBM SPSS Data Collection . . . . .	43
Nodo de fuente de IBM Cognos BI . . . . .	44
Iconos de objeto de cognos . . . . .	45
Importación de datos de Cognos . . . . .	45
Importación de informes de Cognos . . . . .	47
Conexiones de Cognos . . . . .	49
Selección de ubicación de Cognos . . . . .	50
Especificación de parámetros de datos o informes . . . . .	50
Nodo de fuente SAS . . . . .	51
Configuración de opciones para el nodo de fuente SAS . . . . .	52
Nodo de fuente de Excel . . . . .	53
Nodo Fuente XML . . . . .	54
Selección entre múltiples elementos raíz . . . . .	56
Eliminación de espacios no deseados de los datos de fuente XML . . . . .	57
Nodo Datos Usuario . . . . .	58
Configuración de opciones para el nodo Datos Usuario . . . . .	60
Pestañas de nodos de fuente comunes . . . . .	65
Definición de niveles de medición en el nodo Fuente . . . . .	65
Filtrado de campos desde el nodo de fuente . . . . .	67

### **3   Nodos de operaciones con registros**

**69**

Conceptos básicos de las operaciones con registros . . . . .	69
Nodo Seleccionar . . . . .	70
Nodo Muestrear . . . . .	72
Opciones del nodo Muestrear . . . . .	73
Configuración de Conglomerado y estratificación . . . . .	77
Tamaños de muestra por estratos . . . . .	78
Nodo Equilibrar . . . . .	80
Opciones de configuración del nodo Equilibrar . . . . .	81
Nodo Agregar . . . . .	82
Opciones de configuración del nodo Agregar . . . . .	82
Nodo Adición de RFM . . . . .	85
Opciones de configuración del nodo Adición de RFM . . . . .	86
Nodo Ordenar . . . . .	87
Configuración de optimización de ordenación . . . . .	88
Nodo Fundir . . . . .	89
Tipos de uniones . . . . .	90

Especificación de un método de fusión y claves . . . . .	92
Selección de datos para uniones parciales . . . . .	94
Especificación de condiciones para una fusión . . . . .	94
Filtrados de campos del nodo Fundir . . . . .	95
Configuración del orden de entrada y etiquetado . . . . .	97
Configuración de optimización de fusión . . . . .	99
Nodo Añadir . . . . .	100
Opciones de configuración de Añadir . . . . .	101
Nodo Distinguir . . . . .	102
Configuración de optimización de Distinguir . . . . .	104

## **4 Nodos de operaciones con campos**

**106**

Conceptos básicos de las operaciones con campos . . . . .	106
Preparación automática de datos . . . . .	108
Pestaña Campos . . . . .	111
Pestaña Configuración . . . . .	111
Configuración de campos . . . . .	112
Preparar fechas y horas . . . . .	113
Excluir campos . . . . .	114
Preparación de entradas y destinos . . . . .	115
Creación y selección de características . . . . .	117
Nombres de campos . . . . .	119
Pestaña análisis . . . . .	120
Resumen de procesamiento de campo . . . . .	122
Campos . . . . .	123
Resumen de acciones . . . . .	125
Poder predictivo . . . . .	126
Tabla de campos . . . . .	127
Detalles de campo . . . . .	128
Detalles de acción . . . . .	130
Generación de un nodo Derivar . . . . .	133
Nodo Tipo . . . . .	135
Niveles de medida . . . . .	137
Conversión de datos continuos . . . . .	140
¿Qué es la creación de una instancia? . . . . .	141
Valores de campos . . . . .	142
Definición de valores perdidos . . . . .	147
Comprobación de valores de tipo . . . . .	148
Definición del papel de campos . . . . .	149
Copia de atributos de tipo . . . . .	150

Pestaña Configuración de Formato de campo . . . . .	151
Cambio de nombre o filtrado de campos . . . . .	154
Opciones de configuración de filtrado . . . . .	155
Nodo Conjunto . . . . .	161
Configuración de nodo Conjunto . . . . .	162
Nodo Derivar . . . . .	165
Opciones básicas de configuración del nodo Derivar . . . . .	166
Derivación de múltiples campos . . . . .	167
Opciones de configuración de derivación como fórmula . . . . .	169
Opciones de configuración de derivación como marca . . . . .	170
Opciones de configuración de derivación como conjunto . . . . .	173
Opciones de configuración de derivación como estado . . . . .	174
Opciones de configuración de derivación como recuento . . . . .	175
Opciones de configuración de derivación como condicional . . . . .	176
Nueva codificación de valores con el nodo Derivar . . . . .	177
Nodo Rellenar. . . . .	177
Conversión del almacenamiento mediante el nodo Rellenar. . . . .	179
Nodo Anonimizar . . . . .	180
Opciones de configuración del nodo Anonimizar . . . . .	181
Anonimización de valores de campo . . . . .	183
Nodo Reclasificar. . . . .	185
Opciones de configuración del nodo Reclasificar. . . . .	185
Reclasificación de múltiples campos . . . . .	188
Almacenamiento y nivel de medición de los campos reclasificados . . . . .	189
Nodo Intervalos . . . . .	189
Opciones de configuración del nodo Intervalos . . . . .	190
Intervalos de ancho fijo . . . . .	192
Cuantiles (igual frecuencia o igual suma). . . . .	192
Asignar rangos a los casos . . . . .	195
Media/desviación típica. . . . .	196
Intervalos óptimos. . . . .	197
Vista preliminar de intervalos generados . . . . .	199
Nodo Análisis de RFM. . . . .	201
Configuración del nodo Análisis de RFM . . . . .	202
Intervalos del nodo Análisis de RFM . . . . .	204
Nodo Partición . . . . .	205
Opciones del nodo Partición . . . . .	206
Nodo Marcas . . . . .	207
Opciones de configuración del nodo Marcas . . . . .	208
Nodo Reestructurar . . . . .	209
Opciones de configuración del nodo Reestructurar . . . . .	210

Nodo Transponer . . . . .	211
Opciones de configuración del nodo Transponer . . . . .	212
Nodo de intervalos de tiempo . . . . .	216
Especificación de intervalos de tiempo . . . . .	217
Opciones de generación en intervalos de tiempo . . . . .	219
Período de estimación . . . . .	221
Predicciones . . . . .	222
Intervalos admitidos . . . . .	225
Nodo Histórico . . . . .	237
Opciones de configuración del nodo Histórico . . . . .	238
Nodo Reorg. campos . . . . .	239
Opciones de configuración de Reorg. campos . . . . .	239

## **5 Nodos de gráficos**

**243**

Características comunes de nodos de gráficos. . . . .	243
Aspecto estético, superposiciones, paneles y animación. . . . .	244
Uso de la pestaña Resultados . . . . .	249
Uso de la pestaña Anotaciones . . . . .	250
Gráficos 3D . . . . .	250
Nodo Tablero . . . . .	251
Tablero Ficha Básico . . . . .	252
Tablero Pestaña Detallado . . . . .	257
Tipos de visualización de Graphboard preincorporado disponibles. . . . .	259
Creación de visualizaciones de mapas. . . . .	266
Ejemplos de Tableros . . . . .	267
Pestaña Aspecto de Tablero . . . . .	285
Configuración de la ubicación de las plantillas, hojas de estilo y mapas. . . . .	287
Administración de plantillas, hojas de estilo y archivos de mapas . . . . .	289
Conversión y distribución de mapas en formato shapefile . . . . .	290
Conceptos claves de los mapas. . . . .	291
Uso de la utilidad de conversión de mapas . . . . .	292
Distribución de archivos de mapas . . . . .	299
Nodo Gráfico . . . . .	299
Pestaña de nodo Gráfico . . . . .	302
Pestaña Opciones de Gráfico. . . . .	305
Pestaña Aspecto de Gráfico . . . . .	306
Utilización de un gráfico . . . . .	307
Nodo Distribución. . . . .	308
Pestaña Gráfico de Distribución. . . . .	309

Pestaña Aspecto de Distribución . . . . .	310
Utilización de un nodo Distribución . . . . .	311
Nodo Histograma . . . . .	313
Pestaña Gráfico de Histograma . . . . .	314
Pestaña Opciones de Histograma . . . . .	315
Pestaña Aspecto de Histograma . . . . .	316
Uso de histogramas . . . . .	316
Nodo Colección . . . . .	317
Pestaña Gráfico de Colección . . . . .	318
Pestaña Opciones de Colección . . . . .	319
Pestaña Aspecto de Colección . . . . .	320
Utilización de un gráfico de colección . . . . .	321
Nodo G. múltiple . . . . .	322
Pestaña Gráfico de G. múltiple . . . . .	323
Pestaña Aspecto de G. múltiple . . . . .	325
Utilización de un gráfico múltiple . . . . .	326
Nodo Malla . . . . .	326
Pestaña Gráfico de Malla . . . . .	328
Pestaña Opciones de Malla . . . . .	329
Pestaña Aspecto de Malla . . . . .	332
Utilización de un gráfico de malla . . . . .	333
Nodo Gráfico de tiempo . . . . .	338
Pestaña Gráfico de Tiempo . . . . .	339
Pestaña Aspecto de Gráfico de Tiempo . . . . .	340
Utilización de un gráfico de tiempo . . . . .	341
Nodo Evaluación . . . . .	341
Pestaña Gráfico de Evaluación . . . . .	347
Pestaña Opciones de Evaluación . . . . .	348
Pestaña Aspecto de Evaluación . . . . .	350
Lectura de los resultados de una evaluación del modelo . . . . .	352
Uso de un diagrama de evaluación . . . . .	353
Exploración de gráficos . . . . .	355
Uso de tramos . . . . .	356
Usar regiones . . . . .	360
Uso de elementos con marca . . . . .	363
Generación de nodos a partir de gráficos . . . . .	364
Edición de visualizaciones . . . . .	367
Reglas generales de edición de visualizaciones . . . . .	368
Edición y formato de texto . . . . .	369
Cambio de colores, tramas, estilos de guión y transparencia . . . . .	370
Rotación y cambio de la forma y la relación de aspecto de los elementos de punto . . . . .	371
Cambio de tamaño de los elementos gráficos . . . . .	372
Especificación de márgenes y relleno . . . . .	373

Formato de numeración . . . . .	373
Cambio de la configuración del eje y la escala . . . . .	374
Edición de categorías . . . . .	376
Cambio de la orientación de paneles . . . . .	378
Transformación del sistema de coordenadas . . . . .	379
Cambio de estadísticos y elementos gráficos . . . . .	380
Cambio de la posición de la leyenda . . . . .	383
Copia de una visualización y de datos de visualización . . . . .	383
Métodos abreviados de teclado . . . . .	384
Adición de títulos y notas al pie . . . . .	384
Uso de hojas de estilo de gráficos . . . . .	386
Aplicación de hojas de estilo . . . . .	387
Impresión, almacenamiento, copia y exportación de gráficos . . . . .	389

## **6    *Nodos de resultado* **391****

Conceptos básicos sobre nodos de resultados . . . . .	391
Administración de resultados . . . . .	392
Visualización de resultados . . . . .	393
Publicar en Web . . . . .	393
Visualización de resultados en un explorador HTML . . . . .	396
Exportación de resultados . . . . .	396
Selección de casillas y columnas . . . . .	397
Nodo Tabla . . . . .	398
Pestaña Configuración del nodo Tabla . . . . .	398
Pestaña de formato del nodo Tabla . . . . .	399
Pestaña Resultado del nodo de resultados . . . . .	399
Explorador de tablas . . . . .	401
Nodo Matriz . . . . .	403
Pestaña Configuración del nodo Matriz . . . . .	403
Pestaña Aspecto del nodo Matriz . . . . .	404
Explorador de resultados de nodo Matriz . . . . .	406
Nodo Análisis . . . . .	407
Pestaña Análisis del nodo Análisis . . . . .	408
Explorador de resultados de análisis . . . . .	410
Nodo Auditar datos . . . . .	412
Pestaña Configuración del nodo Auditar datos . . . . .	414
Pestaña Calidad de Auditar datos . . . . .	416
Explorador de resultados de auditoría de datos . . . . .	417
Nodo Transformación . . . . .	427
Pestaña Opciones del nodo Transformar . . . . .	428

Pestaña Resultado del nodo Transformar . . . . .	429
Visor de resultados del nodo Transformar . . . . .	429
Nodo Estadísticos . . . . .	432
Pestaña Configuración del nodo Estadísticos . . . . .	433
Explorador de resultados de estadísticos . . . . .	434
Nodo Medias . . . . .	437
Comparación de medias para grupos independientes . . . . .	437
Comparación de medias entre campos relacionados . . . . .	438
Opciones del nodo Medias . . . . .	439
Explorador de resultados del nodo Medias . . . . .	440
Nodo Informe . . . . .	442
Pestaña Plantilla del nodo Informe . . . . .	443
Explorador de resultados del nodo Informe . . . . .	445
Nodo Val. globales . . . . .	446
Pestaña Configuración del nodo Val. globales . . . . .	447
Complementos de IBM SPSS Statistics . . . . .	448

## **7 Nodos de exportación**

**450**

Conceptos básicos sobre nodos de exportación . . . . .	450
Nodo Exportar base de datos . . . . .	451
Pestaña Exportar del nodo Base de datos . . . . .	451
Opciones de fusión Exportar base de datos . . . . .	453
Opciones de Esquema de exportación de base de datos . . . . .	455
Opciones de índice de exportación de base de datos . . . . .	459
Opciones avanzadas de exportación de base de datos . . . . .	461
Programación del cargador masivo . . . . .	464
nodo de exportación Archivo plano . . . . .	472
Pestaña Exportar del nodo Archivo plano . . . . .	472
Nodos de exportación IBM SPSS Data Collection . . . . .	473
Nodo Exportar de IBM Cognos BI . . . . .	475
Conexión de Cognos . . . . .	475
conexión ODBC . . . . .	477
Nodo Exportar SAS . . . . .	479
Pestaña Exportar del nodo Exportar SAS . . . . .	479
Nodo de exportación Excel . . . . .	480
Pestaña Exportar del nodo Excel . . . . .	480
Nodo de exportación XML . . . . .	481
Escritura de datos XML . . . . .	483
Opciones de registros de correspondencia XML . . . . .	483

Opciones de campos de correspondencia XML . . . . .	484
Presentación preliminar de correspondencia XML . . . . .	485

## **8 Nodos de IBM SPSS Statistics 486**

Conceptos básicos de nodos de IBM SPSS Statistics . . . . .	486
Nodo de archivos estadísticos . . . . .	487
Nodo Transformación de Statistics . . . . .	489
Pestaña Sintaxis del nodo Transformación de Statistics . . . . .	489
Sintaxis permitida . . . . .	491
Nodo Modelo de Statistics . . . . .	493
Pestaña Modelo del nodo Modelos de Statistics . . . . .	494
Nodo Modelo de Statistics - Resumen del nugget de modelo . . . . .	495
Nodo Resultados de Statistics . . . . .	497
Pestaña Sintaxis del nodo Resultados de Statistics . . . . .	498
Pestaña Resultados del nodo Resultados de Statistics . . . . .	500
Nodo Exportación de Statistics . . . . .	501
Pestaña Exportar del nodo Exportar Statistics . . . . .	502
Cambio de nombre o filtrado de campos para IBM SPSS Statistics . . . . .	503

## **9 Supernodos 505**

Conceptos básicos de supernodos . . . . .	505
Tipos de supernodos . . . . .	505
Supernodos de origen . . . . .	506
Supernodos de proceso . . . . .	506
Supernodos terminales . . . . .	507
Creación de supernodos . . . . .	508
Anidación de supernodos . . . . .	510
Ejemplos de supernodos válidos . . . . .	511
Ejemplos de supernodos no válidos . . . . .	512
Bloqueo de Supernodos . . . . .	513
Bloqueo y desbloqueo de un supernodo . . . . .	514
Edición de un supernodo bloqueado . . . . .	515
Edición de Supernodos . . . . .	516
Modificación de los tipos de Supernodos . . . . .	516
Anotación y cambio de nombre de Supernodos . . . . .	517
Parámetros de Supernodos . . . . .	518

Supernodos y almacenamiento en caché . . . . .	522
Supernodos y procesos . . . . .	523
Almacenamiento y carga de supernodos . . . . .	524

## ***Apéndice***

<b><i>A Avisos</i></b>	<b>526</b>
------------------------	------------

<b><i>Índice</i></b>	<b>529</b>
----------------------	------------



# ***Acerca de IBM SPSS Modeler***

IBM® SPSS® Modeler es un conjunto de herramientas de minería de datos que permite desarrollar rápidamente modelos predictivos mediante técnicas empresariales y utilizarlos en operaciones empresariales para mejorar la toma de decisiones. Con un diseño que sigue el modelo CRISP-DM, estándar del sector, SPSS Modeler admite el proceso completo de minería de datos, desde los propios datos hasta obtener los mejores resultados empresariales.

SPSS Modeler ofrece una gran variedad de métodos de modelado procedentes del aprendizaje automático, la inteligencia artificial y el estadístico. Los métodos disponibles en la paleta de modelado permiten derivar nueva información procedente de los datos y desarrollar modelos predictivos. Cada método tiene ciertos puntos fuertes y es más adecuado para determinados tipos de problemas.

SPSS Modeler puede adquirirse como producto independiente o utilizarse como cliente junto con SPSS Modeler Server. También hay disponible cierto número de opciones adicionales que se resumen en las siguientes secciones. Si desea obtener más información, consulte <http://www.ibm.com/software/analytics/spss/products/modeler/>.

## ***Productos IBM SPSS Modeler***

La familia de productos IBM® SPSS® Modeler y su software asociado se componen de lo siguiente:

- IBM SPSS Modeler
- IBM SPSS Modeler Server
- IBM SPSS Modeler Administration Console
- IBM SPSS Modeler Batch
- IBM SPSS Modeler Solution Publisher
- IBM SPSS Modeler Server adaptadores para IBM SPSS Collaboration and Deployment Services

## ***IBM SPSS Modeler***

SPSS Modeler es una versión con todas las funcionalidades del producto que puede instalar y ejecutar en su ordenador personal. Puede ejecutar SPSS Modeler en modo local como un producto independiente o utilizarla en modo distribuido junto con IBM® SPSS® Modeler Server para mejorar el rendimiento a la hora de trabajar con grandes conjuntos de datos.

Con SPSS Modeler, puede crear modelos predictivos precisos de forma rápida e intuitiva sin necesidad de programación. Mediante su exclusiva interfaz visual, podrá visualizar fácilmente el proceso de minería de datos. Con ayuda del análisis avanzado incrustado en el producto podrá

detectar patrones y tendencias en sus datos que anteriormente estaban ocultos. Podrá modelar los resultados y comprender los factores que influyen en ellos, lo que le permitirá aprovechar oportunidades comerciales y mitigar los riesgos.

SPSS Modeler está disponible en dos ediciones: SPSS Modeler Professional y SPSS Modeler Premium. Si desea obtener más información, consulte el tema [Ediciones de IBM SPSS Modeler](#) el p. 3.

### ***IBM SPSS Modeler Server***

SPSS Modeler utiliza una arquitectura de cliente/servidor para distribuir peticiones de cliente para operaciones que requieren un uso intensivo de los recursos a un software de servidor de gran potencia, lo que proporciona un rendimiento más rápido con conjuntos de datos de mayor volumen.

SPSS Modeler Server es un producto con licencia independiente que se ejecuta de manera continua en modo de análisis distribuido en un host de servidor junto con una o más instalaciones de IBM® SPSS® Modeler. De este modo, SPSS Modeler Server ofrece un mejor rendimiento cuando se trabaja con grandes conjuntos de datos, ya que las operaciones que requieren un uso intensivo de memoria se pueden realizar en el servidor sin tener que descargar datos al equipo cliente. IBM® SPSS® Modeler Server también ofrece asistencia para las capacidades de optimización de SQL y modelado interno de la base de datos, lo que proporciona mayores ventajas en cuanto al rendimiento y la automatización.

### ***IBM SPSS Modeler Administration Console***

Modeler Administration Console es una aplicación gráfica para administrar muchas de las opciones de configuración de SPSS Modeler Server, las cuales también pueden configurarse a través de un archivo de opciones. La aplicación proporciona una interfaz de usuario de la consola para supervisar y configurar las instalaciones de SPSS Modeler Server y está disponible de forma completamente gratuita para los clientes actuales de SPSS Modeler Server. La aplicación solamente se puede instalar en los ordenadores con Windows; sin embargo, puede administrar un servidor que esté instalado en cualquier plataforma compatible.

### ***IBM SPSS Modeler Batch***

Aunque la minería de datos suele ser un proceso interactivo, también es posible ejecutar SPSS Modeler desde una línea de comandos, sin necesidad de la interfaz gráfica del usuario. Por ejemplo, puede que tenga tareas repetitivas o cuya ejecución sea de larga duración que quiera realizar sin intervención por parte del usuario. SPSS Modeler Batch es una versión especial del producto que ofrece asistencia para todas las capacidades analíticas de SPSS Modeler sin acceder a la interfaz de usuario habitual. Es necesario disponer de una licencia de SPSS Modeler Server para utilizar SPSS Modeler Batch.

## **IBM SPSS Modeler Solution Publisher**

SPSS Modeler Solution Publisher es una herramienta que le permite crear una versión empaquetada de una ruta de SPSS Modeler que se puede ejecutar en un motor de tiempo de ejecución externo o incrustado en una aplicación externa. De este modo, podrá publicar y distribuir rutas completas de SPSS Modeler para utilizarlas en entornos que no tengan SPSS Modeler instalado. SPSS Modeler Solution Publisher se distribuye como parte del servicio IBM SPSS Collaboration and Deployment Services - Scoring, para el que se necesita una licencia independiente. Con esta licencia, recibirá SPSS Modeler Solution Publisher Runtime, que le permite ejecutar las rutas publicadas.

## **IBM SPSS Modeler Server Adaptadores para IBM SPSS Collaboration and Deployment Services**

Tiene a su disposición un determinado número de adaptadores para IBM® SPSS® Collaboration and Deployment Services que permiten que SPSS Modeler y SPSS Modeler Server interactúen con un repositorio de IBM SPSS Collaboration and Deployment Services. De este modo, varios usuarios podrán compartir una ruta de SPSS Modeler distribuida en el repositorio, o bien se podrá acceder a ella desde la aplicación cliente de baja intensidad IBM SPSS Modeler Advantage. Debe instalar el adaptador en el sistema donde se aloje el repositorio.

## **Ediciones de IBM SPSS Modeler**

SPSS Modeler está disponible en las siguientes ediciones.

### **SPSS Modeler Professional**

SPSS Modeler Professional proporciona todas las herramientas que necesita para trabajar con la mayoría de los tipos de datos estructurados, como los comportamientos e interacciones registrados en los sistemas de CRM, datos demográficos, comportamientos de compra y datos de ventas.

### **SPSS Modeler Premium**

SPSS Modeler Premium es un producto con licencia independiente que amplía SPSS Modeler Professional para poder trabajar con datos especializados, como los utilizados para el análisis de entidades o las redes sociales, así como con datos de texto no estructurados. SPSS Modeler Premium está formado por los siguientes componentes:

**IBM® SPSS® Modeler Entity Analytics** incorpora una dimensión completamente nueva al análisis predictivo de IBM® SPSS® Modeler. Mientras que el análisis predictivo trata de predecir comportamientos futuros a partir de datos del pasado, el análisis de entidades se centra en mejorar la coherencia de los datos actuales mediante la resolución de conflictos de identidades dentro de los propios registros. La identidad de un individuo, una organización, un objeto o cualquier otra entidad puede estar expuesta a ambigüedades. La resolución de identidades puede ser vital en diversos campos, entre los que se incluyen la gestión de la relación con el cliente, la detección de fraudes, la lucha contra el blanqueo de dinero y la seguridad nacional e internacional.

**IBM SPSS Modeler Social Network Analysis** transforma la información sobre relaciones en campos que caracterizan el comportamiento social de individuos y grupos. Mediante el uso de datos que describen las relaciones subyacentes de las redes sociales, IBM® SPSS® Modeler Social Network Analysis identifica a los líderes sociales que influyen en el comportamiento de otros en la red. Además, puede determinar qué personas se ven más afectadas por otros participantes de la red. Al combinar estos resultados con otras medidas, puede crear perfiles completos de individuos en los que basar sus modelos predictivos. Los modelos que incluyan esta información social tendrán un mejor rendimiento que los modelos que no la incluyan.

**Text Analytics for IBM® SPSS® Modeler** utiliza tecnologías de lingüística avanzada y Procesamiento del lenguaje natural (PLN) para procesar con rapidez una gran variedad de datos de texto sin estructurar, extraer y organizar los conceptos clave y agruparlos en categorías. Las categorías y conceptos extraídos se pueden combinar con los datos estructurados existentes, como pueden ser datos demográficos, y se pueden aplicar para modelar utilizando el conjunto completo de herramientas de minería de datos de SPSS Modeler para tomar decisiones mejores y más certeras.

## ***Documentación de IBM SPSS Modeler***

Tiene a su disposición documentación en formato de ayuda en línea desde el menú Ayuda de SPSS Modeler. Se incluye documentación para SPSS Modeler, SPSS Modeler Server y SPSS Modeler Solution Publisher, así como el Manual de aplicaciones y otros materiales de apoyo.

La documentación completa de cada producto (incluidas las instrucciones de instalación) en formato PDF está disponible en la carpeta *Documentation* en cada DVD del producto. También es posible descargar los documentos de instalación en Internet en <http://www-01.ibm.com/support/docview.wss?uid=swg27023172>.

La documentación en ambos formatos también está disponible desde el centro de información de SPSS Modeler en <http://publib.boulder.ibm.com/infocenter/spssmodl/v15r0m0/>.

## ***Documentación de SPSS Modeler Professional***

El conjunto de documentación de SPSS Modeler Professional (excluidas las instrucciones de instalación) es el siguiente.

- **Manual del usuario de IBM SPSS Modeler.** Introducción general sobre cómo usar SPSS Modeler, incluyendo cómo crear rutas de datos, tratar valores perdidos, crear expresiones CLEM, trabajar con proyectos e informes y empaquetar rutas para su distribución en IBM SPSS Collaboration and Deployment Services, Predictive Applications o IBM SPSS Modeler Advantage.
- **Nodos Fuente, Proceso y Resultado de IBM SPSS Modeler.** Descripciones de todos los nodos utilizados para leer, procesar y dar salida a datos en diferentes formatos. En la práctica, esto implica todos los nodos que no sean nodos de modelado.
- **Nodos de modelado de IBM SPSS Modeler.** Descripciones de todos los nodos utilizados para crear modelos de minería de datos. IBM® SPSS® Modeler ofrece una variedad de métodos de modelado tomados del aprendizaje de las máquinas, la inteligencia artificial y la estadística.

- **Manual de algoritmos de IBM SPSS Modeler.** Descripciones de los fundamentos matemáticos de los métodos de modelado que se utilizan en SPSS Modeler. Esta guía está disponible únicamente en formato PDF.
- **Manual de aplicaciones de IBM SPSS Modeler.** Los ejemplos de esta guía ofrecen introducciones breves y concisas a métodos y técnicas de modelado específicos. También tiene a su disposición una versión en línea de este manual en el menú Ayuda. Si desea obtener más información, consulte el tema [Ejemplos de aplicaciones](#) el p. 6.
- **Procesos y automatización de IBM SPSS Modeler.** Información sobre la automatización del sistema a través de procesos, incluidas las propiedades que se pueden utilizar para manipular nodos y rutas.
- **IBM SPSS Modeler Manual de distribución.** Información sobre la ejecución de rutas y escenarios de SPSS Modeler como pasos en trabajos de procesamiento en IBM® SPSS® Collaboration and Deployment Services Deployment Manager.
- **Guía del desarrollador de IBM SPSS Modeler CLEF.** CLEF permite integrar programas de otros fabricantes, como rutinas de procesamiento de datos o algoritmos de modelado como nodos en SPSS Modeler.
- **Manual de minería interna de bases de datos de IBM SPSS Modeler.** Este manual incluye información sobre cómo utilizar la potencia de su base de datos, tanto para mejorar su rendimiento como para ampliar su oferta de capacidades analíticas a través de algoritmos de terceros.
- **Guía de administración y rendimiento de IBM SPSS Modeler Server.** Información sobre la configuración y administración de IBM® SPSS® Modeler Server.
- **Manual del usuario de IBM SPSS Modeler Administration Console.** Información sobre cómo instalar y utilizar la interfaz de usuario de la consola para supervisar y configurar SPSS Modeler Server. La consola se implementa como complemento de la aplicación Deployment Manager.
- **Manual de IBM SPSS Modeler Solution Publisher.** SPSS Modeler Solution Publisher es un componente complementario que permite a las organizaciones publicar rutas para su uso fuera del entorno estándar de SPSS Modeler.
- **Manual de CRISP-DM de IBM SPSS Modeler.** Manual que explica paso a paso cómo utilizar la metodología de CRISP-DM en la minería de datos con SPSS Modeler.
- **Manual del usuario de IBM SPSS Modeler Batch.** Guía completa de cómo utilizar IBM SPSS Modeler en modo por lotes, incluida información detallada sobre la ejecución del modo por lotes y argumentos de línea de comandos. Esta guía está disponible únicamente en formato PDF.

## ***Documentación de SPSS Modeler Premium***

El conjunto de documentación de SPSS Modeler Premium (excluidas las instrucciones de instalación) es el siguiente.

- **IBM SPSS Modeler Entity Analytics Manual del usuario.** Información sobre cómo utilizar el análisis de entidades con SPSS Modeler, que cubre la instalación y configuración de repositorios, nodos de análisis de entidades y tareas administrativas.

- **IBM SPSS Modeler Social Network Analysis Manual del usuario.** Una guía para realizar análisis de redes sociales con SPSS Modeler, incluido el análisis de grupos y el análisis de difusión.
- **Text Analytics for SPSS Modeler Manual del usuario.** Información sobre cómo utilizar el análisis de texto con SPSS Modeler, que cubre los nodos de minería de texto, programa interactivo, plantillas y otros recursos.
- Manual del usuario de **Text Analytics for IBM SPSS Modeler Administration Console.** Información sobre cómo instalar y utilizar la interfaz de usuario de la consola para supervisar y configurar IBM® SPSS® Modeler Server para su uso con Text Analytics for SPSS Modeler. La consola se implementa como complemento de la aplicación Deployment Manager.

## ***Ejemplos de aplicaciones***

Mientras que las herramientas de minería de datos de SPSS Modeler pueden ayudar a resolver una amplia variedad de problemas organizativos y empresariales, los ejemplos de la aplicación ofrecen introducciones breves y adaptadas de técnicas y métodos de modelado específicos. Los conjuntos de datos utilizados aquí son mucho más pequeños que los enormes almacenes de datos gestionados por algunos analizadores de datos, pero los conceptos y métodos implicados deberían ser escalables a las aplicaciones reales.

Para acceder a los ejemplos pulsando Ejemplos de aplicación en el menú Ayuda de SPSS Modeler. Los archivos de datos y rutas de muestra se instalan en la carpeta *Demos* en el directorio de instalación del producto. Si desea obtener más información, consulte el tema [Carpeta Demos](#) el p. 7.

**Ejemplos de modelado de base de datos.** Consulte los ejemplos que figuran en el Manual de minería interna de bases de datos de *IBM SPSS Modeler*.

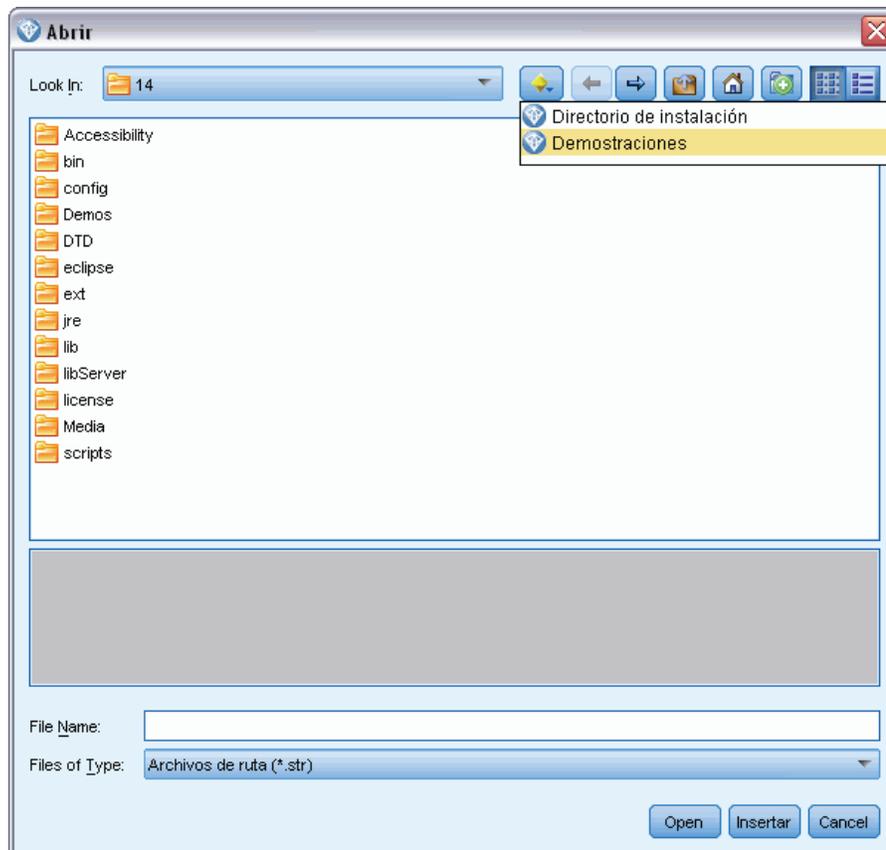
**Ejemplos de procesos.** Consulte los ejemplos que figuran en la Guía de procesos y automatización de *IBM SPSS Modeler*.

## Carpeta Demos

Los archivos de datos y rutas de muestra utilizados con los ejemplos de la aplicación se instalan en la carpeta *Demos* en el directorio de instalación del producto. También puede acceder a esta carpeta desde el grupo de programas IBM SPSS Modeler 15 del menú Inicio de Windows o pulsando *Demos* de la lista de directorios recientes en el cuadro de diálogo Abrir archivo.

Figura 1-1

*Selección de la carpeta Demos desde la lista de directorios utilizados recientemente*



# Nodos de fuente

## Resumen

Los nodos de fuente permiten importar los datos almacenados en distintos formatos, entre los que se incluyen archivos planos, IBM® SPSS® Statistics (.sav), SAS, Microsoft Excel y bases de datos relacionales compatibles con ODBC. También puede generar datos sintéticos mediante el nodo Datos Usuario.

La paleta Fuentes contiene los siguientes nodos:



El nodo Enterprise View crea una conexión a IBM® SPSS® Collaboration and Deployment Services Repository, lo que permite leer los datos de Enterprise View en una ruta y empaquetar un modelo en un escenario al que otros usuarios pueden acceder desde el repositorio. Si desea obtener más información, consulte el tema [Nodo Enterprise View](#) el p. 9.



El nodo Base de datos se puede utilizar para importar datos desde otros paquetes mediante ODBC (del inglés, Open Database Connectivity), incluidos Microsoft SQL Server, DB2, Oracle, etc. Si desea obtener más información, consulte el tema [Nodo de fuente Base de datos](#) el p. 15.



El nodo Archivo variable lee datos desde los archivos de texto de campo libre: campos cuyos registros contienen un número constante de campos, pero un número variado de caracteres. Este nodo resulta también útil para los archivos con texto de cabecera de longitud fija y determinados tipos de anotaciones. Si desea obtener más información, consulte el tema [Nodo Archivo variable](#) el p. 25.



El nodo Archivo fijo importa datos desde archivos de texto de campo fijo; esto es, archivos cuyos campos no están delimitados pero empiezan en la misma posición y tienen una longitud fija. Los datos heredados o generados por la máquina se suelen almacenar en formato de campo fijo. Si desea obtener más información, consulte el tema [Nodo Archivo fijo](#) el p. 29.



El nodo Archivo Statistics lee los datos desde un formato de archivo .sav que utiliza SPSS Statistics y archivos caché guardados en IBM® SPSS® Modeler, que también puede utilizar el mismo formato. Si desea obtener más información, consulte el tema [Nodo de archivos estadísticos](#) en el capítulo 8 el p. 487.



El nodo IBM® SPSS® Data Collection importa datos de encuesta a partir de varios formatos utilizados por el software de investigación de mercado conforme al modelo de datos de Data Collection. Se debe instalar la biblioteca de desarrollador de Data Collection para utilizar este nodo. Si desea obtener más información, consulte el tema [Nodo Data Collection](#) el p. 35.



El nodo de origen de IBM Cognos BI importa datos desde las bases de datos de Cognos BI. Si desea obtener más información, consulte el tema [Importación de datos de Cognos](#) el p. 45.



El nodo Archivo SAS importa datos SAS a SPSS Modeler. Si desea obtener más información, consulte el tema [Nodo de fuente SAS](#) el p. 51.



El nodo Excel importa datos de cualquier versión de Microsoft Excel. No es necesario un origen de datos ODBC. Si desea obtener más información, consulte el tema [Nodo de fuente de Excel](#) el p. 53.



El nodo de origen XML importa datos en formato XML en la ruta. Puede importar un único archivo o todos los archivos en un directorio. Puede especificar opcionalmente un archivo de esquema para leer la estructura XML. Si desea obtener más información, consulte el tema [Nodo Fuente XML](#) el p. 54.



El nodo Datos de usuario proporciona una manera fácil de crear datos sintéticos, ya sea partiendo de cero o modificando los datos existentes. Esto resulta útil, por ejemplo, cuando desee crear un conjunto de datos de comprobación para el modelado. Si desea obtener más información, consulte el tema [Nodo Datos Usuario](#) el p. 58.

Para iniciar una ruta, añada un nodo de fuente al lienzo de rutas. A continuación, pulse dos veces en el nodo para abrir su cuadro de diálogo. Las distintas pestañas del cuadro de diálogo permiten leer datos, ver los campos y valores así como establecer distintas opciones, incluidos filtros, tipos de datos, papel del campo y comprobación de los valores perdidos.

## Nodo Enterprise View

El nodo Enterprise View permite crear y mantener una conexión entre una sesión de IBM® SPSS® Modeler y una Enterprise View en un IBM® SPSS® Collaboration and Deployment Services Repository compartido. Al hacerlo, es posible leer datos de Enterprise View en una ruta de SPSS Modeler y empaquetar un modelo de SPSS Modeler en un escenario al que pueden acceder otros usuarios del repositorio compartido.

Un **escenario** es un archivo que contiene una ruta de SPSS Modeler con modelos y nodos específicos, así como propiedades adicionales que permiten su distribución a IBM SPSS Collaboration and Deployment Services Repository con fines de puntuación o actualización automática del modelo. El uso de los nodos Enterprise View con escenario garantiza que, en una situación con varios usuarios, todos los usuarios trabajen con los mismos datos. Una **conexión** es un enlace de una sesión de SPSS Modeler con Enterprise View en IBM SPSS Collaboration and Deployment Services Repository.

La **Enterprise View** es un conjunto completo de datos que pertenecen a una organización, independientemente de dónde se encuentren físicamente los datos. Cada conexión está formada por una selección específica de una única **vista de aplicación** (subconjunto de Enterprise View adaptado a una aplicación concreta), una **definición de proveedor de datos** (DPD, enlaza las tablas y columnas lógicas de la vista de aplicación con una fuente de datos física) y un **entorno** (identifica las columnas concretas que deben asociarse a los segmentos empresariales definidos). Las definiciones de Enterprise View, vistas de aplicación y DPD se almacenan y versionan en el repositorio, aunque los datos reales se encuentran en una o más bases de datos u otras fuentes externas.

Una vez que se ha establecido una conexión, se especifica la **tabla** de la vista de aplicación con la que se desea trabajar en SPSS Modeler. En una vista de aplicación, una tabla es una vista lógica que consta de algunas o todas las columnas de una o más tablas físicas de una o más bases de datos físicas. Por tanto, el nodo Enterprise View permite ver los registros procedentes de varias tablas de bases de datos como una única tabla en SPSS Modeler.

### **Requisitos**

- Para utilizar el nodo Enterprise View, debe instalar y configurar IBM SPSS Collaboration and Deployment Services Repository en el sitio, con una Enterprise View, unas vistas de aplicación y DPD ya definidos.

*Nota:* Es necesario disponer de una licencia independiente para acceder a un repositorio IBM® SPSS® Collaboration and Deployment Services. Si desea obtener más información, consulte <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

- Además, es necesario instalar IBM® SPSS® Collaboration and Deployment Services Enterprise View Driver en cada ordenador que se utilice para modificar o ejecutar la ruta. Para Windows, basta con instalar el controlador en el ordenador donde se ha instalado IBM® SPSS® Modeler o IBM® SPSS® Modeler Server y no será necesario realizar ninguna configuración adicional del controlador. En UNIX, es necesario añadir una referencia al proceso *pev.sh* al proceso de inicio. Póngase en contacto con el administrador local si desea obtener más información acerca de la instalación del controlador IBM SPSS Collaboration and Deployment Services Enterprise View Driver.
- Una DPD se define respecto a una determinada fuente de datos ODBC. Para usar una DPD de SPSS Modeler, debe tener una fuente de datos ODBC definida en el host del servidor de SPSS Modeler que tenga el mismo nombre (y que conecte con el mismo almacén de datos) que a la que se hace referencia en la DPD.

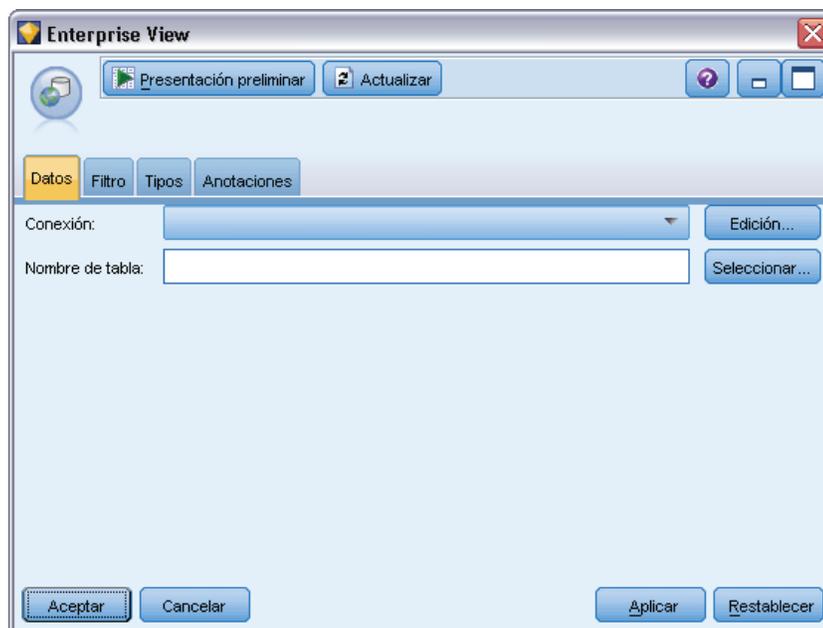
### **Opciones de configuración del nodo Enterprise View**

Puede utilizar las opciones de la pestaña Datos del cuadro de diálogo Enterprise View para:

- Seleccionar una conexión de repositorio existente.
- Editar una conexión de repositorio existente.
- Crear una nueva conexión de repositorio.
- Seleccionar una tabla de vista de aplicación.

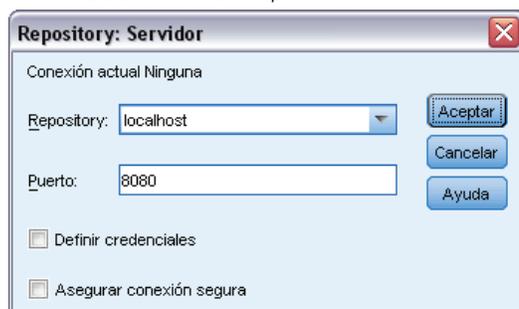
Consulte el Manual del administrador de IBM® SPSS® Collaboration and Deployment Services para obtener más información sobre cómo trabajar con repositorios.

Figura 2-1  
Adición de una conexión a IBM SPSS Collaboration and Deployment Services Repository



**Conexión.** La lista desplegable proporciona opciones para seleccionar una conexión de repositorio existente, editar una conexión existente o añadir una conexión. Si ya ha iniciado sesión en un repositorio utilizando IBM® SPSS® Modeler, al elegir la opción Añadir/editar una conexión aparece el cuadro de diálogo Conexiones de Enterprise View, en el que puede definir o editar los datos necesarios para la conexión actual. Si no ha iniciado sesión, al seleccionar esta opción aparecerá el cuadro de diálogo de inicio de sesión del repositorio.

Figura 2-2  
Inicio de sesión en un repositorio



Para obtener información sobre el inicio de sesión en el repositorio, consulte el *Manual de usuario de SPSS Modeler*.

Una vez que se ha establecido una conexión a un repositorio, dicha conexión permanecerá hasta que salga de SPSS Modeler. Otros nodos de la misma ruta pueden compartir la misma conexión, pero será necesario crear una nueva conexión para cada nueva ruta.

Si se inicia sesión correctamente, aparecerá el cuadro de diálogo Conexiones de Enterprise View.

**Nombre de tabla.** Este campo está vacío inicialmente y no puede rellenarse hasta que se cree una conexión. Si conoce el nombre de la tabla de vista de aplicación a la que desea acceder, introdúzcalo en el campo Nombre de tabla. En caso contrario, pulse en el botón Seleccionar para abrir un cuadro de diálogo que muestre las tablas de vista de aplicación disponibles.

## Conexiones de Enterprise View

Este cuadro de diálogo permite definir o editar los detalles necesarios para la conexión de repositorio. Puede especificar las siguientes opciones:

- Vista de aplicación y versión
- Entorno
- Definición de proveedor de datos (DPD)
- Descripción de la conexión

Figura 2-3  
Selección de una vista de aplicación



**Conexiones.** Muestra las conexiones de repositorio existentes.

- **Añadir una nueva conexión.** Muestra el cuadro de diálogo Recuperar objeto, desde el que puede buscar y seleccionar una vista de aplicación del repositorio.
- **Copiar la conexión seleccionada.** Realiza una copia de la conexión seleccionada, lo que evita tener que volver a buscar la misma vista de aplicación.
- **Eliminar la conexión seleccionada.** Elimina la conexión seleccionada de la lista.

**Detalles de la conexión.** Muestra la vista de aplicación, la etiqueta de versión, el entorno, la DPD y texto descriptivo de la conexión actualmente seleccionada en el panel de conexiones.

- **Vista de aplicación.** La lista desplegable muestra la vista de aplicación seleccionada, si la hay. Si se han realizado conexiones con otras vistas de aplicación en la sesión actual, también aparecerán en la lista desplegable. Pulse en el botón Examinar que aparece al lado para buscar otras vistas de aplicación en el repositorio.
- **Etiqueta de la versión.** Los campos desplegables muestran todas las etiquetas de versión definidas para la vista de aplicación especificada. Las etiquetas de versión ayudan a identificar las versiones de los objetos del repositorio. Por ejemplo, puede haber dos versiones de una vista de aplicación determinada. Con las etiquetas, puede especificar la etiqueta PRUEBA para la versión utilizada en el entorno de desarrollo y la etiqueta PRODUCCIÓN para la versión utilizada en el entorno de producción. Seleccione la etiqueta adecuada.

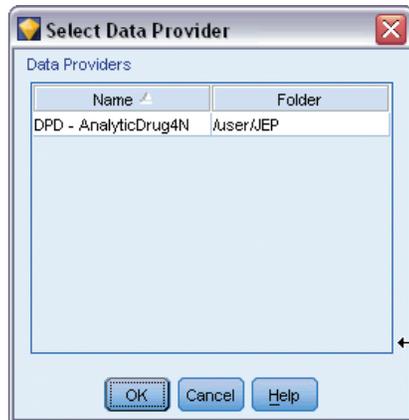
*Nota:* las etiquetas no deberían incluir el carácter “[”]; de lo contrario, el nombre de la tabla no se mostrará en el pestaña Datos del cuadro de diálogo Enterprise View.

- **Entorno.** El campo desplegable muestra todos los entornos válidos. La configuración del entorno determina las DPD que estarán disponibles, especificando por tanto las columnas concretas que deberán asociarse a los segmentos empresariales definidos. Por ejemplo, cuando se selecciona Analytic, solamente se devuelven las columnas de la vista de aplicación definidas como Analytic. El entorno por defecto es Analytic; también puede seleccionar Operational.
- **Proveedor de datos.** La lista desplegable muestra los nombres de hasta diez definiciones de proveedor de datos para la vista de aplicación seleccionada. Solamente se muestran las DPD que hacen referencia a la vista de aplicación seleccionada. Pulse en el botón Examinar que aparece al lado para ver el nombre y la ruta de todas las DPD relacionadas con la vista de aplicación actual.
- **Descripción.** Texto descriptivo sobre la conexión de repositorio. Este texto se utilizará para el nombre de la conexión; al pulsar en Aceptar el texto aparecerá en la lista desplegable Conexión y en la barra de título del cuadro de diálogo de Enterprise View, así como en la etiqueta del nodo Enterprise View, en el lienzo.

## ***Selección de la DPD***

El cuadro de diálogo Seleccionar proveedor de datos muestra el nombre y la ruta de todas las DPD que hacen referencia a la vista de aplicación actual.

Figura 2-4  
Selección de una DPD



Las vistas de aplicación pueden tener varias DPD para servir de ayuda en las diferentes fases de un proyecto. Por ejemplo, los datos históricos utilizados para generar un modelo pueden proceder de una base de datos, mientras que los datos operativos proceden de otra.

Una DPD se define respecto a una determinada fuente de datos ODBC. Para usar una DPD de IBM® SPSS® Modeler, debe tener una fuente de datos ODBC definida en el host del servidor de SPSS Modeler que tenga el mismo nombre (y que conecte con el mismo almacén de datos) que a la que se hace referencia en la DPD.

- Para elegir la DPD con la que desea trabajar, elija su nombre en la lista y pulse en Aceptar.

## Selección de la tabla

El cuadro de diálogo Seleccionar tabla muestra todas las tablas a las que se hace referencia en la vista de aplicación actual. Este cuadro de diálogo aparece vacío si no se ha establecido ninguna conexión a IBM SPSS Collaboration and Deployment Services Repository.

Figura 2-5  
Selección de una tabla



- Para elegir la tabla con la que desea trabajar, elija su nombre en la lista y pulse en Aceptar.

## Nodo de fuente Base de datos

El nodo de fuente Base de datos se puede utilizar para importar datos desde otros paquetes mediante ODBC (del inglés, Open Database Connectivity), incluidos Microsoft SQL Server, DB2, Oracle, etc.

Para leer o escribir en una base de datos, debe tener una fuente de datos ODBC instalada y configurada para la base de datos pertinente, con los permisos de lectura o escritura, según sea necesario. IBM® SPSS® Data Access Pack incluye un conjunto de controladores ODBC que puede utilizarse con este fin. Estos controladores están disponibles en IBM SPSS Data Access Pack DVD o en el sitio de descarga. Si tiene alguna pregunta acerca de la creación o configuración de permisos de las fuentes de datos ODBC, póngase en contacto con el administrador de la base de datos.

El soporte para bases de datos en IBM® SPSS® Modeler se clasifica en tres niveles, cada uno representando un nivel diferente para retrotracción y optimización de SQL, dependiendo del proveedor de la base de datos. Los diferentes niveles de soporte se implementan mediante un número de configuración del sistema, que puede personalizarse como parte de un contrato de Servicios.

Los tres niveles de soporte de bases de datos son:

Tabla 2-1

*Niveles de soporte para bases de datos*

Nivel de soporte	Descripción
Nivel 1	Toda la retrotracción SQL posible está disponible, con optimización específica para bases de datos SQL.
Nivel 2	La mayoría de la retrotracción SQL está disponible, con optimización no específica para bases de datos SQL.
Nivel 3	No hay retrotracción u optimización para SQL, solamente la lectura y la escritura de datos en la base de datos.

### Controladores ODBC compatibles

Si desea obtener la información más reciente acerca de las bases de datos y los controladores ODBC compatibles y cuyo uso se ha comprobado con SPSS Modeler 15, vea las matrices de compatibilidad de productos disponibles en el sitio de asistencia corporativo (<http://www.ibm.com/support>).

### Dónde instalar los controladores

Tenga en cuenta que todos los controladores ODBC deben instalarse y configurarse en todos los ordenadores en los que pueda realizarse el procesamiento.

- Si desea ejecutar IBM® SPSS® Modeler en modo local (independiente), deberá instalar los controladores en el ordenador local.
- Si ejecuta SPSS Modeler en modo distribuido en un servidor IBM® SPSS® Modeler Server remoto, los controladores ODBC se deberán instalar en el ordenador en el que se encuentre SPSS Modeler Server. Para SPSS Modeler Server en sistemas UNIX, consulte también “Configuración de controladores ODBC en sistemas UNIX” más adelante en esta sección.

- Si necesita poder acceder a las mismas fuentes de datos desde SPSS Modeler y SPSS Modeler Server, deberá instalar los controladores ODBC en ambos ordenadores.
- Si desea ejecutar SPSS Modeler a través de Terminal Services, deberá instalar los controladores de ODBC en el servidor de Terminal Services en el que se haya instalado SPSS Modeler.
- Si desea utilizar el tiempo de ejecución de IBM® SPSS® Modeler Solution Publisher para ejecutar las rutas publicadas en otro ordenador, también deberá instalar y configurar los controladores ODBC en dicho ordenador.

**Nota:** si utiliza SPSS Modeler Server en UNIX para acceder a una base de datos Teradata, debe utilizar ODBC Driver Manager instalado con el controlador Teradata ODBC. Para implementar este cambio en SPSS Modeler Server indique el valor de ODBC\_DRIVER\_MANAGER\_PATH junto al principio del proceso modelersrv.sh, donde indican los comentarios. Esta variable de entorno se debe configurar con la ubicación de ODBC Driver Manager que se incluye junto con el controlador Teradata ODBC (/usr/odbc/lib en una instalación normal del controlador Teradata ODBC). Deberá reiniciar SPSS Modeler Server para que los cambios surtan efecto. Para obtener más información acerca de las plataformas SPSS Modeler Server compatibles en el acceso de Teradata y la versión del controlador ODBC de Teradata compatible, consulte el sitio de asistencia corporativo <http://www.ibm.com/support>.

### **Configuración de controladores ODBC en sistemas UNIX**

De manera predeterminada, Driver Manager de DataDirect no está configurado para SPSS Modeler Server en sistemas UNIX. Para configurar UNIX para que cargue Driver Manager de DataDirect, introduzca los siguientes comandos:

```
cd directorio_instalación_modeler_server/bin
rm -f libspssodbc.so
ln -s libspssodbc_datadirect.so libspssodbc.so
```

Al realizar esta acción, se elimina el enlace predeterminado y se crea un enlace a Driver Manager de DataDirect.

Siga los siguientes pasos generales para acceder a los datos desde una base de datos:

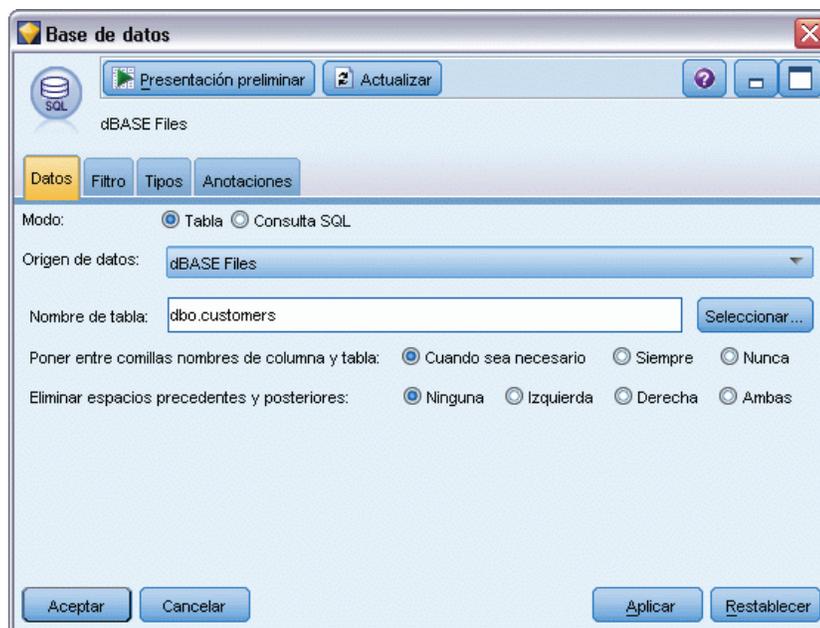
- ▶ Instale un controlador ODBC y configure una fuente de datos en la base de datos que desee utilizar.
- ▶ En el cuadro de diálogo del nodo Base de datos, conéctese a una base de datos mediante el modo Tabla o Consulta SQL.
- ▶ Seleccione una tabla de la base de datos.
- ▶ Mediante el cuadro de diálogo del nodo Base de datos, puede modificar los tipos de usos y filtrar los campos de datos.

Estos pasos se describen de forma más detallada en los siguientes temas.

### **Configuración de las opciones de nodos Base de datos**

Puede utilizar las opciones de la pestaña Datos del cuadro de diálogo del nodo de fuente Base de datos para poder acceder a una base de datos y leer los datos desde la tabla seleccionada.

Figura 2-6  
Carga de datos seleccionando una tabla



**Modo.** Seleccione Tabla para conectarse a una tabla mediante los controles de los cuadros de diálogo.

Seleccione Consulta SQL para realizar consultas a la base de datos seleccionada a continuación mediante SQL. Si desea obtener más información, consulte el tema [Realización de consultas a la base de datos](#) el p. 23.

**Fuente de datos.** Para los modos Tabla y Consulta SQL, puede introducir un nombre en el campo Fuente de datos o seleccionar Añadir nueva conexión de base de datos de la lista desplegable.

Las siguientes opciones se utilizan para conectarse a una base de datos y seleccionar una tabla mediante el cuadro de diálogo:

**Nombre de tabla.** Si conoce el nombre de la tabla a la que desea acceder, introdúzcalo en el campo Nombre de tabla. En caso contrario, pulse en el botón Seleccionar para abrir un cuadro de diálogo que muestre las tablas disponibles.

**Poner entre comillas nombres de columna y tabla.** Indique si desea que los nombres de columna y tabla aparezcan entre comillas cuando las consultas se envíen a la base de datos (en el caso, por ejemplo, de que contengan espacios o signos de puntuación).

- La opción Cuando sea necesario pondrá entre comillas los nombres de campos y tablas *sólo* si incluyen caracteres no estándar. Los caracteres no estándar incluyen caracteres no ASCII, caracteres de espacio y los no alfanuméricos distintos del punto (.).
- Seleccione Nunca si *no* desea que los nombres de campos y tablas aparezcan entre comillas.
- Seleccione Siempre si desea que *todos* los nombres de campos y tablas aparezcan entre comillas.

**Eliminar espacios precedentes y posteriores.** Seleccione las opciones para descartar los espacios precedentes y posteriores en las cadenas.

*Nota:* Las comparaciones entre cadenas que utilizan y no utilizan retrotracción SQL pueden generar diferentes resultados donde existen espacios precedentes.

**Lectura de cadenas vacías desde Oracle.** Al leer de o escribir en una base de datos de Oracle, debe tener en cuenta que, a diferencia de IBM® SPSS® Modeler y la mayoría del resto de base de datos, Oracle trata y almacena valores de cadenas vacíos tanto como valores equivalentes hasta nulos. Esto significa que los mismos datos extraídos desde una base de datos de Oracle pueden comportarse de manera diferente a cuando se extraen desde otra base de datos o de un archivo, y pueden devolver resultados diferentes.

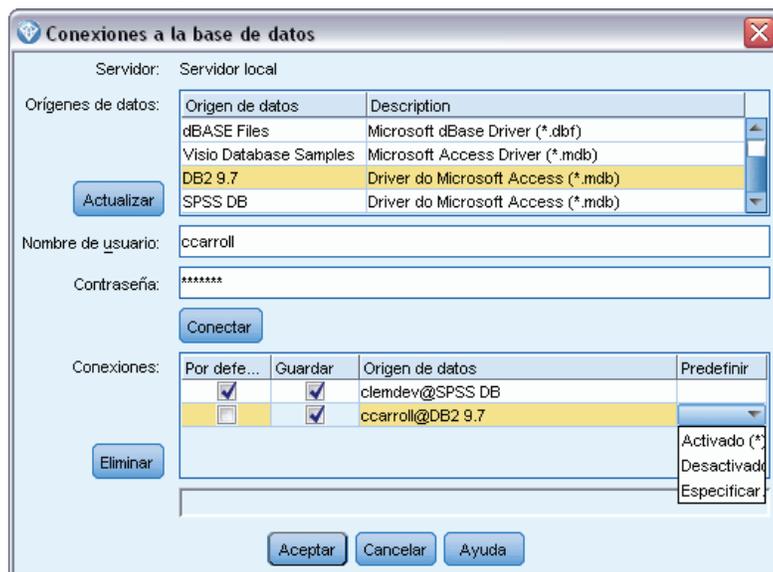
## Adición de conexión a la base de datos

Para poder abrir una base de datos, primero debe seleccionar la fuente de datos a la que desea conectarse. En la pestaña Datos, seleccione Añadir nueva conexión de base de datos de la lista desplegable Fuente de datos.

Esto abrirá el cuadro de diálogo Conexiones a la base de datos.

*Nota:* para conocer otro modo de abrir este cuadro de diálogo, en el menú principal, seleccione: Herramientas > Bases de datos...

Figura 2-7  
Cuadro de diálogo Conexiones a la base de datos



**Fuentes de datos.** Muestra las fuentes de datos disponibles. Desplácese hacia abajo si no encuentra la base de datos deseada. Una vez haya seleccionado una fuente de datos y haya introducido las contraseñas, pulse en Conectar. Pulse en Actualizar para actualizar la lista.

**Nombre de usuario.** Si la fuente de datos está protegida por una contraseña, introduzca su nombre de usuario.

**Contraseña.** Si la fuente de datos está protegida por una contraseña, introduzca su contraseña.

**Conexiones.** Muestra bases de datos conectadas actualmente.

- **Por defecto.** Puede seleccionar una conexión como el valor por defecto. Al hacerlo, la fuente de la base de datos o los nodos de exportación tendrán esta conexión predefinida como su fuente de datos, aunque puede editarse si lo desea.
- **Guardar.** Puede seleccionar una o más conexiones que quiera volver a mostrar en sesiones posteriores.
- **Fuente de datos.** Las cadenas de conexión para las bases de datos conectadas actualmente.
- **Predefinido.** Indica (con el carácter \*) si se han especificado valores predefinidos para la conexión de base de datos. Para especificar valores predefinidos, pulse en esta columna en la fila correspondiente a la conexión de base de datos y seleccione Especificar en la lista. Si desea obtener más información, consulte el tema [Especificación de valores predefinidos para una conexión de base de datos](#) el p. 19.

Para eliminar las conexiones, seleccione una de la lista y pulse en Eliminar.

Una vez haya completado sus selecciones, pulse en Aceptar.

## ***Especificación de valores predefinidos para una conexión de base de datos***

Para algunas bases de datos, puede especificar cierto número de ajustes por defecto para una conexión de base de datos. Todos los ajustes se aplican a la exportación de base de datos.

Los tipos de bases de datos que admiten esta función son los siguientes:

- IBM InfoSphere Warehouse ejecutándose en DB2 9.1 o posterior. Si desea obtener más información, consulte el tema [Configuración para IBM DB2 InfoSphere Warehouse](#) el p. 19.
- SQL Server 2008 o posterior, Enterprise Edition y Developer Edition. Si desea obtener más información, consulte el tema [Configuración para SQL Server](#) el p. 20.
- Oracle 10g y 11gR1 o posterior, Enterprise Edition o Personal Edition. Si desea obtener más información, consulte el tema [Configuración para Oracle](#) el p. 20.
- IBM Netezza, IBM DB2 para z/OS y Teradata se conectan a una base de datos o esquema de un modo similar. Si desea obtener más información, consulte el tema [Configuración de IBM Netezza, IBM DB2 para z/OS y Teradata](#) el p. 22.

Si está conectado a una base de datos o esquema que no admite esta función, verá el mensaje No se puede configurar ningún valor predefinido para esta conexión de base de datos.

### ***Configuración para IBM DB2 InfoSphere Warehouse***

Estos ajustes se muestran para IBM InfoSphere Warehouse ejecutándose en DB2 9.1 o posterior.

**Espacio de tabla.** El espacio de tabla que debe utilizarse para la exportación. Los administradores de la base de datos pueden crear o configurar espacios de tabla con particiones. Recomendamos que seleccione uno de estos espacios de tabla (en lugar del valor por defecto) para utilizarlo para la exportación de base de datos.

**Utilizar compresión.** Si está seleccionado, crea tablas para la exportación con compresión (por ejemplo, el equivalente a `CREATE TABLE MYTABLE(...) COMPRESS YES;` en SQL).

**No registrar actualizaciones.** Si está seleccionado, evita el registro al crear tablas e insertar datos (el equivalente a `CREATE TABLE MYTABLE(...) NOT LOGGED INITIALLY;` en SQL).

### ***Configuración para SQL Server***

Estos ajustes se muestran para SQL Server 2008 o posterior, Enterprise Edition y Developer Edition.

**Utilizar compresión.** Si está seleccionado, crea tablas para la exportación con compresión.

**Compresión para.** Elija el nivel de compresión.

- **Fila.** Activa la compresión a nivel de fila (por ejemplo, el equivalente a `CREATE TABLE MYTABLE(...) WITH (DATA_COMPRESSION = ROW);` en SQL).
- **Página.** Activa la compresión a nivel de página (por ejemplo, `CREATE TABLE MYTABLE(...) WITH (DATA_COMPRESSION = PAGE);` en SQL).

### ***Configuración para Oracle***

#### ***Configuración de Oracle 10g***

Estos ajustes se muestran para Oracle 10g Enterprise Edition o Personal Edition.

**Utilizar compresión.** Si está seleccionado, crea tablas para la exportación con compresión. Para esta versión de la base de datos, únicamente está disponible la compresión básica (por ejemplo, `CREATE TABLE MYTABLE(...) COMPRESS;` en SQL).

#### ***Configuración de Oracle 11gR1***

Estos ajustes se muestran para Oracle 11g R1 Enterprise Edition o Personal Edition.

**Utilizar compresión.** Si está seleccionado, crea tablas para la exportación con compresión.

**Compresión para.** Elija el nivel de compresión.

- **Por defecto.** Activa la compresión por defecto (por ejemplo, `CREATE TABLE MYTABLE(...) COMPRESS;` en SQL). En este caso, tiene el mismo efecto que la opción Operaciones de carga directa.

- **Operaciones de carga directa.** Activa la compresión únicamente para operaciones de inserción masiva (ruta directa) (por ejemplo, `CREATE TABLE MYTABLE(...)COMPRESS FOR DIRECT_LOAD OPERATIONS;` en SQL).
- **Todas las operaciones.** Activa la compresión para todas las operaciones (por ejemplo, `CREATE TABLE MYTABLE(...)COMPRESS FOR ALL OPERATIONS;` en SQL).

### **Configuración de Oracle 11gR2: opción Básica**

Estos ajustes se muestran para Oracle 11g R2 Enterprise Edition o Personal Edition que utilice la opción Básica.

**Utilizar compresión.** Si está seleccionado, crea tablas para la exportación con compresión.

**Compresión para.** Elija el nivel de compresión.

- **Por defecto.** Activa la compresión por defecto (por ejemplo, `CREATE TABLE MYTABLE(...) COMPRESS;` en SQL). En este caso, tiene el mismo efecto que la opción Básica.
- **Básica.** Activa la compresión básica (por ejemplo, `CREATE TABLE MYTABLE(...) COMPRESS BASIC;` en SQL).

### **Configuración de Oracle 11gR2: opción Avanzada**

Estos ajustes se muestran para Oracle 11g R2 Enterprise Edition o Personal Edition que utilice la opción Avanzada.

**Utilizar compresión.** Si está seleccionado, crea tablas para la exportación con compresión.

**Compresión para.** Elija el nivel de compresión.

- **Por defecto.** Activa la compresión por defecto (por ejemplo, `CREATE TABLE MYTABLE(...) COMPRESS;` en SQL). En este caso, tiene el mismo efecto que la opción Básica.
- **Básica.** Activa la compresión básica (por ejemplo, `CREATE TABLE MYTABLE(...) COMPRESS BASIC;` en SQL).
- **OLTP.** Activa la compresión OLTP (por ejemplo, `CREATE TABLE MYTABLE(...)COMPRESS FOR OLTP;` en SQL).
- **Consulta baja/alta.** (Servidores Exadata únicamente) Activa la compresión de columna híbrida para la consulta (por ejemplo, `CREATE TABLE MYTABLE(...)COMPRESS FOR QUERY LOW;` o `CREATE TABLE MYTABLE(...)COMPRESS FOR QUERY HIGH;` en SQL). La compresión para consulta es de utilidad en entornos de almacenamiento de datos; HIGH proporciona una relación de compresión mayor que LOW.
- **Archivo bajo/alto.** (Servidores Exadata únicamente) Activa la compresión de columna híbrida para el archivo (por ejemplo, `CREATE TABLE MYTABLE(...)COMPRESS FOR ARCHIVE LOW;` o `CREATE TABLE MYTABLE(...)COMPRESS FOR ARCHIVE HIGH;` en SQL). La compresión para archivo es de utilidad para comprimir datos que se almacenarán durante largos periodos de tiempo; HIGH proporciona una relación de compresión mayor que LOW.

### Configuración de IBM Netezza, IBM DB2 para z/OS y Teradata

Si especifica valores predefinidos de IBM Netezza, IBM DB2 para z/OS o Teradata, se le solicitará que seleccione lo siguiente:

**Utilizar base de datos/esquema del adaptador de puntuación del servidor.** Si selecciona esta opción, se habilitará la opción Base de datos/esquema del adaptador de puntuación del servidor.

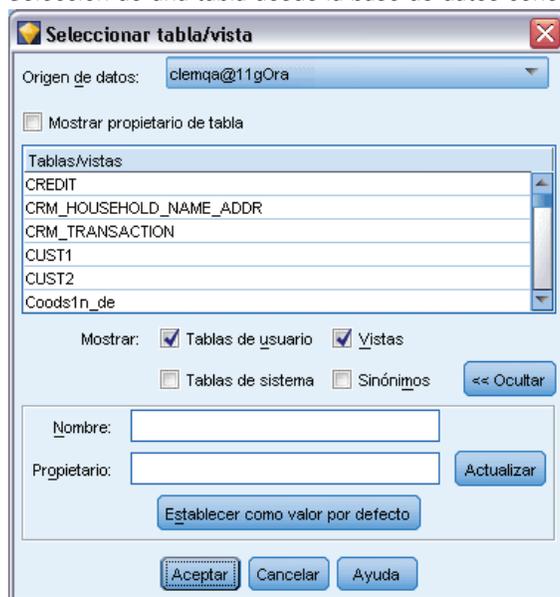
**Base de datos/esquema del adaptador de puntuación del servidor.** En la lista desplegable, seleccione la conexión que necesite.

### Selección de una tabla de la base de datos

Una vez se haya conectado a una fuente de datos, puede elegir importar campos desde una tabla o vista determinada. Desde la pestaña Datos del cuadro de diálogo Base de datos, puede introducir el nombre de una tabla en el campo Nombre de tabla o pulsar en Seleccionar para abrir un cuadro de diálogo que muestre las tablas y vistas disponibles.

Figura 2-8

Selección de una tabla desde la base de datos conectada actualmente



**Mostrar propietario de tabla.** Indique si una fuente de datos necesita que se especifique el propietario de una tabla antes de acceder a ella. Anule la selección de esta opción para las fuentes de datos que no tienen este requisito.

*Nota:* las bases de datos de Oracle y SAS suelen requerir que muestre al propietario de tabla.

**Tablas/vistas.** Seleccione la tabla o vista que se vaya a importar.

**Mostrar.** Muestra las columnas de la fuente de datos a la que está conectado actualmente. Pulse en una de las opciones siguientes para personalizar la vista de las tablas disponibles:

- Pulse en **Tablas de usuario** para ver las tablas normales de la base de datos creadas por los usuarios de la base de datos.
- Pulse en **Tablas de sistema** para ver las tablas de la base de datos que posee el sistema (por ejemplo, las tablas que ofrecen información acerca de la base de datos, como detalles de índices). Esta opción puede utilizarse para ver las pestañas utilizadas en las bases de datos de Excel. (Tenga en cuenta que también existe un nodo de fuente de Excel independiente. Si desea obtener más información, consulte el tema [Nodo de fuente de Excel](#) el p. 53.)
- Pulse en **Vistas** para ver las tablas virtuales basadas en una consulta que implica una o más tablas normales.
- Pulse en **Sinónimos** para ver los sinónimos creados en la base de datos para las tablas existentes.

**Filtros de nombre/propietario.** Estos campos le permiten filtrar la lista de tablas mostradas por nombre o propietario. Por ejemplo, escriba **SYS** para mostrar solamente las tablas con este propietario. Para las búsquedas con comodín, se puede utilizar un guión bajo ( `_` ) para representar un carácter simple y un signo de porcentaje ( `%` ) puede representar una secuencia de cero o más caracteres.

**Establecer como valor por defecto.** Guarda la configuración actual como el valor por defecto para el usuario actual. Esta configuración se restaurará en el futuro cuando un usuario abra un nuevo cuadro de diálogo de selector de tablas *para el mismo nombre de fuente de datos e inicio de sesión solamente del usuario*.

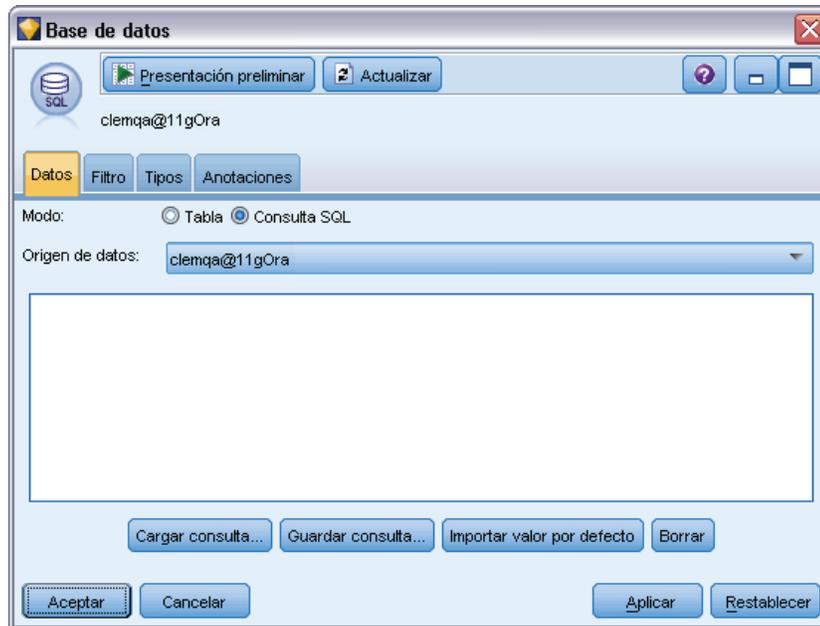
## ***Realización de consultas a la base de datos***

Una vez se haya conectado a una fuente de datos, podrá elegir importar campos mediante consultas SQL. En el cuadro de diálogo principal, seleccione **Consulta SQL** como modo de conexión. Esto añadirá una ventana de editor de consultas al cuadro de diálogo. Mediante el editor de consultas, puede crear o cargar una o más consultas SQL cuyo conjunto de resultados se leerá en la ruta de datos.

Si especifica múltiples consultas SQL, sepárelas con punto y coma ( `;` ) y asegúrese de que no hay varias declaraciones **SELECT**.

Para cancelar y cerrar la ventana de editor de consultas, seleccione **Tabla** como modo de conexión.

Figura 2-9  
Carga de datos mediante consultas SQL



Puede incluir parámetros de ruta de SPSS Modeler (un tipo de variable definida por el usuario) en la consulta SQL. Si desea obtener más información, consulte el tema [Uso de parámetros de ruta en una consulta SQL](#) el p. 24.

**Cargar consulta.** Pulse para abrir el explorador de archivos, que puede utilizar para cargar una consulta guardada anteriormente.

**Guardar consulta.** Pulse para abrir el cuadro de diálogo Guardar consulta, que puede utilizar para guardar la consulta actual.

**Importar valor por defecto.** Pulse para importar una instrucción SQL SELECT de ejemplo generada de forma automática mediante la tabla y columnas seleccionadas en el cuadro de diálogo.

**Borrar.** Borre el contenido del área de trabajo. Utilice esta opción cuando desee comenzar de nuevo.

### ***Uso de parámetros de ruta en una consulta SQL***

Al escribir una consulta SQL para importar campos, puede incluir parámetros de ruta de SPSS Modeler que se hayan definido anteriormente. Se admiten todos los tipos de parámetros de ruta.

La siguiente tabla muestra cómo se interpretarán algunos ejemplos de parámetros de ruta en la consulta SQL.

Tabla 2-2  
Ejemplos de parámetros de ruta

Nombre de parámetro de ruta (ejemplo)	Almacenamiento	Valor del parámetro de ruta	Interpretado como
PString	Cadena	ss	'ss'
PInt	Entero	5	5
PReal	Real	5.5	5.5
PTime	Hora	23:05:01	t{'23:05:01'}
PDate	Fecha	2011-03-02	d{'2011-03-02'}
PTimeStamp	Marca de tiempo	2011-03-02 23:05:01	ts{'2011-03-02 23:05:01'}
PColumn	Desconocido	IntValue	IntValue

En la consulta SQL, especifica un parámetro de ruta del mismo modo que en una expresión CLEM, concretamente por '\$P-<parameter\_name>', donde <parameter\_name> es el nombre que se ha definido para el parámetro de ruta.

Cuando se hace referencia a un campo, el tipo de almacenamiento debe definirse como Desconocido, y el valor del parámetro debe ir entre comillas si es necesario. Por lo tanto, utilizando los ejemplos mostrados en la tabla, si introdujo la consulta SQL:

```
select "IntValue" from Table1 where "IntValue" < '$P-PInt';
```

se evaluará como:

```
select "IntValue" from Table1 where "IntValue" < 5;
```

Si necesita hacer referencia al campo IntValue mediante el parámetro PColumn, deberá especificar la consulta de la manera siguiente para obtener el mismo resultado:

```
select "IntValue" from Table1 where "'$P-PColumn'" < '$P-PInt';
```

## Nodo Archivo variable

Puede utilizar los nodos Archivo variable para leer los datos desde los archivos de texto de campo libre (archivos cuyos registros contienen un número constante de campos pero un número variado de caracteres), también conocido como archivos de texto delimitado. Este tipo de nodo resulta también útil para los archivos con texto de cabecera de longitud fija y determinados tipos de anotaciones. Los registros se leerán uno a uno y pasarán por la ruta hasta que se lea todo el archivo.

### Notas para la lectura de los datos de texto delimitado

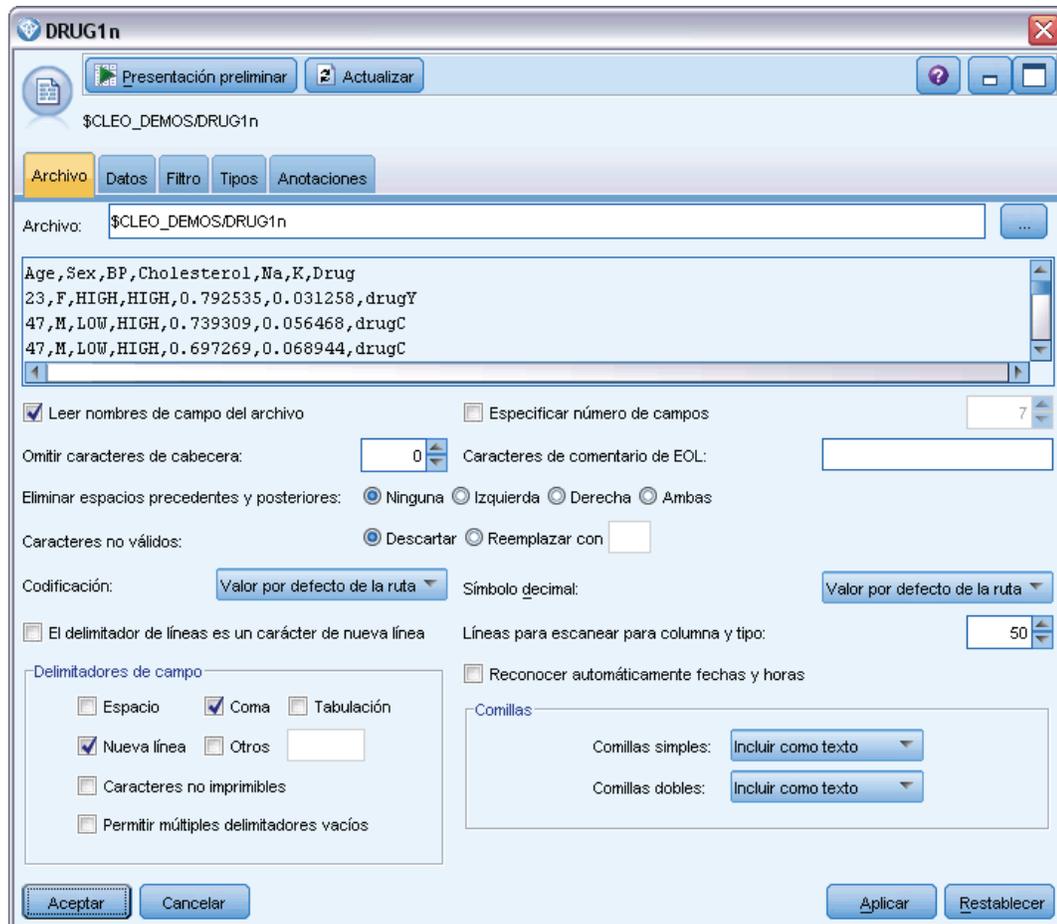
- Los registros se deben delimitar por un carácter de nueva línea en el extremo de cada línea. El carácter de nueva línea no debe utilizarse con ningún otro fin (como dentro de cualquier valor o nombre de campo). Los espacios precedentes y posteriores se deberían eliminar para ahorrar espacio, aunque no es esencial. De forma opcional, se puede eliminar a través del nodo.
- Los campos se deben delimitar por una coma u otro carácter que idealmente se utilice solamente como delimitador, es decir, que no aparezca en valores o nombres de campos. En caso de que no sea posible, todos los campos de texto se pueden escribir entre comillas dobles,

siempre que ninguno de los nombres de campo o valores de texto contenga comillas dobles. Si los nombres o valores contienen comillas dobles, se pueden delimitar los campos de texto con comillas simples, como alternativa, siempre que éstas no se utilicen dentro de valores. Si no se pueden utilizar comillas simples ni dobles, se deberán cambiar los valores de texto para eliminar o sustituir el carácter delimitador o las comillas simples o dobles.

- Cada fila, incluida la fila de la cabecera, debe contener el mismo número de campos.
- La primera línea debería contener los nombres de campo. Si este no es el caso, anule la selección de Leer nombres de campo del archivo para dar a cada campo un nombre genérico como *Campo1*, *Campo2*, etc.
- La segunda línea debería contener el primer registro de los datos. No debe haber línea en blanco ni comentarios.
- Los valores numéricos no deben incluir el separador de millares ni el símbolo de agrupación—sin la coma en 3,000.00, por ejemplo. El indicador decimal (punto) solamente se debe utilizar cuando sea apropiado.
- Los valores de fecha y hora deben estar en uno de los formatos reconocidos en el cuadro de diálogo Opciones de ruta, como DD/MM/YYYY o HH:MM:SS. Todos los campos de fecha y hora del archivo deberían seguir el mismo formato y cualquier campo que contenga una fecha debe utilizar el mismo formato para todos los valores dentro de ese campo.

## Opciones de configuración para el nodo Archivo variable

Figura 2-10  
Cuadro de diálogo del nodo Archivo variable



**Archivo.** Especifique el nombre del archivo. Puede introducir un nombre de archivo o pulsar en el botón de puntos suspensivos (...) para seleccionar un archivo. La ruta de archivo se mostrará cuando haya seleccionado un archivo y su contenido se mostrará con los delimitadores del panel inferior.

El texto de muestra representado desde la fuente de datos, se puede copiar y pegar en los controles siguientes: Caracteres de comentario de EOL y delimitadores especificados por el usuario. Utilice Ctrl-C y Ctrl-V para copiar y pegar.

**Leer nombres de campo del archivo.** Seleccionada por defecto, esta opción trata la primera fila del archivo de datos como etiquetas para la columna. Si la primera fila no es una cabecera, anule la selección para proporcionar a cada campo de forma automática un nombre genérico, como *Campo1*, *Campo2*, para el número de campos del conjunto de datos.

**Especificar número de campos.** Especifique el número de campos de cada registro. El número de campos puede detectarse de forma automática siempre que los registros hayan finalizado en nueva línea. También puede establecer el número de forma manual.

**Omitir caracteres de cabecera.** Especifique los caracteres que desea ignorar al principio del primer registro.

**Caracteres de comentario de EOL.** Especifique los caracteres, como # o !, para indicar anotaciones en los datos. Si uno de estos caracteres aparece en el archivo de datos, se ignorará todo lo que aparezca hasta aquí, sin incluir el siguiente carácter de nueva línea.

**Eliminar espacios precedentes y posteriores.** Seleccione las opciones para descartar los espacios precedentes y posteriores en las cadenas de importación.

*Nota:* Las comparaciones entre cadenas que utilizan y no utilizan retrotracción SQL pueden generar diferentes resultados donde existen espacios precedentes.

**Caracteres no válidos.** Seleccione Descartar para eliminar los caracteres no válidos de origen de datos. Seleccione Reemplazar con para reemplazar los caracteres no válidos con el símbolo especificado (sólo un carácter). Los caracteres no válidos son los caracteres nulos o los que no existen en el método de codificación especificado.

**Codificación.** Especifica el método de codificación de textos utilizado. Puede elegir entre valor por defecto del sistema, el valor por defecto de la ruta o UTF-8.

- El valor por defecto del sistema se especifica en el Panel de control de Windows o, si lo ejecuta en modo distribuido, en el equipo servidor.
- La ruta por defecto se especifica en el cuadro de diálogo Propiedades de ruta.

**Símbolo decimal.** Seleccione el tipo de separador decimal utilizado en la fuente de datos. El Valor por defecto de la ruta es el carácter seleccionado de la pestaña Opciones del cuadro de diálogo de propiedades de la ruta. En el resto de los casos, seleccione Punto (.) o Coma (,) para leer todos los datos en este cuadro de diálogo utilizando el carácter elegido como el separador decimal.

**El delimitador de línea es un carácter de nueva línea.** Para utilizar el carácter de nueva línea como el delimitador de línea en lugar de un delimitador de campo, seleccione esta opción. Por ejemplo, esto puede ser de utilidad si hay un número impar de delimitadores en una fila que hace que haya ajuste de línea. Tenga en cuenta que seleccionar esta opción significa que no podrá seleccionar Nueva línea en la lista Delimitadores.

**Delimitadores.** Mediante las casillas de verificación mostradas para este control, puede especificar los caracteres, como la coma (,), que desea que definan los límites de los campos del archivo. Asimismo, puede especificar más de un delimitador, como as “,|” para los registros que utilizan varios delimitadores. El delimitador por defecto es la coma.

*Nota:* si la coma se ha definido también como el separador decimal, los valores por defecto no funcionarán aquí. En aquellos casos en los que la coma sea el delimitador del campo y el separador decimal, seleccione Otros en la lista Delimitadores. A continuación, especifique una coma en el campo de entrada.

Seleccione Permitir múltiples delimitadores vacíos para tratar varios caracteres delimitadores vacíos adyacentes como un solo delimitador. Por ejemplo, si un valor de datos aparece seguido de cuatro espacios y otro valor de datos, este grupo será tratado como dos campos en lugar de cinco.

**Líneas que explorar para columna y tipo.** Especifique cuántas líneas y columnas se van a explorar para los tipos de datos especificados.

**Reconocer automáticamente fechas y horas.** Para que IBM® SPSS® Modeler intente reconocer automáticamente entradas de datos como fechas u horas, active esta casilla de verificación. Por ejemplo, significa que una entrada del tipo 07-11-1965 se identificará como fecha y una del tipo 02:35:58 se identificará como hora; sin embargo, las entradas ambiguas como 07111965 ó 023558 se mostrarán como enteros ya que no hay delimitadores entre los números.

*Nota:* para evitar posibles problemas de datos cuando utilice archivos de datos de versiones anteriores de SPSS Modeler, esta casilla se desactiva por defecto para la información guardada en versiones anteriores a 13.

**Comillas.** Si utiliza las listas desplegables, podrá especificar las comillas simples y dobles tratadas en la importación. Puede seleccionar Descartar todas las comillas, Incluir como texto incluyéndolas en el valor del campo o Emparejar y descartar para hacer coincidir las parejas de comillas y eliminarlas. Si alguna comilla no coincide, recibirá un mensaje de error. Tanto Descartar como Emparejar y descartar almacenan el valor del campo (sin comillas) como una cadena.

Mientras trabaja en este cuadro de diálogo, pulse en Actualizar en cualquier momento para volver a cargar los campos desde la fuente de datos. Esto resulta útil cuando se modifican las conexiones de datos en el nodo de fuente o cuando se trabaja entre pestañas en el cuadro de diálogo.

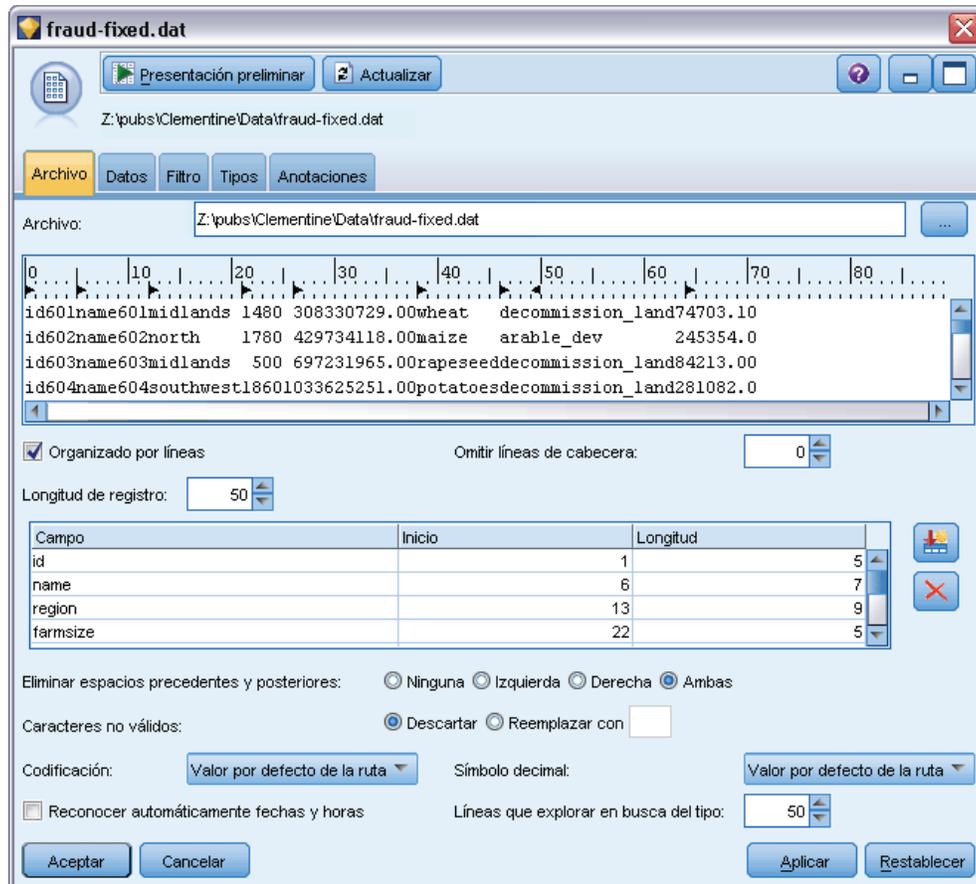
## ***Nodo Archivo fijo***

Puede utilizar los nodos Archivo fijo para importar los datos desde los archivos de texto de campo fijo (archivos cuyos campos no están delimitados pero empiezan en la misma posición y tienen una longitud fija). Los datos heredados o generados por la máquina se suelen almacenar en formato de campo fijo. Mediante la pestaña Archivo del nodo Archivo fijo, podrá especificar fácilmente la posición y longitud de las columnas en los datos.

### ***Configuración de opciones para el nodo Archivo fijo***

La pestaña Archivo del nodo Archivo fijo le permite introducir datos en IBM® SPSS® Modeler y especificar la posición de las columnas y la longitud de los registros. Con el panel de presentación preliminar de los datos en el centro del cuadro de diálogo, puede pulsar para añadir flechas que especifiquen los puntos de separación entre los campos.

Figura 2-11  
Especificación de columnas en los datos de campo fijo



**Archivo.** Especifique el nombre del archivo. Puede introducir un nombre de archivo o pulsar en el botón de puntos suspensivos (...) para seleccionar un archivo. La ruta de archivo se mostrará cuando haya seleccionado un archivo y su contenido se mostrará con los delimitadores en el panel inferior.

El panel de presentación preliminar de los datos se puede utilizar para especificar la posición y longitud de la columna. La regla situada en la parte superior de la ventana de presentación preliminar ayuda a medir la longitud de las variables y especifica el punto de separación entre ellas. Puede especificar las líneas del punto de separación pulsando en el área de la regla por encima de los campos. Los puntos de separación se pueden mover arrastrándolos y se pueden descartar arrastrándolos fuera de la región de presentación preliminar de los datos.

- Cada línea del punto de separación añade de forma automática un nuevo campo a la tabla de campos inferior.
- Las posiciones de inicio que indican las flechas se añaden de forma automática a la columna Inicio de la tabla inferior.

**Organizado por líneas.** Indique si desea omitir el carácter de nueva línea al final de cada registro.

**Omitir líneas de cabecera.** Especifique las líneas que desea ignorar al principio del primer registro. Esto resulta útil para ignorar las cabeceras de columna.

**Longitud de registro.** Especifique el número de caracteres de cada registro.

**Campo.** Todos los campos definidos para este archivo de datos se muestran aquí. Hay dos formas de definir los campos:

- Especifique los campos de forma interactiva utilizando el panel anterior de presentación preliminar de los datos.
- Especifique los campos de forma manual añadiendo las filas de campos vacíos a la tabla inferior. Pulse en el botón a la derecha del panel de campos para añadir nuevos campos. A continuación, en el campo vacío, escriba un nombre de campo, una posición de inicio y una longitud. Estas opciones añadirán flechas de forma automática al panel de presentación preliminar de los datos, lo cual se puede ajustar fácilmente.

Para eliminar un campo definido previamente, seleccione el campo en la lista y pulse en el botón rojo de eliminación.

**Inicio.** Especifique la posición del primer carácter del campo. Por ejemplo, si el segundo campo de un registro comienza en el decimosexto carácter, deberá introducir 16 como punto de partida.

**Longitud.** Especifique cuántos caracteres hay en el valor más largo de cada campo. Esto determina el punto de corte para el siguiente campo.

**Eliminar espacios precedentes y posteriores.** Seleccione descartar los espacios precedentes y posteriores en las cadenas de importación.

*Nota:* Las comparaciones entre cadenas que utilizan y no utilizan retrotracción SQL pueden generar diferentes resultados donde existen espacios precedentes.

**Caracteres no válidos.** Seleccione Descartar para eliminar los caracteres no válidos de la entrada de datos. Seleccione Reemplazar con para reemplazar los caracteres no válidos con el símbolo especificado (solamente un carácter). Los caracteres no válidos son los caracteres nulos (0) o los que no existen en la codificación actual.

**Codificación.** Especifica el método de codificación de textos utilizado. Puede elegir entre valor por defecto del sistema, el valor por defecto de la ruta o UTF-8.

- El valor por defecto del sistema se especifica en el Panel de control de Windows o, si lo ejecuta en modo distribuido, en el equipo servidor.
- La ruta por defecto se especifica en el cuadro de diálogo Propiedades de ruta.

**Símbolo decimal.** Seleccione el tipo de separador decimal utilizado en la fuente de datos. El Valor por defecto de la ruta es el carácter seleccionado de la pestaña Opciones del cuadro de diálogo de propiedades de la ruta. En el resto de los casos, seleccione Punto (.) o Coma (,) para leer todos los datos en este cuadro de diálogo utilizando el carácter elegido como el separador decimal.

**Reconocer automáticamente fechas y horas.** Para que SPSS Modeler intente reconocer automáticamente entradas de datos como fechas u horas, active esta casilla de verificación. Por ejemplo, significa que una entrada del tipo 07-11-1965 se identificará como fecha y una del tipo 02:35:58 se identificará como hora; sin embargo, las entradas ambiguas como 07111965 ó 023558 se mostrarán como enteros ya que no hay delimitadores entre los números.

*Nota:* para evitar posibles problemas de datos cuando utilice archivos de datos de versiones anteriores de SPSS Modeler, esta casilla se desactiva por defecto para la información guardada en versiones anteriores a 13.

**Líneas que explorar en busca del tipo.** Especifique cuántas líneas se van a explorar para los tipos de datos especificados.

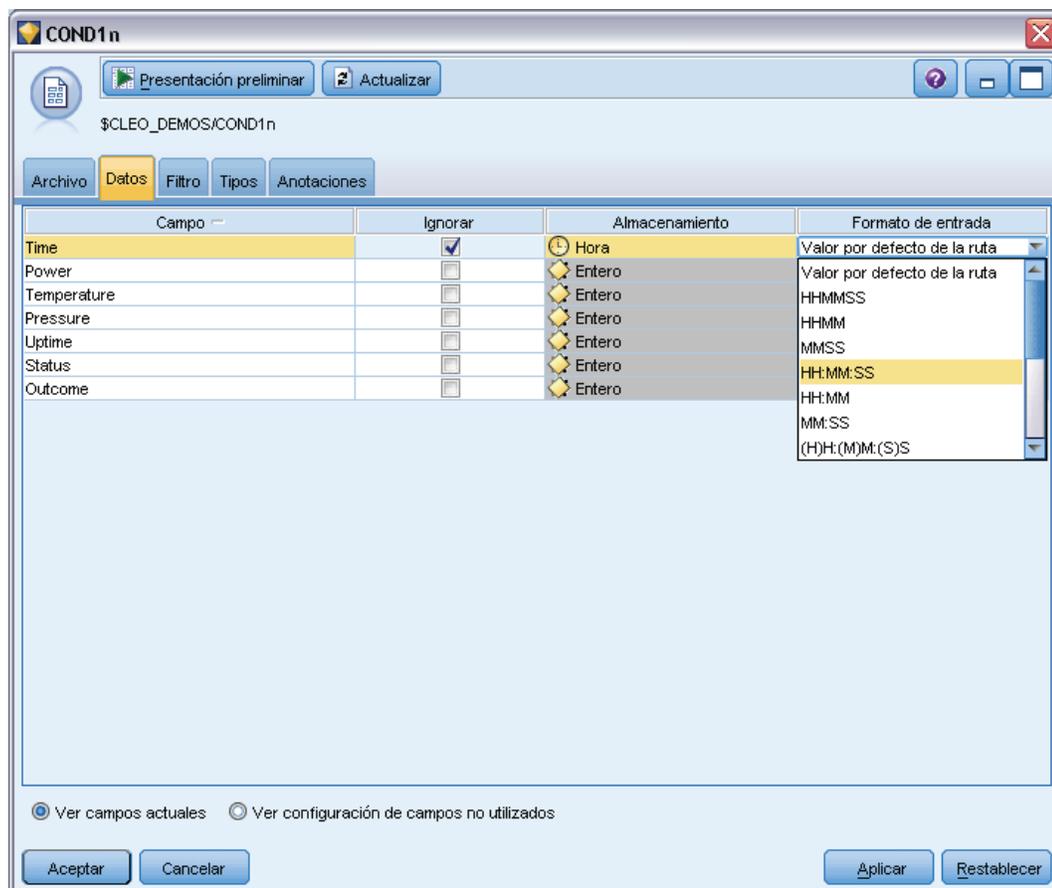
Mientras trabaja en este cuadro de diálogo, pulse en Actualizar en cualquier momento para volver a cargar los campos desde la fuente de datos. Esto resulta útil cuando se modifican las conexiones de datos en el nodo de fuente o cuando se trabaja entre pestañas en el cuadro de diálogo.

## ***Configuración de almacenamiento de campos y formato***

Las opciones de la pestaña Datos para los nodos Archivo fijo, Archivo variable, Fuente XML y Datos usuario permiten especificar el tipo de almacenamiento para los campos, como se importan o crean en IBM® SPSS® Modeler. Para los nodos Archivo fijo, Archivo variable y Datos Usuario, también puede especificar el formato de campo u otros metadatos.

Para leer datos de otras fuentes, el almacenamiento viene determinado de manera automática aunque se puede modificar mediante una función de conversión, como `to_integer`, en un nodo Rellenar o un nodo Derivar.

Figura 2-12  
Sustitución del tipo de almacenamiento y asignación de formato al campo en la importación



**Campo.** Utilice la columna *Campo* para ver y seleccionar campos en el conjunto de datos actual.

**Ignorar.** Seleccione la casilla de verificación en la columna *Ignorar* para activar las opciones de las columnas *Almacenamiento* y *Formato de entrada*.

### **Almacenamiento de datos**

Almacenamiento describe la forma en la que los datos se almacenan en un campo. Por ejemplo, un campo con valores 1 y 0 almacena datos enteros. Esto es diferente al nivel de medición, que describe el uso de los datos y no afecta al almacenamiento. Por ejemplo, puede que desee definir el nivel de medición para un campo de enteros con valores de 1 y 0 para una *Marca*. Esto suele indicar que 1=*Verdadero* y 0=*Falso*. Aunque el almacenamiento se debe determinar en el origen, el nivel de medición se puede cambiar utilizando un nodo *Tipo* en cualquier punto de la ruta. Si desea obtener más información, consulte el tema [Niveles de medida](#) en el capítulo 4 el p. 137.

Los tipos de almacenamiento disponibles son los siguientes:

- **Cadena.** Se utiliza para campos que contienen datos no numéricos, también llamados datos alfanuméricos. Una cadena puede incluir cualquier secuencia de caracteres, como *fred*, *Clase 2* o *1234*. Tenga en cuenta que los números de las cadenas no se pueden utilizar en los cálculos.

- **Entero** Un campo cuyos valores son números enteros.
- **Real.** Los valores son números que pueden incluir decimales (no está limitado a números enteros). El formato de presentación se especifica en el cuadro de diálogo Propiedades de ruta y puede quedar anulado para campos individuales en un nodo Tipo (ficha Formato).
- **Fecha.** Valores de fecha especificados en un formato estándar como año, mes y día (por ejemplo, 26.09.07). El formato determinado se especifica en el cuadro de diálogo Propiedades de ruta.
- **Hora.** Tiempo medido como una duración. Por ejemplo, una llamada de servicio que dure 1 hora, 26 minutos y 38 segundos se puede representar como 01:26:38, según el formato de hora actual que se especifica en el cuadro de diálogo Propiedades de ruta.
- **Marca de tiempo.** Valores que incluyen un componente de fecha y hora, por ejemplo 2007–09–26 09:04:00, siguen dependiendo de los formatos actuales de fecha y hora del cuadro de diálogo Propiedades de ruta. Tenga en cuenta que es posible que los valores de marca de tiempo se tengan que escribir entre comillas dobles para garantizar que se interpretan como un valor único en lugar de valores de fecha y hora diferentes. (Es aplicable, por ejemplo cuando introduce valores en un nodo Datos Usuario.)

**Conversiones del almacenamiento.** Puede convertir el almacenamiento de un campo mediante una serie de funciones de conversión, como `to_string` y `to_integer`, en un nodo Rellenar. Si desea obtener más información, consulte el tema [Conversión del almacenamiento mediante el nodo Rellenar](#) en el capítulo 4 el p. 179. Tenga en cuenta que las funciones de conversión (y muchas otras funciones que requieren un tipo específico de entrada como un valor de fecha y hora) dependen de los formatos actuales especificados en el cuadro de diálogo Propiedades de ruta. Por ejemplo, si desea convertir un campo de cadena con los valores *Ene 2003*, *Feb 2003*, (etc.) para almacenar datos, seleccione MES AAAA como formato de fecha por defecto de la ruta. Las funciones de conversión también están disponibles en el nodo Derivar para realizar la conversión temporal durante un cálculo de derivación. También puede utilizar el nodo Derivar para realizar otras manipulaciones, como volver a codificar los campos con valores categóricos. Si desea obtener más información, consulte el tema [Nueva codificación de valores con el nodo Derivar](#) en el capítulo 4 el p. 177.

**Lectura de datos mezclados.** Tenga en cuenta que al leer los campos con almacenamiento numérico (entero, real, hora, marca de tiempo o fecha), cualquier valor no numérico se establece como nulo o perdido por el sistema. Esto ocurre porque, a diferencia de otras aplicaciones, SPSS Modeler no permite la combinación de tipos de almacenamiento dentro de un campo. Para evitarlo, cualquier campo con datos mezclados debe leerse como cadenas, ya sea cambiando el tipo de almacenamiento en el nodo de origen o en la aplicación externa, según sea necesario.

### ***Formato de entrada de campo (únicamente los nodos Archivo fijo, Archivo variable y Datos usuario)***

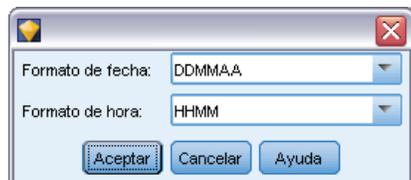
Para todos los tipos de almacenamiento, excepto Cadena y Entero, puede especificar las opciones de formato para el campo seleccionado mediante la lista desplegable. Por ejemplo, al fusionar datos desde distintas configuraciones regionales, puede que necesite especificar un punto (.) como el separador decimal para un campo, mientras que otro necesitará una coma como separador.

Las opciones de entrada especificadas en el nodo de fuente omiten las opciones de formato especificadas en el cuadro de diálogo de propiedades de la ruta; sin embargo, no permanecen más adelante en la ruta. Pretenden realizar el análisis de la entrada correctamente, según el

conocimiento que se tenga de los datos. Los formatos especificados se utilizan como un manual para realizar el análisis de los datos tal y como se leen en SPSS Modeler y no para determinar cómo se debe asignar el formato una vez leídos en SPSS Modeler. Para especificar el formato en una base por campo en otro lugar de la ruta, utilice la pestaña Formato de un nodo Tipo. Si desea obtener más información, consulte el tema [Pestaña Configuración de Formato de campo](#) en el capítulo 4 el p. 151.

Figura 2-13

*Especificación de los formatos de fecha y hora para los campos de marca de tiempo*



Las opciones varían en función del tipo de almacenamiento. Por ejemplo, para el tipo de almacenamiento Real, puede seleccionar Punto (.) o Coma (,) como separador decimal. Para los campos de marca de tiempo, se abrirá un cuadro de diálogo independiente cuando seleccione Especificar de la lista desplegable. Si desea obtener más información, consulte el tema [Opciones de configuración de Formato de campo](#) en el capítulo 4 el p. 152.

Para todos los tipos de almacenamiento, puede seleccionar también Valor por defecto de la ruta de forma que pueda utilizar los valores por defecto de la ruta para la importación. La configuración de la ruta se especifica en el cuadro de diálogo de propiedades de la ruta.

### **Opciones adicionales**

Hay otras opciones que se pueden especificar mediante la pestaña Datos:

- Si desea ver la configuración del almacenamiento para los datos que ya no están conectados a través del nodo actual (por ejemplo, datos de entrenamiento), seleccione Ver configuración de campos no utilizados. Puede borrar los campos heredados pulsando en Borrar.
- Mientras trabaja en este cuadro de diálogo, pulse en Actualizar en cualquier momento para volver a cargar los campos desde la fuente de datos. Esto resulta útil cuando se modifican las conexiones de datos en el nodo de fuente o cuando se trabaja entre pestañas en el cuadro de diálogo.

## **Nodo Data Collection**

Los nodos de fuente Data Collection importan datos de encuesta basándose en Survey Reporter Developer Kit de IBM® SPSS® Data Collection utilizado por el software de investigación de mercados de IBM Corp. Este formato distingue los **datos de casos**, las respuestas reales a las preguntas recogidas durante una encuesta, de los **metadatos** que describen cómo se recopilan y organizan los datos de casos. Los metadatos constan de información como textos de preguntas, nombres de variables y descripciones, definiciones de variable de respuestas múltiples, traducciones de las cadenas de texto y la definición de la estructura de los datos de casos.

*Nota:* este nodo requiere Data Collection Survey Reporter Developer Kit, que se distribuye con los productos de software de Data Collection en IBM Corp.. Para obtener más información, consulte la página Web de Data Collection en <http://www.ibm.com/software/analytics/spss/products/data-collection/survey-reporter-dev-kit/>. Además de la instalación de Developer Kit, no se requieren configuraciones adicionales.

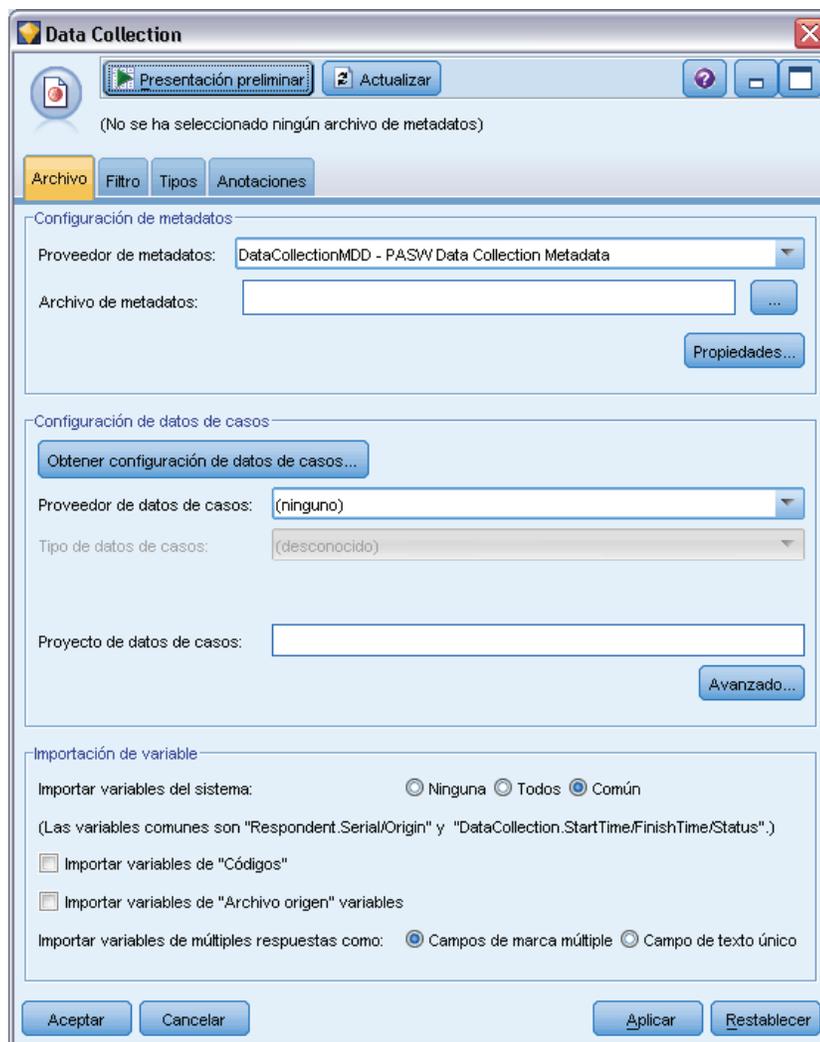
### **Comentarios**

- Los datos de encuesta se leen solamente desde el formato VDATA tabular y plano o desde las fuentes en el formato HDATA jerárquico si incluyen una fuente de metadatos (requiere Data Collection 4.5 o superiores).
- Los tipos se instancian con la información de los metadatos.
- Cuando se importan los datos de encuesta a IBM® SPSS® Modeler, las preguntas se representan como campos, con un registro para cada encuestado.

### **Opciones de archivo de importación de Data Collection**

La pestaña Archivo del nodo Data Collection permite especificar opciones para los metadatos y los datos de casos que desea importar.

Figura 2-14  
Opciones de archivo del nodo de fuente Data Collection



### Configuración de metadatos

*Nota:* Para ver la lista completa de tipos de archivos de proveedores, es necesario que instale Survey Reporter Developer Kit de IBM® SPSS® Data Collection, disponible con el software Data Collection. Si desea obtener más información, consulte la página Web de Data Collection en <http://www.ibm.com/software/analytics/spss/products/data-collection/survey-reporter-dev-kit/>.

**Proveedor de metadatos.** Los datos de encuesta se pueden importar desde distintos formatos compatibles con el software Data Collection Survey Reporter Developer Kit. Entre los tipos de proveedor disponibles se encuentran:

- DataCollectionMDD. Lee metadatos de un archivo de definición de cuestionario (.mdd). Es el formato estándar de Data Collection Data Model.

- Base de datos de ADO. Lee datos de casos y metadatos de archivos ADO. Especifique el nombre y la ubicación del archivo *.adoinfo* que contiene los metadatos. El nombre interno de este DSC es *mrADODsc*.
- Base de datos de In2data. Lee metadatos y datos de casos de In2data. El nombre interno de este DSC es *mrI2dDsc*.
- Archivo de registro de recopilación de datos. Lee metadatos de un archivo de registro estándar de Data Collection. Normalmente, los archivos de registro tienen una extensión de nombre de archivo *.tmp*. Sin embargo, algunos archivos de registro pueden tener otra extensión de nombre de archivo. Si es necesario, puede cambiar el nombre del archivo para que tenga una extensión de nombre de archivo *.tmp*. El nombre interno de este DSC es *mrLogDsc*.
- Archivo de definiciones de Quancept. Convierte los metadatos en un proceso de Quancept. Especifique el nombre del archivo *.qdi* de Quancept. El nombre interno de este DSC es *mrQdiDrsDsc*.
- Base de datos de Quanvert. Lee metadatos y datos de casos de Quanvert. Especifique el nombre y la ubicación del archivo *.qvinfo* o el archivo *.pkd*. El nombre interno de este DSC es *mrQvDsc*.
- Base de datos de participación de recopilación de datos. Lee tablas de muestra y de historial del proyecto y crea variables categóricas derivadas correspondientes a las columnas de esas tablas. El nombre interno de este DSC es *mrSampleReportingMDSC*.
- Archivo Statistics. Lee metadatos y datos de casos de un archivo IBM® SPSS® Statistics *.sav* de SPSS. Escribe datos de casos en un archivo SPSS Statistics *.sav* para su análisis en SPSS Statistics. Escribe metadatos desde un archivo *.sav* de SPSS Statistics a un archivo *.mdd*. El nombre interno de este DSC es *mrSavDsc*.
- Archivo Surveycraft. Lee metadatos y datos de caso de SurveyCraft. Especifique el nombre del archivo *.vq* de SurveyCraft. El nombre interno de este DSC es *mrSCDsc*.
- Archivo de procesos de recopilación de datos. Lee metadatos en un archivo *mrScriptMetadata*. Normalmente, estos archivos tienen una extensión de nombre de archivo *.mdd* o *.dms*. El nombre interno de este DSC es *mrScriptMDSC*.
- Archivo XML triple-S Lee metadatos de un archivo triple-S en formato XML. El nombre interno de este DSC es *mrTripleSDsc*.

**Propiedades de metadatos.** Opcionalmente, seleccione Propiedades para especificar la versión de encuesta que se va a importar así como el idioma, contexto y tipo de etiqueta que se van a utilizar. Si desea obtener más información, consulte el tema [Propiedades de metadatos de importación de IBM SPSS Data Collection](#) el p. 40.

### **Configuración de datos de casos**

*Nota:* Para ver la lista completa de tipos de archivos de proveedores, es necesario que instale Survey Reporter Developer Kit de Data Collection, disponible con el software Data Collection. Si desea obtener más información, consulte la página Web de Data Collection en <http://www.ibm.com/software/analytics/spss/products/data-collection/survey-reporter-dev-kit/>.

**Obtener configuración de datos de casos.** Al leer metadatos solamente desde archivos *.mdd*, pulse en Obtener configuración de datos de casos para determinar las fuentes de datos de casos que están asociadas a los metadatos seleccionados, junto con la configuración específica necesaria

para acceder a una fuente determinada. Esta opción solamente se encuentra disponible para archivos *.mdd*.

**Proveedor de datos de casos.** Los siguientes tipos de proveedor son compatibles:

- Base de datos de ADO. Lee datos de casos mediante la interfaz de Microsoft ADO. Seleccione UDL de OLE-DB para el tipo de datos de casos y especifique una cadena de conexión en el campo UDL de casos de datos. Si desea obtener más información, consulte el tema [Cadena de conexión a la base de datos](#) el p. 41. El nombre interno de este componente es *mrADODsc*.
- Archivo de texto delimitado (Excel). Lee datos de casos de un archivo delimitado por comas (.CSV), de manera que se pueda abrir con Excel. El nombre interno es *mrCsvDsc*.
- Archivo de datos de recopilación de datos. Lee datos de casos de un archivo de formato de datos de Data Collection nativo (Data Collection 4.5 en adelante). El nombre interno es *mrDataFileDsc*.
- Base de datos de In2data. Lee metadatos y datos de casos de un archivo de base de datos de In2data (*.i2d*). El nombre interno es *mrI2dDsc*.
- Archivo de registro de recopilación de datos. Lee datos de casos de un archivo de registro de Data Collection estándar. Normalmente, los archivos de registro tienen una extensión de nombre de archivo *.tmp*. Sin embargo, algunos archivos de registro pueden tener otra extensión de nombre de archivo. Si es necesario, puede cambiar el nombre del archivo para que tenga una extensión de nombre de archivo *.tmp*. El nombre interno es *mrLogDsc*.
- Archivo de datos de Quantum. Lee datos de casos de cualquier archivo ASCII con formato Quantum (*.dat*). El nombre interno es *mrPunchDsc*.
- Archivo de datos de Quancept. Lee datos de casos de un archivo *.drs*, *.drz* o *.dru* de Quancept. El nombre interno es *mrQdiDrsDsc*.
- Base de datos de Quanvert. Lee datos de casos de un archivo *qvinfo* o *.pkd* de Quanvert. El nombre interno es *mrQvDsc*.
- Base de datos de recopilación de datos (MS SQL Server). Lee datos de casos en una base de datos relacional de Microsoft SQL Server. Si desea obtener más información, consulte el tema [Cadena de conexión a la base de datos](#) el p. 41. El nombre interno es *mrRdbDsc2*.
- Archivo Statistics. Lee datos de casos desde un archivo SPSS Statistics *.sav*. El nombre interno es *mrSavDsc*.
- Archivo Surveycraft. Lee datos de casos de un archivo *.qdt* de SurveyCraft. Tanto los archivos *.vq* como los archivos *.qdt* deben estar en el mismo directorio, con acceso de lectura y escritura para ambos archivos. Los archivos no se crean así por defecto al utilizar SurveyCraft, por lo que se deben mover para importar los datos de SurveyCraft. El nombre interno es *mrScDsc*.
- Archivo de datos triple-S. Lee datos de casos de un archivo de datos Triple-S, en formato de longitud fija o delimitado por comas. El nombre interno es *mr TripleDsc*.
- XML de recopilación de datos. Lee datos de casos de un archivo de datos XML de Data Collection. Normalmente, este formato se puede utilizar para transferir datos de casos de una ubicación a otra. El nombre interno es *mrXmlDsc*.

**Tipo de datos de casos.** Especifica si los datos de casos se leen desde un archivo, carpeta, UDL de OLE-DB o DSN de ODBC, y actualiza las opciones del cuadro de diálogo que correspondan. Las opciones válidas dependen del tipo de proveedor. Puede especificar las opciones de conexión

ODBC u OLE-DB para los proveedores de bases de datos. Si desea obtener más información, consulte el tema [Cadena de conexión a la base de datos](#) el p. 41.

**Proyecto de datos de casos.** Al leer datos de casos de una base de datos de Data Collection, puede escribir el nombre del proyecto. Para el resto de tipos de datos de casos, esta configuración se deberá dejar en blanco.

### **Variable Importar**

**Importar variables del sistema.** Especifica si las variables del sistema se importan, incluidas las variables que indican el estado de intervención (en proceso, completado, fecha de finalización, etc.). Puede seleccionar Ninguna, Todos o Común.

**Importar variables “Códigos”.** Controla la importación de variables que representan los códigos utilizados para “otras” respuestas abiertas para las variables categóricas.

**Importar variables “SourceFile”.** Controla la importación de variables que contienen nombres de archivo de imágenes de respuestas exploradas.

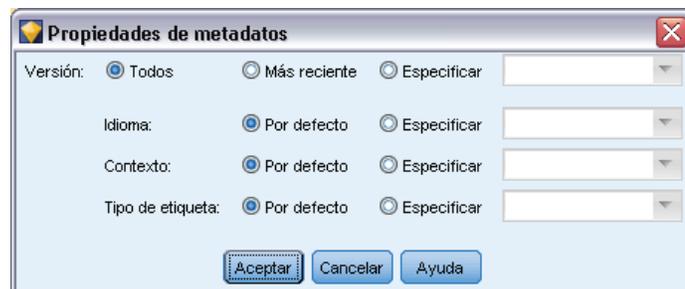
**Importar variables de respuestas múltiples como.** Las variables de respuestas múltiples se pueden importar como varios campos de marcas (un conjunto de dicotomía múltiple), que es el método por defecto de las nuevas rutas. Las rutas creadas en las versiones de IBM® SPSS® Modeler anteriores a la 12.0 importaban respuestas múltiples en un solo campo, con valores separados por comas. El método anterior aún se admite para permitir que las rutas existentes se ejecuten igual que anteriormente, pero se recomienda la actualización de las rutas más antiguas para utilizar el nuevo método. Si desea obtener más información, consulte el tema [Importación de conjuntos de respuestas múltiples](#) el p. 42.

## **Propiedades de metadatos de importación de IBM SPSS Data Collection**

Al importar datos de encuesta de IBM® SPSS® Data Collection, puede especificar la versión de encuesta que se va a importar así como el idioma, contexto y tipo de etiqueta que se van a utilizar. Tenga en cuenta que solamente se puede importar un idioma, contexto y tipo de etiqueta al mismo tiempo.

Figura 2-15

*Propiedades de metadatos de importación de IBM SPSS Data Collection*



**Versión.** Cada versión de encuesta puede considerarse como una instantánea de los metadatos que se utilizan para recopilar un conjunto de datos de casos determinado. Como se realizan cambios en los cuestionarios, se pueden crear varias versiones. Puede importar la última versión, todas las versiones o una versión específica.

- **Todas las versiones.** Seleccione esta opción si desea utilizar una combinación (superconjunto) de todas las versiones disponibles. (En ocasiones, esto se denomina superversión). Cuando existe un conflicto entre las versiones, las versiones más recientes suelen tener prioridad sobre las versiones anteriores. Por ejemplo, si una etiqueta de categoría es diferente en alguna de las versiones, se utilizará el texto de la última versión.
- **Última versión.** Seleccione esta opción si desea utilizar la versión más reciente.
- **Especificar versión.** Seleccione esta opción si desea utilizar una versión de encuesta determinada.

Seleccionar todas las versiones es útil cuando desea exportar datos de casos para más de una versión y se han hecho cambios en las definiciones de variable y categoría, lo que supone que los datos de casos recopilados con una versión no son válidos en otra versión. Seleccionar todas las versiones para las que desea exportar los datos de casos significa que podrá exportar los datos de casos recopilados con las distintas versiones al mismo tiempo sin que se produzcan errores de validación debido a las diferencias entre versiones. Sin embargo, según los cambios de versión, puede producirse algún error de validación.

**Idioma.** Se pueden almacenar preguntas y textos asociados en varios idiomas en los metadatos. Puede utilizar el idioma por defecto para la encuesta o especificar un idioma determinado. Si un elemento no está disponible en el idioma especificado, se utilizará el establecido por defecto.

**Contexto.** Seleccione el contexto de usuario que desee utilizar. El contexto de usuario controla los textos que se muestran. Por ejemplo, seleccione *Pregunta* para mostrar textos de preguntas o *Análisis* para mostrar textos más cortos adecuados para visualizarlos al analizar los datos.

**Tipo de etiqueta.** Enumera los tipos de etiqueta que se han definido. El valor por defecto es *etiqueta*, que se utiliza para textos de preguntas en el contexto de usuario *Pregunta* y las descripciones de variables en el contexto de usuario *Análisis*. Se pueden definir otros tipos de etiquetas para las instrucciones, descripciones, etc.

## ***Cadena de conexión a la base de datos***

Al utilizar el nodo IBM® SPSS® Data Collection para importar datos de casos de una base de datos a través de OLE-DB o ODBC, seleccione *Edición* en la pestaña *Archivo* para acceder al cuadro de diálogo de cadena de conexión, que permite personalizar la cadena de conexión que se pasa al proveedor para ajustar la conexión.

Figura 2-16  
Cadena de conexión de importación de IBM SPSS Data Collection

### Propiedades avanzadas

Al utilizar el nodo IBM® SPSS® Data Collection para importar datos de casos de una base de datos que requiere un inicio de sesión explícito, seleccione Avanzado para proporcionar un ID de usuario y una contraseña para acceder a la fuente de datos.

Figura 2-17  
Propiedades avanzadas de importación de IBM SPSS Data Collection

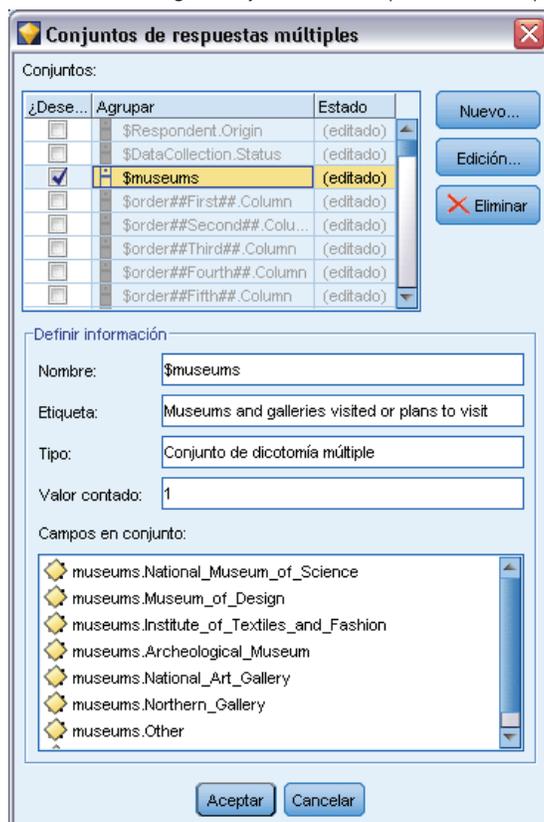
### Importación de conjuntos de respuestas múltiples

Las variables de respuestas múltiples se pueden importar de IBM® SPSS® Data Collection como conjuntos de dicotomía múltiple con un campo de marca independiente para cada posible valor de la variable. Por ejemplo, si se pide a los encuestados que seleccionen qué museos han visitado de una lista, el conjunto incluiría un campo de marca independiente para cada museo enumerado.

Figura 2-18  
Pregunta de respuestas múltiples

Tras importar los datos, puede añadir o editar conjuntos de respuestas múltiples desde cualquier nodo que incluya una pestaña Filtro. Si desea obtener más información, consulte el tema [Edición de conjuntos de respuestas múltiples](#) en el capítulo 4 el p. 158.

Figura 2-19  
Cuadro de diálogo Conjuntos de respuestas múltiples



### ***Importación de respuestas múltiples en un único campo (para rutas creadas en versiones anteriores)***

En las versiones de IBM® SPSS® Modeler anteriores en lugar de importar respuestas múltiples como se ha descrito antes, se importaban en un solo campo, con valores separados por comas. Este método aún se admite para compatibilizar las rutas existentes, pero se recomienda que cualquier ruta se actualice para utilizar el nuevo método.

## ***Notas de importación de columna de IBM SPSS Data Collection***

Las columnas de los datos de IBM® SPSS® Data Collection se leen en IBM® SPSS® Modeler tal y como se resume en la siguiente tabla.

Tipo de columna de Data Collection	Almacenamiento de SPSS Modeler	Nivel de medida
Marca booleana (sí/no)	Cadena	Marca (valores 0 y 1)
Categorico	Cadena	Nominal

Tipo de columna de Data Collection	Almacenamiento de SPSS Modeler	Nivel de medida
Marca de fecha o tiempo	Marca de tiempo	Continuo
Dobles (valor de punto flotante en un rango especificado)	Real	Continuo
Largo (valor de punto flotante en un rango especificado)	Entero	Continuo
Texto (descripción de texto libre)	Cadena	Sin tipo
Nivel (indica cuadrículas o bucles de una pregunta)	No sucede en VDATA y no se importa a SPSS Modeler	
Objeto (datos binarios como un facsímil que muestra texto escrito a mano o una grabación de voz)	No se importa a SPSS Modeler	
Ninguno (tipo desconocido)	No se importa a SPSS Modeler	
Columna de encuestados en serie (asocia un ID único a cada encuestado)	Entero	Sin tipo

Para evitar posibles incoherencias entre las etiquetas de valor que se leen desde los metadatos y los valores reales, todos los valores de los metadatos se convierten a minúsculas. Por ejemplo, el valor de la etiqueta *E1720\_years* se convertiría a *e1720\_years*.

## ***Nodo de fuente de IBM Cognos BI***

El nodo de fuente de IBM Cognos BI le permite introducir los datos de la base de datos de Cognos BI o informes de lista individuales en la sesión de minería de datos. De este modo, puede combinar las funciones de inteligencia empresarial de Cognos con las capacidades de análisis predictivo de IBM® SPSS® Modeler. Puede importar datos relacionales, relacionales modelados dimensionalmente (DMR) y OLAP.

Desde una conexión de servidor de Cognos, primero seleccione una ubicación a la que importar datos o informes. Una ubicación contiene un modelo de Cognos y todas las carpetas, consultas, informes, vistas, métodos abreviados, URL y definiciones de trabajo asociados a dicho modelo. Un modelo de Cognos define reglas empresariales, descripciones de datos, relaciones de datos, dimensiones y jerarquías empresariales y otras tareas administrativas.

Si está importando datos, seleccione los objetos que quiere importar desde el paquete seleccionado. Entre los objetos que puede importar se incluyen asuntos de consulta (que representan tablas de bases de datos) o elementos de consulta individuales (que representan las columnas de tabla). Si desea obtener más información, consulte el tema [Iconos de objeto de cognos](#) el p. 45.

Si el paquete tiene filtros definidos, puede importar uno o más de ellos. Si un filtro que ha importado está asociado con datos importados, ese filtro se aplica antes de importar los datos.

*Nota:* los datos que se van a importar deben tener el formato UTF-8.

Si importa un informe, seleccione un paquete o una carpeta de un paquete que contenga uno o más informes. A continuación, seleccione el informe individual que desee importar. *Nota:* solo se pueden importar los informes de lista individuales; las listas múltiples no son compatibles.

Si los parámetros se han definido, para un objeto de datos o un informe, puede especificar valores para estos parámetros antes de importar el objeto o el informe.

### ***Iconos de objeto de cognos***

Los diferentes tipos de objeto que puede importar de una base de datos de Cognos BI se representan por diferentes iconos, tal y como muestra la siguiente tabla.

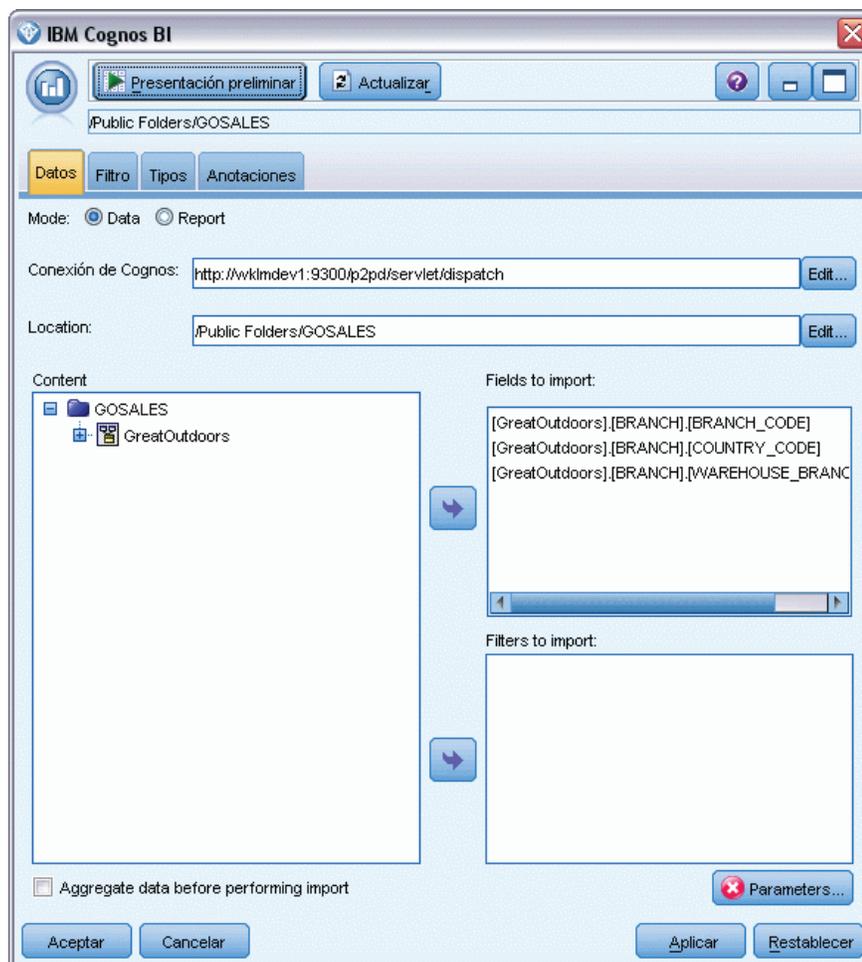
Tabla 2-3  
*Iconos de objeto de cognos*

<b>Icono</b>	<b>Objeto</b>
	Paquete
	Espacio de nombre
	Asunto de consulta
	Elemento de consulta
	Dimensión de medición
	Medida
	Dimensión
	Jerarquía de nivel
	Nivel
	Filtro
	Informe
	Cálculo autónomo

### ***Importación de datos de Cognos***

Para importar datos de una base de datos de IBM Cognos BI, asegúrese de que Modo está definido como Datos y complete el cuadro de diálogo de la siguiente forma.

Figura 2-20  
 Importación de datos de Cognos



**Conexión.** Pulse en el botón Editar para mostrar un cuadro de diálogo donde puede definir los detalles de una nueva conexión de Cognos desde la que importar datos o informes. Si ya ha iniciado sesión en un servidor de Cognos mediante IBM® SPSS® Modeler, también puede editar los detalles de la conexión actual. Si desea obtener más información, consulte el tema [Conexiones de Cognos](#) el p. 49.

**Ubicación.** Cuando haya establecido la conexión de servidor de Cognos, pulse en el botón Editar junto a este campo para mostrar una lista de paquetes disponibles desde los que importar contenido. Si desea obtener más información, consulte el tema [Selección de ubicación de Cognos](#) el p. 50.

**Contenido.** Muestra el nombre del paquete seleccionado, junto con los espacios de nombre asociados al paquete. Pulse dos veces en un espacio de nombre para mostrar los objetos que puede importar. Los diferentes tipos de objeto se muestran con diferentes iconos. Si desea obtener más información, consulte el tema [Iconos de objeto de cognos](#) el p. 45.

Para seleccionar un objeto para importar, seleccione el objeto y pulse la flecha superior de las dos que se encuentran a la derecha para mover el objeto al panel Campos que van a importarse. Al seleccionar un asunto de consulta se importan todos sus elementos de consulta. Si pulsa dos veces en un asunto de consulta, lo expandirá, de modo que podrá seleccionar uno o más de sus elementos de consulta individuales. Puede realizar varias selecciones con Ctrl-pulsar (seleccionar elementos individuales), Mayús-pulsar (seleccionar un bloque de elementos) y Ctrl-A (seleccionar todos los elementos).

Para seleccionar un filtro que se aplicará (si el paquete tiene filtros definidos), navegue hasta el filtro en el panel de contenido, seleccione el filtro y pulse la flecha inferior de las dos que se encuentran a la derecha para mover el filtro al panel Filtros que se aplicarán. Puede realizar múltiples selecciones dejando pulsada la tecla Ctrl (para seleccionar filtros individuales) y la tecla Mayús (para seleccionar un bloque de filtros).

**Campos que van a importarse.** Enumera los objetos de base de datos que ha seleccionado para importar a SPSS Modeler para su procesamiento. Si ya no necesita un objeto específico, selecciónelo y pulse en la flecha izquierda para devolverlo al panel Contenido. Puede realizar varias selecciones del mismo modo que para Contenido.

**Filtros que se aplicarán.** Enumera los filtros que haya seleccionado para aplicar a los datos antes de importarse. Si ya no necesita un filtro concreto específico, selecciónelo y pulse en la flecha izquierda para devolverlo al panel Contenido. Puede realizar varias selecciones del mismo modo que para Contenido.

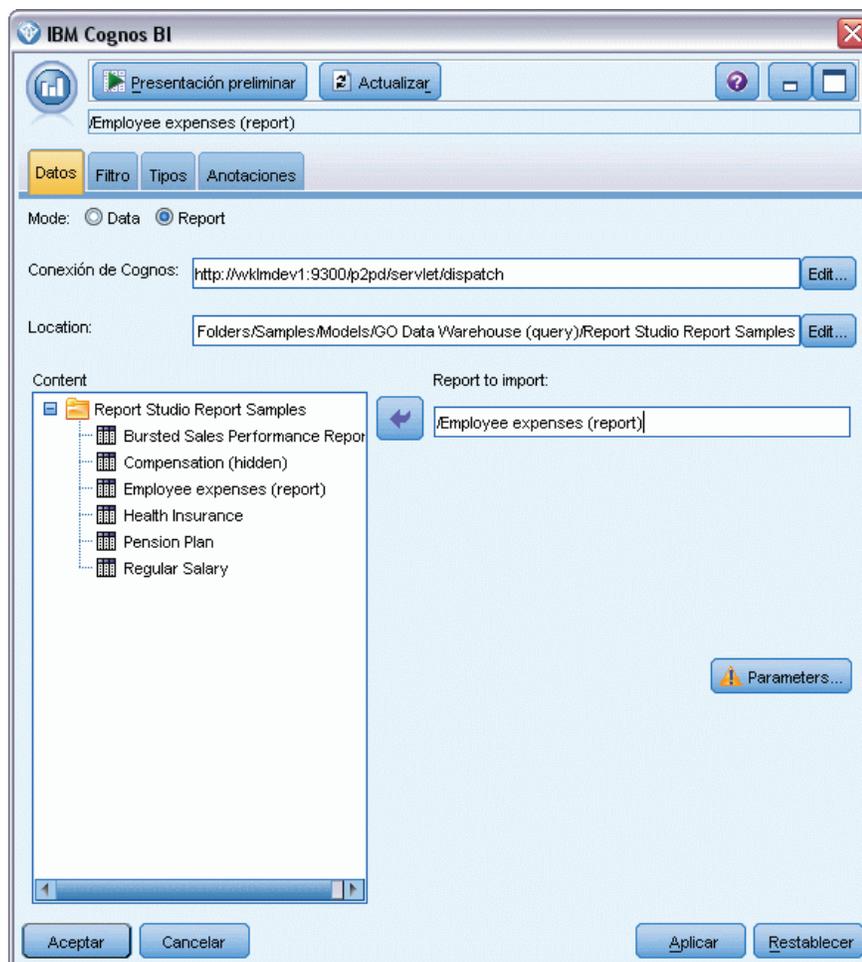
**Parámetros.** Si este botón está activado, el objeto seleccionado tiene parámetros definidos. Puede utilizar los parámetros para realizar ajustes (por ejemplo, realizar un cálculo parametrizado) antes de importar los datos. Si los parámetros se definen pero no se proporciona el parámetro por defecto, el botón muestra un triángulo de advertencia. Pulse el botón para mostrar los parámetros y, opcionalmente, editarlos. Si el botón está desactivado, el informe no tiene parámetros definidos.

**Agregar datos antes de importar.** Seleccione esta casilla si desea importar datos agregados en lugar de datos sin procesar.

## ***Importación de informes de Cognos***

Para importar un informe predefinido de una base de datos de IBM Cognos BI, asegúrese de que Modo está definido como Informe y complete el cuadro de diálogo de la siguiente forma. *Nota:* solo se pueden importar los informes de lista individuales; las listas múltiples no son compatibles.

Figura 2-21  
Importación de informes de Cognos



**Conexión.** Pulse en el botón Editar para mostrar un cuadro de diálogo donde puede definir los detalles de una nueva conexión de Cognos desde la que importar datos o informes. Si ya ha iniciado sesión en un servidor de Cognos mediante IBM® SPSS® Modeler, también puede editar los detalles de la conexión actual. Si desea obtener más información, consulte el tema [Conexiones de Cognos](#) el p. 49.

**Ubicación.** Cuando haya establecido la conexión de servidor de Cognos, pulse en el botón Editar junto a este campo para mostrar una lista de paquetes disponibles desde los que importar contenido. Si desea obtener más información, consulte el tema [Selección de ubicación de Cognos](#) el p. 50.

**Contenido.** Muestra el nombre del paquete o de la carpeta seleccionada que contiene los informes. Desplácese hasta un informe específico, selecciónelo y pulse la flecha derecha para traer el informe al campo Informe que se importará.

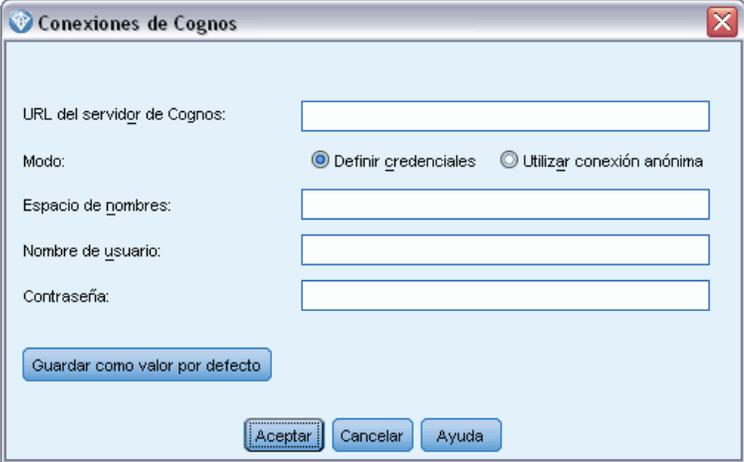
**Informe que se importará.** Indica el informe que ha seleccionado para importar a SPSS Modeler. Si ya no necesita el informe, selecciónelo y pulse en la flecha izquierda para devolverlo al panel Contenido o traer un informe diferente a este campo.

**Parámetros.** Si este botón está activado, el informe seleccionado tiene los parámetros definidos. Puede utilizar parámetros para realizar ajustes antes de importar el informe (por ejemplo, especificando una fecha de inicio y de fin para los datos del informe). Si los parámetros se definen pero no se proporciona el parámetro por defecto, el botón muestra un triángulo de advertencia. Pulse el botón para mostrar los parámetros y, opcionalmente, editarlos. Si el botón está desactivado, el informe no tiene parámetros definidos.

## Conexiones de Cognos

El cuadro de diálogo Conexiones de Cognos le permite seleccionar el servidor de Cognos BI desde el que importar o exportar objetos de base de datos.

Figura 2-22  
Selección de servidor Cognos



**URL del servidor Cognos.** Escriba la URL del servidor de Cognos BI desde la que realizará la importación o exportación. Este es el valor de la propiedad de entorno “URI de envío externo” de la configuración IBM Cognos del servidor de Cognos BI. Póngase en contacto con el administrador del sistema de Cognos si no está seguro de qué URL utilizar.

**Modo.** Seleccione Establecer credenciales si desea registrarse con un espacio de nombre, nombre de usuario y contraseña de Cognos específicos (por ejemplo, como administrador). Seleccione Utilizar conexión anónima para registrarse sin credenciales de usuario, en cuyo caso no cumplimentará los otros campos.

**Espacio de nombre.** Especifique el proveedor de autenticación de seguridad de Cognos utilizado para registrarse en el servidor. El proveedor de autenticación se utiliza para definir y mantener usuarios, grupos y papeles, así como para controlar el proceso de autenticación.

**Nombre de usuario.** Introduzca el nombre de usuario de Cognos con el que se inicia sesión en el servidor.

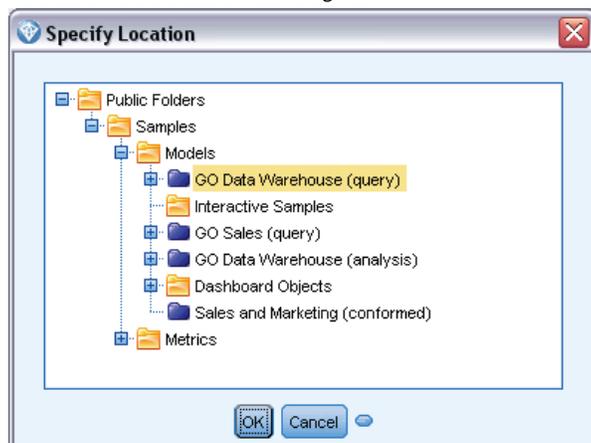
**Contraseña.** Introduzca la contraseña asociada al nombre de usuario especificado.

**Guardar como valor por defecto.** Pulse en este botón para almacenar esta configuración como por defecto, para no tener que volver a introducirla cada vez que abra el nodo.

## ***Selección de ubicación de Cognos***

El cuadro de diálogo Especificar ubicación permite seleccionar un paquete de Cognos desde el que importar datos o un paquete o carpeta desde el que importar informes.

Figura 2-23  
*Selección de ubicación de Cognos*



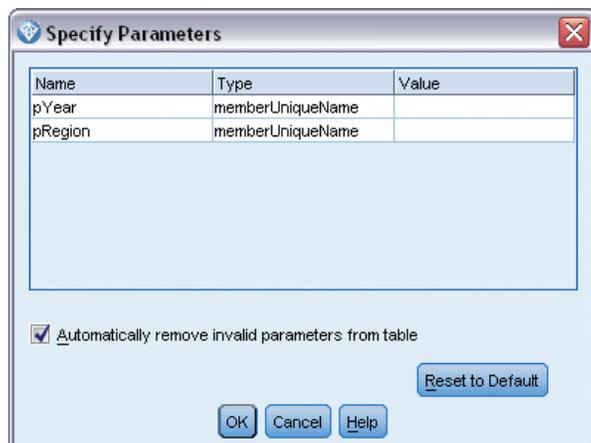
**Carpetas públicas.** Si importa datos, enumera los paquetes y carpetas disponibles desde el servidor seleccionado. Seleccione el paquete que quiera utilizar y pulse en Aceptar. Solamente puede seleccionar un paquete por cada nodo de fuente de Cognos BI.

Si importa informes, enumera los paquetes y carpetas que contienen los informes disponibles desde el servidor seleccionado. Seleccione un paquete o carpeta de informe y pulse en Aceptar. Solamente puede seleccionar un paquete o carpeta de informe por nodo de fuente de Cognos BI, aunque las carpetas de informe pueden contener otras carpetas de informes e informes individuales.

## ***Especificación de parámetros de datos o informes***

Si los parámetros se han definido en Cognos BI, para un objeto de datos o un informe, puede especificar valores para estos parámetros antes de importar el objeto o el informe. Un ejemplo de parámetros para un informe podría ser las fechas de inicio y fin de los contenidos del informe.

Figura 2-24  
Parámetros de Cognos



**Nombre.** El nombre del parámetro, tal y como se especifica en la base de datos de Cognos BI.

**Tipo.** Una descripción del parámetro.

**Valor.** El valor que se asignará al parámetro. Para introducir o editar un valor, pulse dos veces en su casilla de la tabla. Los valores no se validan aquí, por lo que cualquier valor no válido se detecta en el periodo de ejecución.

**Eliminar automáticamente de la tabla de parámetros no válidos.** Esta opción se selecciona de manera predeterminada y eliminará cualquier parámetro no válido encontrado dentro del objeto de datos o informe.

## Nodo de fuente SAS

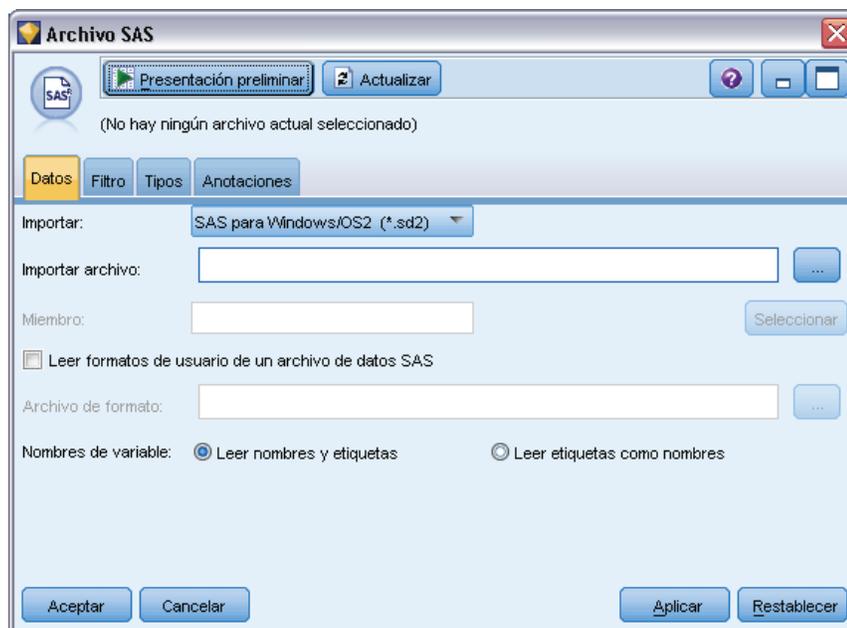
*Nota:* Esta función está disponible en SPSS Modeler Professional y SPSS Modeler Premium.

El nodo de fuente SAS le permite introducir los datos SAS en la sesión de minería de datos. Puede importar cuatro tipos de archivos:

- SAS para Windows/OS2 (.sd2)
- SAS para UNIX (.ssd)
- Archivo de transporte SAS (.tpt)
- SAS versión 7/8/9 (.sas7bdat)

Una vez se han importado los datos, se mantienen todas las variables y no se cambia ningún tipo de variable. Se han seleccionado todos los casos.

Figura 2-25  
Importación de un archivo SAS



### **Configuración de opciones para el nodo de fuente SAS**

**Importar.** Seleccione el tipo de archivo SAS que se va a transportar. Puede seleccionar SAS para Windows/OS2 (.sd2), SAS para UNIX (.SSD), Archivo de transporte SAS (.tpt) o SAS versión 7/8/9 (.sas7bdat).

**Importar archivo.** Especifique el nombre del archivo. Puede introducir un nombre de archivo o pulsar en el botón de puntos suspensivos (...) para buscar la ubicación del archivo.

**Miembro.** Seleccione un miembro para importar desde el archivo de transporte SAS seleccionado anteriormente. Puede introducir un nombre de archivo o pulsar en Seleccionar para buscar en todos los miembros del archivo.

**Leer formatos de usuario de un archivo de datos SAS.** Seleccione leer formatos de usuario. Los archivos SAS almacenan datos y formatos de datos (como etiquetas de variables) en archivos diferentes. La mayoría de las veces, deseará importar también los formatos. No obstante, si tiene un conjunto de datos de grandes dimensiones, es posible que desee anular la selección de esta opción para ahorrar memoria.

**Archivo de formato.** Si se necesita un archivo de formato, este cuadro de diálogo estará activado. Puede introducir un nombre de archivo o pulsar en el botón de puntos suspensivos (...) para buscar la ubicación del archivo.

**Nombres de variable.** Seleccione un método para gestionar etiquetas y nombres de variables en la importación desde un archivo SAS. Los metadatos seleccionados para incluirlos aquí permanecen durante el trabajo en IBM® SPSS® Modeler y se pueden volver a exportar para su uso en SAS.

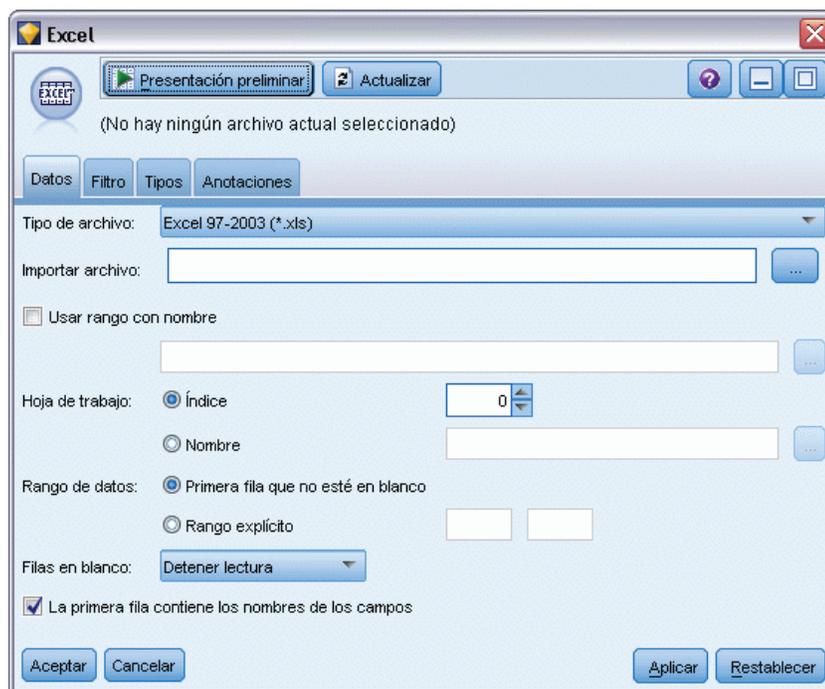
- **Leer nombres y etiquetas.** Seleccione para leer las etiquetas y nombres de variables en SPSS Modeler. Esta opción está seleccionada por defecto y los nombres de variables se muestran en el nodo Tipo. Las etiquetas se pueden mostrar en el generador de expresiones, gráficos, exploradores de modelos y otros tipos de resultados, dependiendo de las opciones especificadas en el cuadro de diálogo de propiedades de la ruta.
- **Leer etiquetas como nombres.** Seleccione para leer las etiquetas descriptivas de variables desde el archivo SAS en lugar de los nombres abreviados de campos y utilice estas etiquetas como nombres de variables en SPSS Modeler.

## Nodo de fuente de Excel

El nodo de fuente de Excel permite importar datos de cualquier versión de Microsoft Excel.

Figura 2-26

Nodo de fuente de Excel



**Tipo de archivo.** Seleccione el tipo de archivo de Excel que está importando.

**Importar archivo.** Especifica el nombre y la ubicación del archivo de hoja de cálculo que se va a importar.

**Usar rango con nombre.** Permite especificar un rango con nombre de casillas tal y como se define en la hoja de cálculo de Excel. Pulse en el botón de puntos suspensivos (...) para seleccionar de la lista de rangos disponibles. Si se utiliza un rango con nombre, no se podrá aplicar otra configuración de rango de datos y hoja de trabajo y, en consecuencia, estará desactivada.

**Seleccionar hoja de trabajo.** Especifique la hoja de trabajo que se va a importar, por índice o por nombre.

- **Por índice.** Especifique el valor de índice para la hoja de trabajo que desea importar, comenzando por 0 para la primera hoja de trabajo, 1 para la segunda hoja de trabajo y así sucesivamente.
- **Por nombre.** Especifique el nombre de la hoja de trabajo que desea importar. Pulse en el botón de puntos suspensivos (...) para seleccionar de la lista de hojas de trabajo disponibles.

**Intervalo en hoja de trabajo.** Puede importar datos que comiencen por la primera fila que no esté en blanco o con un rango de casillas explícito.

- **El intervalo comienza en la primera fila no en blanco.** Busca la primera casilla que no esté en blanco y la utiliza como la esquina superior izquierda del rango de datos.
- **Rango explícito de casillas.** Le permite especificar un intervalo explícito por fila y columna. Por ejemplo, para especificar el intervalo de Excel A1:D5, puede introducir A1 en el primer campo y D5 en el segundo (o, de forma alternativa, R1C1 y R5C4). Se devolverán todas las filas del rango especificado, incluidas las filas en blanco.

**En filas en blanco.** Si se encuentra más de una fila en blanco, puede seleccionar si desea Parar de leer o seleccionar Devolver filas en blanco para continuar leyendo todos los datos hasta el final de la hoja de trabajo, incluidas las filas en blanco.

**Primera fila tiene nombres de columnas.** Indica que la primera fila del rango concreto debería usarse como nombres de campo (columna). Si no se selecciona, los nombres de campo se generan automáticamente.

### **Almacenamiento de campo y Nivel de medición**

Al leer valores de Excel, los campos con almacenamiento numérico se leen con un nivel de medición de *Continuo* por defecto y los campos de cadena se leen como *Nominales*. Puede cambiar el nivel de medición de forma manual (continuo en lugar de nominal) en la pestaña Tipo, pero el almacenamiento viene determinado de forma automática (aunque puede cambiarse mediante la función de conversión, como `to_integer`, en un nodo Rellenar o un nodo Derivar si es necesario). Si desea obtener más información, consulte el tema [Configuración de almacenamiento de campos y formato](#) el p. 32.

Por defecto, los campos con una mezcla de valores numéricos y de cadena se leen como números, lo que significa que cualquier valor de cadena se establecerá en valores nulos (valores perdidos del sistema) en IBM® SPSS® Modeler. Esto ocurre porque, a diferencia de Excel, SPSS Modeler no permite tipos de almacenamiento mezclado dentro de un campo. Para evitarlo, puede establecer de forma manual el formato de casillas a Texto en la hoja de cálculo de Excel, lo que hará que todos los valores (incluidos los números) se leen como cadenas.

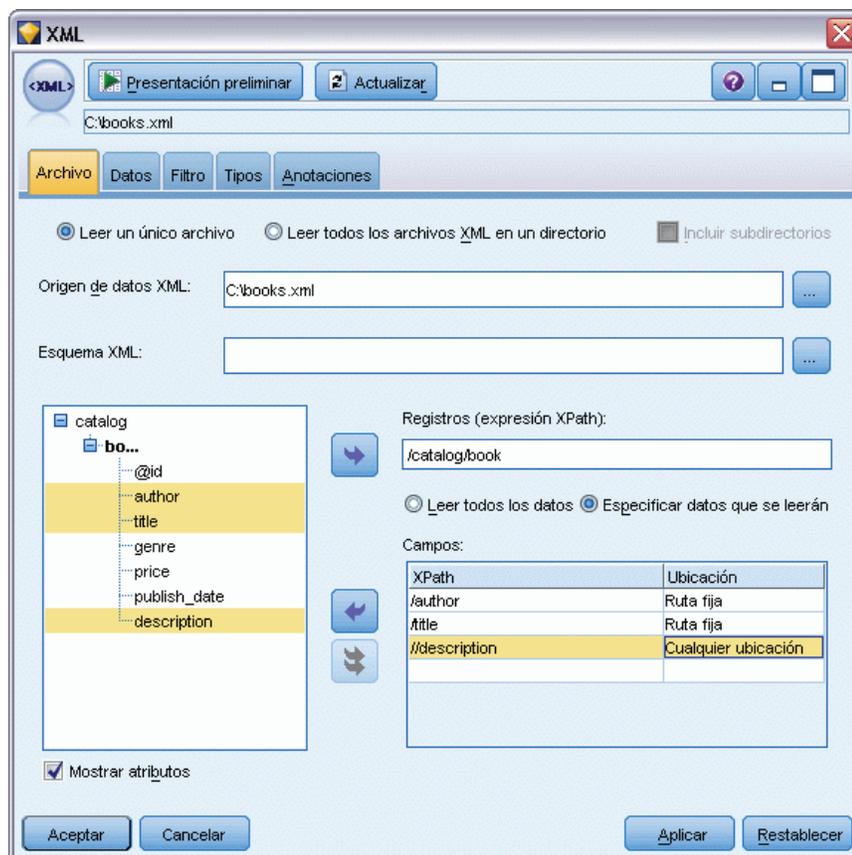
## **Nodo Fuente XML**

*Nota:* Esta función está disponible en SPSS Modeler Professional y SPSS Modeler Premium.

El nodo Fuente XML le permite importar datos de un archivo en formato XML a una ruta de IBM® SPSS® Modeler. XML es un lenguaje estándar de intercambio de datos y, para la mayoría de organizaciones es el formato que eligen para ello. Por ejemplo, es posible que una agencia tributaria quiera analizar datos de devoluciones de impuestos que se han enviado en línea que tienen datos en formato XML.

La importación de datos XML a una ruta de SPSS Modeler le permite realizar una amplia gama de funciones de análisis predictivos en la fuente. Los datos XML se analizan en formato tabular en el que las columnas se corresponden un los diferentes niveles de anidación de los elementos y atributos XML. Los elementos XML se muestran en formato de XPath (consulte <http://www.w3.org/TR/xpath20/>).

Figura 2-27  
Importación de datos XML



**Lectura de un archivo único.** Por defecto, SPSS Modeler lee un único archivo, que especifique en el campo Fuente de datos XML.

**Lectura de todos los archivos XML en un directorio.** Seleccione esta opción si desea leer todos los archivos XML de un directorio concreto. Especifique la ubicación en el campo Directorio que aparece. Seleccione la casilla de verificación Incluir subdirectorios para leer más archivos XML de todos los subdirectorios del directorio especificado.

**Fuente de datos XML.** Introduzca la ruta y el nombre completo del archivo de fuente XML que dese importar o utilice el botón Examinar para buscar el archivo.

**Esquema XML.** (Opcional) Especifique la ruta completa y el nombre del archivo de un archivo XSD o DTD desde el que se leerá la estructura XML o utilice el botón Examinar para buscar el archivo. Si deja este campo en blanco, se leerá la estructura desde el archivo de fuente XML.

Un archivo XSD o DTD puede tener más de un elemento raíz. En este caso, cuando seleccione un campo diferente, aparecerá un cuadro de diálogo en el que podrá seleccionar el elemento raíz que desee utilizar. Si desea obtener más información, consulte el tema [Selección entre múltiples elementos raíz](#) el p. 56.

**Estructura XML.** Un árbol de jerarquía que muestra la estructura del archivo de fuente XML (o el esquema, si lo ha especificado en el campo Esquema XML). Para definir un límite del registro, seleccione un elemento y pulse en el botón de flecha derecha para copiar el elemento en el campo Registros.

**Mostrar atributos.** Muestra u oculta los atributos de los elementos XML en el campo estructura XML.

**Registros (expresión XPath).** Muestra la sintaxis XPath de un elemento copiado desde el campo de estructura XML. Este elemento se resalta en la estructura XML y define los límites del registro. Cada vez que este elemento se encuentra en el archivo de fuente se crea un nuevo registro. Si el campo está vacío, el primer elemento filial de la raíz se utiliza como el límite del registro.

**Leer todos los datos.** Por defecto, todos los datos del archivo de fuente se leen en la ruta.

**Especificar los datos que se leerán.** Seleccione esta opción si desea importar elementos individuales, atributos o ambos. Si selecciona esta opción activará la pestaña Campos donde puede especificar los datos que desea importar.

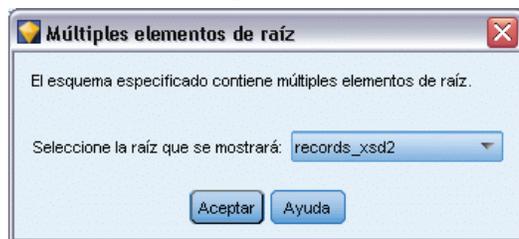
**Campos.** Esta tabla contienen los elementos y atributos seleccionados para importar, si ha seleccionado la opción Especificar los datos que se leerán. Puede introducir la sintaxis de XPath o un elemento o atributo directamente en la columna XPath o seleccionar un elemento o atributo en la estructura XML y pulsar en el botón de flecha derecha para copiar el elemento en la tabla. Para copiar todos los elementos filiales y atributos de un elemento, seleccione el elemento en la estructura XML y pulse en el botón de doble flecha a la derecha.

- **XPath.** La sintaxis XPath de los elementos que se importarán.
- **Ubicación.** La ubicación de la estructura XML de los elementos que se importarán. Ruta fijada muestra la ruta del elemento relativo al elemento resaltado en la estructura XML (o el primer elemento filial en la raíz, si no se ha resaltado ningún elemento). Cualquier ubicación denota un elemento del nombre asignado a cualquier ubicación de la estructura XML. Personalizado aparece si introduce una ubicación directamente en la columna XPath.

## ***Selección entre múltiples elementos raíz***

Mientras un archivo XML formado adecuadamente solamente puede tener un único elemento raíz, un archivo XSD o DTD puede contener múltiples raíces. Si una de las raíces coincide con la del archivo de fuente XML, se utilizará ese elemento raíz, mientras que si no es así deberá seleccionar uno para utilizarlo.

Figura 2-28  
Selección entre múltiples elementos raíz



**Seleccione la raíz que desea mostrar.** Seleccione el elemento raíz que desea utilizar. El predeterminado es el primer elemento raíz en la estructura XSD o DTD.

### ***Eliminación de espacios no deseados de los datos de fuente XML***

Es posible utilizar la combinación de caracteres [CR][LF] para implementar saltos de línea en los datos de la fuente XML. En algunos casos estos saltos de línea pueden aparecer en medio de una cadena de texto, por ejemplo:

```
<descripción> Una descripción detallada de la creación de aplicaciones[CR][LF]  
con XML.</descripción>
```

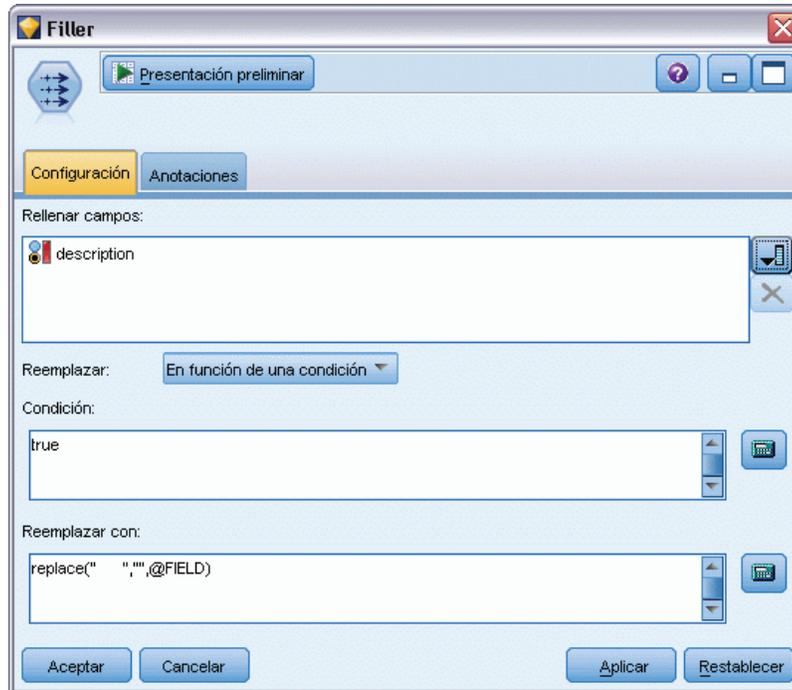
Estos saltos de línea pueden no ser visibles cuando se abre el archivo en algunas aplicaciones, por ejemplo un navegador web. Sin embargo, cuando se leen los datos en la ruta a través del nodo de fuente XML, los saltos de línea se convierten en una serie de caracteres de espacio, por ejemplo:

Figura 2-29  
Registro XML con saltos de línea que aparecen como espacios



Puede corregir esto usando un nodo Rellenar para quitar los espacios no deseados:

Figura 2-30  
Nodo Rellenar con los ajustes para eliminar espacios

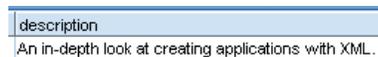


Aquí hay un ejemplo de cómo puede conseguirlo:

- ▶ Conecte un nodo Rellenar al nodo Fuente XML.
- ▶ Abra el nodo Rellenar y utilice el selector de campos para seleccionar el campo con espacios no deseados.
- ▶ Defina Reemplazar como En función de una condición y defina Condición como true.
- ▶ En el campo Reemplazar con, introduzca `replace(" ", "", @FIELD)` y pulse Aceptar.
- ▶ Conecte un nodo Tabla al nodo Rellenar y ejecute la ruta.

En los resultados del nodo Tabla, el texto aparecerá como sigue:

Figura 2-31  
Registro XML con los espacios no deseados eliminados



## Nodo Datos Usuario

El nodo Datos Usuario le facilita la creación de los datos sintéticos, bien partiendo desde cero o modificando los datos existentes. Esto resulta útil, por ejemplo, cuando desee crear un conjunto de datos de comprobación para el modelado.

### Creación de datos partiendo desde cero

El nodo Datos Usuario está disponible en la paleta Fuentes y se puede añadir directamente al lienzo de rutas.

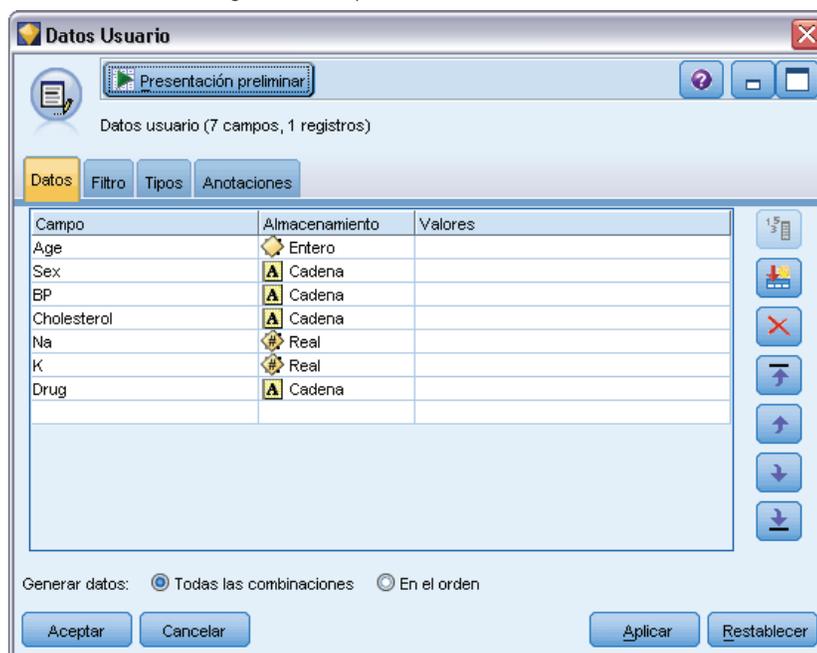
- ▶ Pulse en la pestaña Fuentes de la paleta de nodos.
- ▶ Arrastre y suelte o pulse dos veces para añadir el nodo Datos Usuario al lienzo de rutas.
- ▶ Pulse dos veces para abrir el cuadro de diálogo y especifique los campos y valores.

*Nota:* los nodos Datos Usuario que se seleccionan desde la paleta Fuentes estarán completamente en blanco, sin campos ni información de datos. Lo que le permite crear datos sintéticos partiendo totalmente desde cero.

### Generación de datos a partir de una fuente de datos existente

Figura 2-32

Nodo Datos Usuario generado a partir de un nodo de ruta



También puede generar un nodo Datos Usuario a partir de un nodo no terminal de la ruta:

- ▶ Indique en qué punto de la ruta desea reemplazar el nodo.
- ▶ Pulse con el botón derecho en el nodo que introducirá sus datos en el nodo Datos Usuario y seleccione Generar nodo de datos de usuario en el menú.
- ▶ El nodo Datos usuario aparece con todos los procesos que se encuentran debajo adjuntos, sustituyendo en este punto el nodo existente en la ruta de datos. Una vez generado, el nodo hereda toda la estructura de datos y la información sobre el tipo de campo (si está disponible) de los metadatos.

*Nota:* si los datos no se han ejecutado a través de todos los nodos de la ruta, los nodos no estarán completamente instanciados, lo que significa que puede que los valores de datos y almacenamiento no estén disponibles cuando se realice la sustitución por un nodo Datos Usuario.

### ***Configuración de opciones para el nodo Datos Usuario***

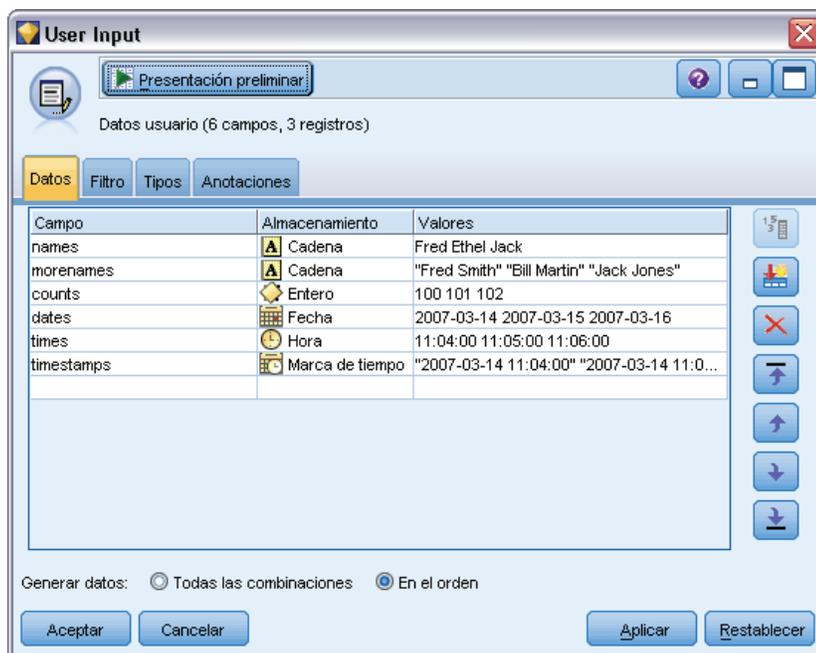
El cuadro de diálogo para el nodo Datos Usuario contiene varias herramientas que puede utilizar para introducir valores y definir la estructura de datos para los datos sintéticos. Para un nodo generado, la tabla de la pestaña Datos contiene nombres de campos de la fuente de datos original. Para un nodo añadido desde la paleta Fuentes, la tabla está vacía. Si utiliza las opciones de la tabla, podrá realizar las siguientes tareas:

- Añadir nuevos campos mediante el botón Añadir un nuevo campo, situado en la parte derecha de la tabla.
- Cambiar el nombre de los campos existentes.
- Especificar el almacenamiento de datos para cada campo.
- Especificar valores.
- Cambiar el orden de los campos en la visualización.

#### ***Introducción de datos***

Puede especificar valores o insertarlos desde el conjunto de datos original para cada campo, mediante el botón Selector de valores, situado en la parte derecha de la tabla. Consulte las reglas descritas a continuación, si desea obtener más información acerca de la especificación de valores. También puede dejar el campo vacío; los campos vacíos se rellenan con el espacio nulo del sistema (\$null\$).

Figura 2-33  
Especificación del tipo de almacenamiento para campos en un nodo Datos Usuario generado



Para especificar valores de cadena, simplemente introdúzcalos en la columna de valores separados por espacios:

Fred Ethel Martin

Cadenas que incluyen espacios que se pueden escribir entre comillas dobles:

"Bill Smith" "Fred Martin" "Jack Jones"

Para los campos numéricos, puede introducir varios valores de la misma forma (con espacios entre ellos):

10 12 14 16 18 20

O puede especificar la misma serie de números estableciendo los límites (10, 20) y los pasos entre ellos (2). Con este método, se escribirá lo siguiente:

10,20,2

Estos dos métodos se pueden combinar insertando uno dentro de otro, como se indica a continuación:

1 5 7 10,20,2 21 23

Esta entrada generará los siguientes valores:

1 5 7 10 12 14 16 18 20 21 23

Los valores de fecha y hora se pueden introducir mediante el formato por defecto actual seleccionado en el cuadro de diálogo Propiedades de ruta, por ejemplo:

11:04:00 11:05:00 11:06:00

2007-03-14 2007-03-15 2007-03-16

Para los valores de marca de tiempo, que tienen un componente de fecha y hora, se deben utilizar comillas dobles:

"2007-03-14 11:04:00" "2007-03-14 11:05:00" "2007-03-14 11:06:00"

Para obtener información adicional consulte los comentarios sobre el almacenamiento de datos a continuación.

**Generar datos.** Permite especificar cómo se generan los registros al ejecutar la ruta.

- **Todas las combinaciones.** Genera registros que contienen todas las combinaciones posibles de los valores de campo, por lo tanto, cada valor de campo aparecerá en varios registros. En ocasiones puede generar más datos de los deseados, por lo que es recomendable seguir este nodo con un nodo de muestra.
- **En el orden.** Genera registros en el orden en que se han especificado los valores de campo. Cada valor de campo aparece solamente en un registro. El número de registros será igual al número más grande de valores de un único campo. Cuando los campos tienen un número menor, se insertan valores no definidos (\$null\$).

Por ejemplo, las entradas siguientes generarán los registros mostrados en las siguientes tablas.

- **Edad.** 30,60,10
- **PS.** BAJA
- **Colesterol.** NORMAL ALTA
- **Medicamento.** (se deja vacío)

Generar datos está definido como Todas las combinaciones:

Edad	PS	Colesterol	Medicamento
30	BAJA	NORMAL	\$null\$
30	BAJA	SUPERIOR	\$null\$
40	BAJA	NORMAL	\$null\$
40	BAJA	SUPERIOR	\$null\$
50	BAJA	NORMAL	\$null\$
50	BAJA	SUPERIOR	\$null\$
60	BAJA	NORMAL	\$null\$
60	BAJA	SUPERIOR	\$null\$

Generar datos está definido como En el orden:

Edad	PS	Colesterol	Medicamento
30	BAJA	NORMAL	\$null\$
40	\$null\$	SUPERIOR	\$null\$

Edad	PS	Colesterol	Medicamento
50	\$null\$	\$null\$	\$null\$
60	\$null\$	\$null\$	\$null\$

### Almacenamiento de datos

Almacenamiento describe la forma en la que los datos se almacenan en un campo. Por ejemplo, un campo con valores 1 y 0 almacena datos enteros. Esto es diferente al nivel de medición, que describe el uso de los datos y no afecta al almacenamiento. Por ejemplo, puede que desee definir el nivel de medición para un campo de enteros con valores de 1 y 0 para una *Marca*. Esto suele indicar que 1=*Verdadero* y 0=*Falso*. Aunque el almacenamiento se debe determinar en el origen, el nivel de medición se puede cambiar utilizando un nodo Tipo en cualquier punto de la ruta. Si desea obtener más información, consulte el tema [Niveles de medida](#) en el capítulo 4 el p. 137.

Los tipos de almacenamiento disponibles son los siguientes:

- **Cadena.** Se utiliza para campos que contienen datos no numéricos, también llamados datos alfanuméricos. Una cadena puede incluir cualquier secuencia de caracteres, como *fred*, *Clase 2* o *1234*. Tenga en cuenta que los números de las cadenas no se pueden utilizar en los cálculos.
- **Entero** Un campo cuyos valores son números enteros.
- **Real.** Los valores son números que pueden incluir decimales (no está limitado a números enteros). El formato de presentación se especifica en el cuadro de diálogo Propiedades de ruta y puede quedar anulado para campos individuales en un nodo Tipo (ficha Formato).
- **Fecha.** Valores de fecha especificados en un formato estándar como año, mes y día (por ejemplo, 26.09.07). El formato determinado se especifica en el cuadro de diálogo Propiedades de ruta.
- **Hora.** Tiempo medido como una duración. Por ejemplo, una llamada de servicio que dure 1 hora, 26 minutos y 38 segundos se puede representar como 01:26:38, según el formato de hora actual que se especifica en el cuadro de diálogo Propiedades de ruta.
- **Marca de tiempo.** Valores que incluyen un componente de fecha y hora, por ejemplo 2007-09-26 09:04:00, siguen dependiendo de los formatos actuales de fecha y hora del cuadro de diálogo Propiedades de ruta. Tenga en cuenta que es posible que los valores de marca de tiempo se tengan que escribir entre comillas dobles para garantizar que se interpretan como un valor único en lugar de valores de fecha y hora diferentes. (Es aplicable, por ejemplo cuando introduce valores en un nodo Datos Usuario.)

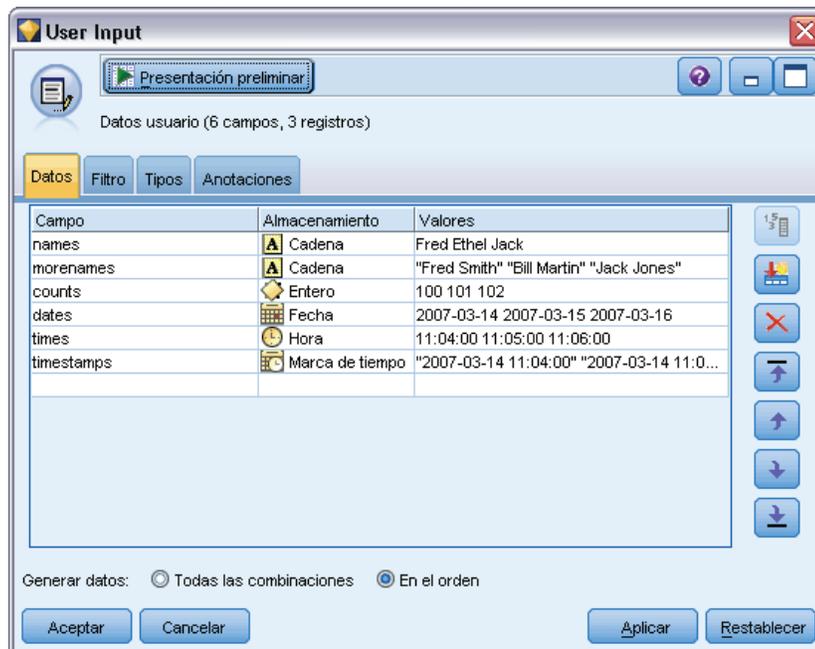
**Conversiones del almacenamiento.** Puede convertir el almacenamiento de un campo mediante una serie de funciones de conversión, como `to_string` y `to_integer`, en un nodo Rellenar. Si desea obtener más información, consulte el tema [Conversión del almacenamiento mediante el nodo Rellenar](#) en el capítulo 4 el p. 179. Tenga en cuenta que las funciones de conversión (y muchas otras funciones que requieren un tipo específico de entrada como un valor de fecha y hora) dependen de los formatos actuales especificados en el cuadro de diálogo Propiedades de ruta. Por ejemplo, si desea convertir un campo de cadena con los valores *Ene 2003*, *Feb 2003*, (etc.) para almacenar datos, seleccione MES AAAA como formato de fecha por defecto de la ruta. Las funciones de conversión también están disponibles en el nodo Derivar para realizar la conversión temporal durante un cálculo de derivación. También puede utilizar el nodo Derivar para realizar otras manipulaciones, como volver a codificar los campos con valores categóricos.

Si desea obtener más información, consulte el tema [Nueva codificación de valores con el nodo Derivar](#) en el capítulo 4 el p. 177.

**Lectura de datos mezclados.** Tenga en cuenta que al leer los campos con almacenamiento numérico (entero, real, hora, marca de tiempo o fecha), cualquier valor no numérico se establece como nulo o perdido por el sistema. Esto ocurre porque, a diferencia de otras aplicaciones, IBM® SPSS® Modeler no permite la combinación de tipos de almacenamiento dentro de un campo. Para evitarlo, cualquier campo con datos mezclados debe leerse como cadenas, ya sea cambiando el tipo de almacenamiento en el nodo de origen o en la aplicación externa, según sea necesario.

*Nota:* los nodos Datos Usuario generados pueden contener ya información de almacenamiento, recogida del nodo de fuente si está instanciado. Un nodo que no está instanciado no contiene información acerca del tipo de uso ni del almacenamiento.

Figura 2-34  
Especificación del tipo de almacenamiento para campos en un nodo Datos Usuario generado



### **Reglas para la especificación de valores**

Para los campos simbólicos, debe dejar espacios entre varios valores, como:

ALTA NORMAL BAJA

Para los campos numéricos, puede introducir varios valores de la misma forma (con espacios entre ellos):

10 12 14 16 18 20

O puede especificar la misma serie de números estableciendo los límites (10, 20) y los pasos entre ellos (2). Con este método, se escribirá lo siguiente:

10,20,2

Estos dos métodos se pueden combinar insertando uno dentro de otro, como se indica a continuación:

1 5 7 10,20,2 21 23

Esta entrada generará los siguientes valores:

1 5 7 10 12 14 16 18 20 21 23

## ***Pestañas de nodos de fuente comunes***

Las siguientes opciones se pueden especificar para todos los nodos de fuente pulsando en la pestaña correspondiente:

- **Pestaña Datos.** Se utiliza para cambiar el tipo de almacenamiento por defecto.
- **Pestaña Filtro.** Se utiliza para eliminar los campos de datos o cambiarles el nombre. Esta pestaña ofrece la misma funcionalidad que el nodo Filtro. Si desea obtener más información, consulte el tema [Opciones de configuración de filtrado](#) en el capítulo 4 el p. 155.
- **Pestaña Tipos.** Se utiliza para definir niveles de medición. Esta pestaña ofrece la misma funcionalidad que el nodo Tipo.
- **Pestaña Anotaciones.** Se utiliza para todos los nodos, esta pestaña proporciona opciones para cambiar el nombre de los nodos, ofrecer herramientas personalizadas y almacenar una anotación más extensa.

## ***Definición de niveles de medición en el nodo Fuente***

Las propiedades de campo se pueden especificar en un nodo de origen o en un nodo Tipo independiente. La funcionalidad es similar en ambos nodos. Se trata de las siguientes propiedades:

- **Campo.** Pulse dos veces en cualquier nombre de campo para especificar etiquetas de valor y de campo para los datos de IBM® SPSS® Modeler. Por ejemplo, los metadatos de campo importados de IBM® SPSS® Statistics se pueden ver o modificar aquí. Del mismo modo, puede crear nuevas etiquetas para campos y sus valores. Las etiquetas que se especifican aquí se muestran en SPSS Modeler dependiendo de las selecciones realizadas en el cuadro de diálogo de propiedades de la ruta.
- **Medida.** Este es el nivel de medición que se usa para describir características de los datos en un campo determinado. Si se conocen todos los detalles de un campo, éste se denomina **completamente instanciado**. Si desea obtener más información, consulte el tema [Niveles de medida](#) en el capítulo 4 el p. 137.

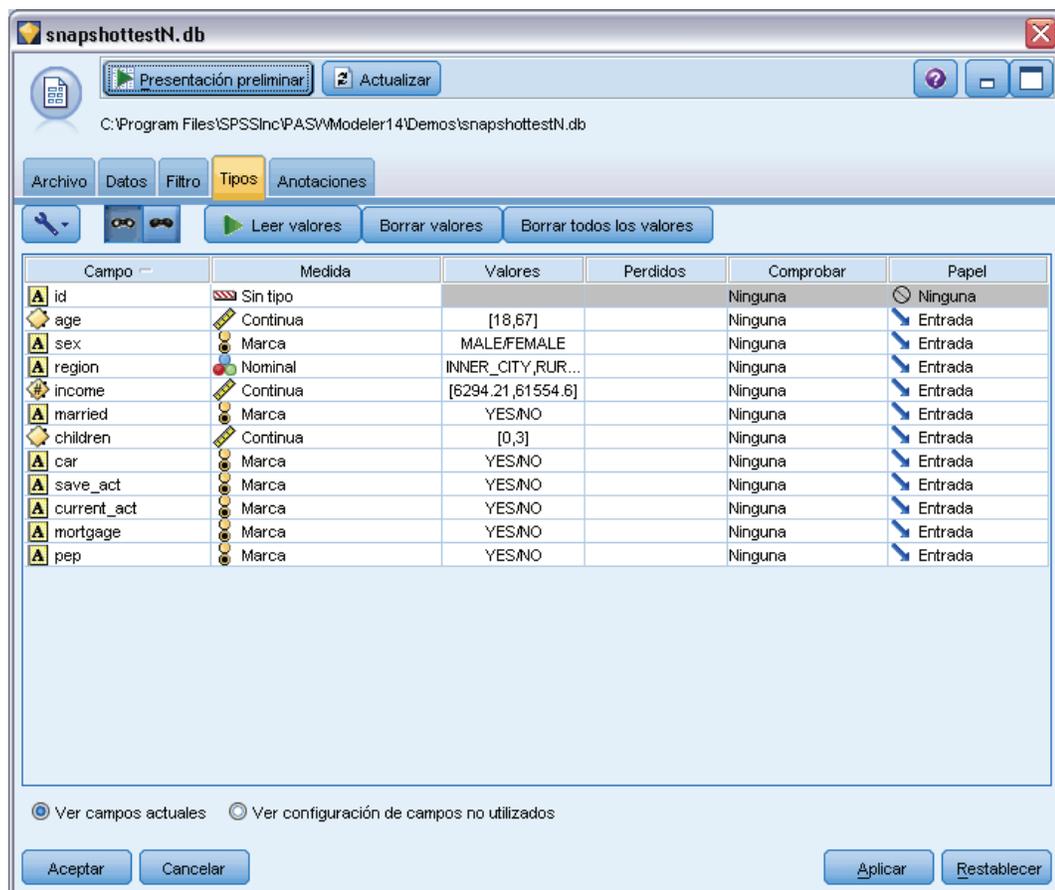
*Nota:* El nivel de medición de un campo difiere del tipo de almacenamiento de un campo, lo cual indica si los datos están almacenados como cadenas, números enteros, números reales, fechas, horas o marcas de tiempo.

- **Valores.** Esta columna le permite especificar las opciones para leer los valores de datos desde el conjunto de datos, o utilizar la opción Especificar para especificar los niveles de medición y los valores en un cuadro de diálogo separado. También puede optar por pasar los campos sin leer sus valores. Si desea obtener más información, consulte el tema [Valores de campos](#) en el capítulo 4 el p. 142.

- **Perdidos.** Se utiliza para especificar cómo se tratan los valores perdidos de un campo. Si desea obtener más información, consulte el tema [Definición de valores perdidos](#) en el capítulo 4 el p. 147.
- **Comprobar.** En esta columna puede configurar opciones que garanticen que los valores de campo se ajusten a los valores o rangos especificados. Si desea obtener más información, consulte el tema [Comprobación de valores de tipo](#) en el capítulo 4 el p. 148.
- **Papel.** Se usa para indicar a los nodos de modelado si los campos serán Entrada (campos predictores) o Objetivo (campos pronosticados) para un proceso de aprendizaje automático. Ambas y Ninguna también son papeles disponibles, junto con Partición, que indica un campo utilizado para crear particiones de registros en muestras independientes para entrenamiento, comprobación y validación. El valor Split especifica que se generarán modelos diferentes para cada valor posible del campo. Si desea obtener más información, consulte el tema [Definición del papel de campos](#) en el capítulo 4 el p. 149.

Si desea obtener más información, consulte el tema [Nodo Tipo](#) en el capítulo 4 el p. 135.

Figura 2-35  
Opciones de la pestaña Tipos



### ***Cuándo crear una instancia en el nodo de fuente***

Puede conocer dos formas acerca del almacenamiento de datos y los valores de los campos. Esta **creación de instancia** se puede producir en el nodo de fuente, cuando introduzca primero los datos en IBM® SPSS® Modeler, o insertando el nodo Tipo en la ruta de datos.

La creación de una instancia en el nodo de fuente resulta útil cuando:

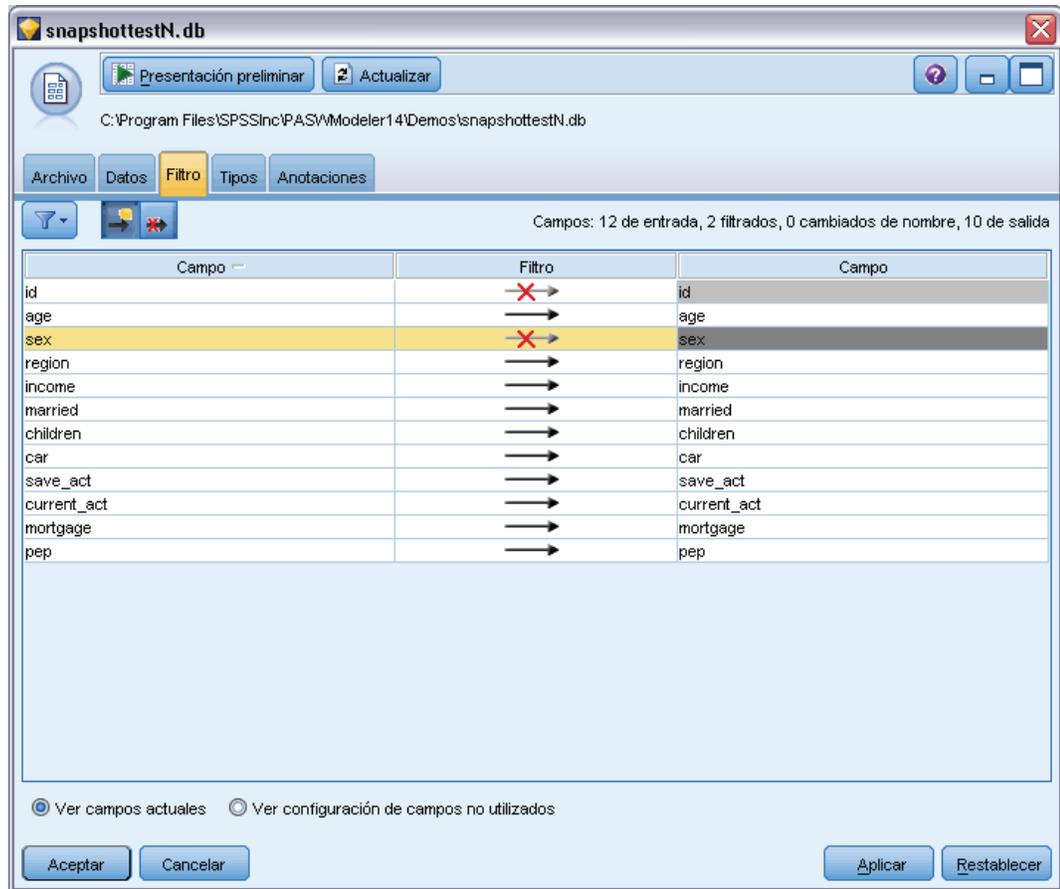
- El conjunto de datos es pequeño.
- Tiene previsto derivar nuevos campos mediante el generador de expresiones (la creación de una instancia hace que los valores de campos estén disponibles desde el generador de expresiones).

Normalmente, si el conjunto de datos no es muy grande y no tiene intención de añadir más adelante ningún campo a la ruta, la creación de una instancia en el nodo de fuente es el método más conveniente.

### ***Filtrado de campos desde el nodo de fuente***

La pestaña Filtro en el cuadro de diálogo del nodo de fuente le permitirá excluir los campos de las operaciones que se encuentren debajo, según el examen inicial de los datos. Esto resulta útil, por ejemplo, cuando hay campos duplicados en los datos o si ya conoce los datos como para excluir los campos irrelevantes. Si lo prefiere, puede añadir un nodo Filtro independiente en la ruta. La funcionalidad es similar en ambos casos. Si desea obtener más información, consulte el tema [Opciones de configuración de filtrado](#) en el capítulo 4 el p. 155.

Figura 2-36  
Filtrado de campos desde el nodo de fuente



# Nodos de operaciones con registros

## Conceptos básicos de las operaciones con registros

Los nodos de operaciones con registros se usan para realizar cambios en los datos a nivel de registro. Estas operaciones son importantes durante las fases **Comprensión de los datos** y **Preparación de los datos** del análisis porque permiten adaptar los datos a las necesidades particulares de su negocio.

Por ejemplo, según los resultados de la auditoría de datos realizada con el nodo Auditar datos (paleta Resultado), puede que desee fusionar los registros de las compras realizadas por los clientes durante los últimos tres meses. Con el nodo Fundir, puede fusionar registros basándose en los valores de un campo clave, como el *ID de cliente*. Así mismo, puede descubrir que es imposible administrar una base de datos con información sobre visitas al sitio Web con más de un millón de registros. Con el nodo Muestrear, puede seleccionar un subconjunto de datos para utilizarlo en el modelado.

La paleta Operaciones con registros contiene los siguientes nodos:



El nodo Seleccionar selecciona o descarta un subconjunto de registros de la ruta de datos en función de una condición específica. Por ejemplo, podría seleccionar los registros que pertenezcan a una región de ventas determinada. Si desea obtener más información, consulte el tema [Nodo Seleccionar](#) el p. 70.



El nodo Muestrear selecciona un subconjunto de registros. Se admite una variedad de tipos de muestras, entre las que se incluyen las muestras estratificadas, conglomeradas y no aleatorias (estructuradas). El muestreo puede ser de gran utilidad para mejorar el rendimiento y para seleccionar grupos de registros o transacciones relacionadas para un análisis. Si desea obtener más información, consulte el tema [Nodo Muestrear](#) el p. 72.



El nodo Equilibrar corrige los desequilibrios de un conjunto de datos para que cumpla una condición determinada. La directiva de equilibrado ajusta la proporción de registros si una condición es verdadera por el factor determinado. Si desea obtener más información, consulte el tema [Nodo Equilibrar](#) el p. 80.



El nodo Agregar reemplaza una secuencia de registros de entrada con registros de salida agregados y resumidos. Si desea obtener más información, consulte el tema [Nodo Agregar](#) el p. 82.



El nodo Adición de RFM (actualidad, frecuencia, monetario) permite tomar datos de transacciones históricas de clientes, deshacerse de los datos no utilizados y combinar todos los datos de transacciones restantes en una única fila que indica cuándo hizo negociaciones con los clientes por última vez, cuántas transacciones hicieron y el valor monetario total de dichas transacciones. Si desea obtener más información, consulte el tema [Nodo Adición de RFM](#) el p. 85.



Los nodos Ordenar organizan registros en orden ascendente o descendente atendiendo a los valores de uno o varios campos. Si desea obtener más información, consulte el tema [Nodo Ordenar](#) el p. 87.



El nodo Fundir toma varios registros de entrada y crea un registro de salida único que contiene todos o algunos de los campos de entrada. Es útil para fusionar datos desde diferentes orígenes, como datos de clientes internos y datos demográficos adquiridos. Si desea obtener más información, consulte el tema [Nodo Fundir](#) el p. 89.



El nodo Añadir concatena conjuntos de registros. Es útil para combinar conjuntos de datos con estructuras parecidas, pero con datos diferentes. Si desea obtener más información, consulte el tema [Nodo Añadir](#) el p. 100.



El nodo Distinguir se puede usar para eliminar registros duplicados pasando el primero de los registros distintos a la ruta de datos o descartando el primer registro y pasando cualquier duplicado a la ruta de datos en su lugar. Si desea obtener más información, consulte el tema [Nodo Distinguir](#) el p. 102.

Algunos de los nodos de la paleta Operaciones con registros requieren el uso de una expresión CLEM. Si está familiarizado con CLEM, puede introducir una expresión en el campo. Sin embargo, todos los campos de expresión ofrecen un botón que abre el generador de expresiones CLEM, que ayuda a crear estas expresiones de forma automática.

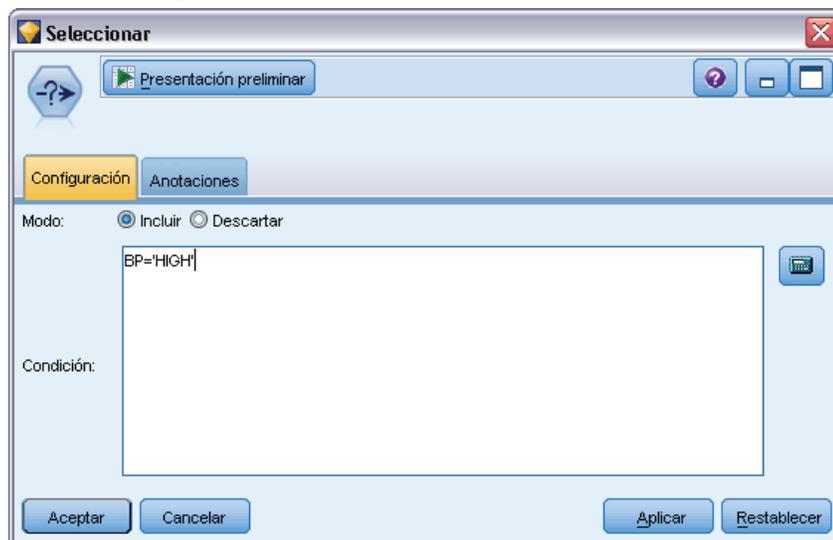
Figura 3-1  
*Botón Generador de expresiones*



## ***Nodo Seleccionar***

Puede usar el nodo Seleccionar para elegir o descartar un subconjunto de registros de la ruta de datos basándose en una condición especial, como PS (presión sanguínea) = "HIGH" (alta).

Figura 3-2  
Cuadro de diálogo del nodo Seleccionar



**Moda.** Especifica si los registros que cumplen la condición se incluirán o se excluirán de la ruta de datos.

- **Incluir.** Seleccione esta opción para incluir los registros que reúnan la condición de selección.
- **Descartar.** Seleccione esta opción para excluir los registros que reúnan la condición de selección.

**Condición.** Muestra la condición de selección que se usará para comprobar cada registro, lo que se especifica usando una expresión CLEM. Puede escribir una expresión en la ventana o bien usar el generador de expresiones pulsando en el botón de calculadora (Generador de expresiones) ubicado en la parte derecha de la ventana.

Si selecciona descartar los registros basados en una condición, como los siguientes:

`(var1='valor1' y var2='valor2')`

el nodo Seleccionar por defecto también descarta los registros con valores nulos para todos los campos de selección. Para evitarlo, añada la siguiente condición a la original:

y no `(@NULL(var1) y @NULL(var2))`

Los nodos Seleccionar también se usan para seleccionar una proporción de registros. No obstante, por lo general, se usaría un nodo diferente para esta operación. Sin embargo, si la condición que desea especificar es más compleja que los parámetros que se proporcionan, puede crearla usando el nodo Seleccionar. Por ejemplo, podría usar una condición como ésta:

`BP = "HIGH" and random(10) <= 4`

Con ella se seleccionará aproximadamente el 40% de los registros que muestran alta presión sanguínea y pasarlos a la parte de abajo para sus análisis posteriores.

## Nodo Muestrear

Puede utilizar los nodos Muestrear para seleccionar un subconjunto de análisis o especificar una proporción de registros que descartar. Se admite una variedad de tipos de muestras, entre las que se incluyen las muestras estratificadas, conglomeradas y no aleatorias (estructuradas). El muestreo se puede utilizar por varias razones:

- Para mejorar el rendimiento por modelos de estimación en un subconjunto de datos. Los modelos estimados de una muestra son normalmente tan precisos como los derivados del conjunto de datos completo y pueden serlo más si el rendimiento mejorado permite experimentar con diferentes métodos que de lo contrario no intentaría.
- Para seleccionar grupos de registros relacionados o transacciones de análisis, como la selección de todos los elementos en un carro de la compra en línea (o cesta de la compra) o todas las propiedades en una vecindad específica.
- Para identificar las unidades o los casos de inspección aleatoria en busca de la garantía de calidad, la prevención de fraudes o la seguridad.

*Nota:* Si simplemente desea realizar una partición de sus datos en muestras de comprobación y entrenamiento con fines de validación, puede utilizar un nodo Partición. Si desea obtener más información, consulte el tema [Nodo Partición](#) en el capítulo 4 el p. 205.

### **Tipos de muestras**

**Muestras conglomeradas.** Grupos o conglomerados de muestra en lugar de unidades individuales. Por ejemplo, supongamos que tiene un archivo de datos con un registro por alumno. Si ha agrupado por colegio y el tamaño de la muestra es 50 %, se seleccionará el 50 % de colegios y se tomarán todos los alumnos de cada colegio seleccionado. Se rechazarán los alumnos de los colegios no seleccionados. Como media, esperaría una selección del 50 % de alumnos aproximadamente, pero dado que los colegios tienen distintos tamaños, el porcentaje no será exacto. De forma similar, podría agrupar elementos del carro de la compra por ID de transacciones para asegurarse de que se conservan todos los elementos de las transacciones seleccionadas. Para un ejemplo que agrupe las propiedades por ciudad, consulte la ruta de la muestra `complexsample_property.str`.

**Muestras estratificadas.** Selecciona muestras de forma independiente dentro de subgrupos no superpuestos de población o estratos. Por ejemplo, puede asegurar que se toman hombres y mujeres en proporciones iguales o que se representan todas las regiones o grupos socioeconómicos dentro de una población urbana. También puede especificar un tamaño de muestra diferente para cada estrato (por ejemplo, si piensa que un grupo se ha representado en inferioridad de condiciones en los datos originales). Para un ejemplo que estratifique las propiedades por provincia, consulte la ruta de la muestra `complexsample_property.str`.

**Muestreo sistemático o 1-de cada-n.** Cuando es difícil obtener la muestra aleatoriamente, las unidades se pueden muestrear sistemáticamente (con un intervalo fijo) o secuencialmente.

**Ponderaciones muestrales.** Las ponderaciones de muestreo se calculan automáticamente al extraer una muestra compleja y corresponde aproximadamente a la “frecuencia” que cada unidad muestreada representa en los datos originales. Por consiguiente, la suma de las ponderaciones sobre la muestra debe calcular el tamaño de los datos originales.

**Marco de muestreo**

Un marco de muestreo define el origen potencial de casos que se deben incluir en una muestra o estudio. En algunos casos, puede ser factible identificar cada miembro individual de una población e incluir cualquiera de ellos en una muestra, por ejemplo, cuando realizan muestreos de elementos extraídos de una línea de producción. La mayoría de las veces no podrá acceder a cada caso posible. Por ejemplo, no puede estar seguro de quién votará en una elección hasta que se haya producido la elección. En este caso, podría utilizar el registro electoral como marco de muestreo, incluso si algunas personas registradas no votan y algunas personas podrían votar a pesar de no haberse incluido en la lista en el momento que comprobó el registro. Cualquiera fuera del marco de muestreo no tiene posibilidades de poder incluirse en la muestra. La cuestión de si su marco de muestreo es lo bastante cercano a la población que intenta evaluar debe estudiarse para cada caso real.

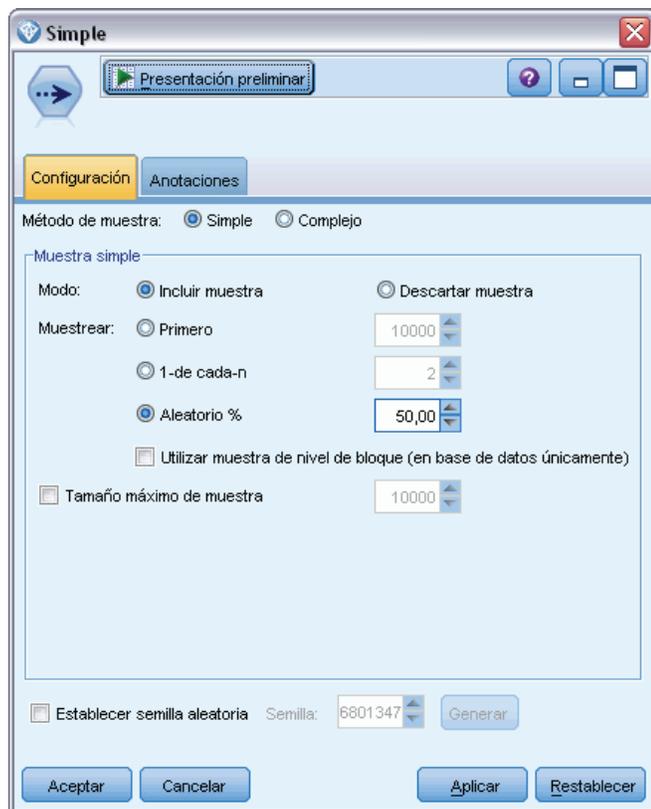
**Opciones del nodo Muestrear**

Puede seleccionar el método Simple o Complejo según se adapte mejor a sus necesidades.

**Opciones de muestreo simple**

El método Simple permite seleccionar un porcentaje aleatorio de registros, seleccionar registros contiguos o seleccionar cada  $n^{\circ}$  registro.

Figura 3-3  
Opciones de muestreo simple



**Moda.** Seleccione si desea pasar (incluir) o descartar (excluir) registros de las siguientes maneras:

- **Incluir muestra.** Incluye los registros seleccionados en la ruta de datos y descarta todos los demás. Por ejemplo, si configura el modo como Incluir muestra y establece el valor de 1-de cada-n en 5, se incluirá cada quinto registro, produciendo un conjunto de datos que representa un quinto del tamaño original aproximadamente. Es el modo por defecto cuando se realiza un muestreo de datos y el único modo posible con el método complejo.
- **Descartar muestra.** Excluye los registros seleccionados e incluye todos los demás. Por ejemplo, si configura el modo en Descartar muestra y establece el valor de 1-de cada-n en 5, se descartará cada quinto informe. Este nodo sólo está disponible con el método simple.

**Muestra.** Seleccione el método de muestreo de entre las opciones siguientes:

- **Primeros.** Seleccione esta opción para usar muestro de datos contiguos. Por ejemplo, si el tamaño máximo de la muestra se define como 10000, se seleccionarán los primeros 10.000 registros.
- **1-de cada-n.** Seleccione esta opción para muestrear datos pasando o descartando cada *enésimo* registro. Por ejemplo, si *n* se define como 5, se seleccionarán cada cinco registros.
- **Aleatorio %.** Seleccione esta opción para muestrear un porcentaje aleatorio de los datos. Por ejemplo, si el porcentaje está definido como 20, se pasará a la ruta de datos o se descartará el 20% de los datos, dependiendo del modo seleccionado. Use el campo para especificar un

porcentaje de muestreo. También puede especificar un valor de semilla usando el control Establecer semilla aleatoria.

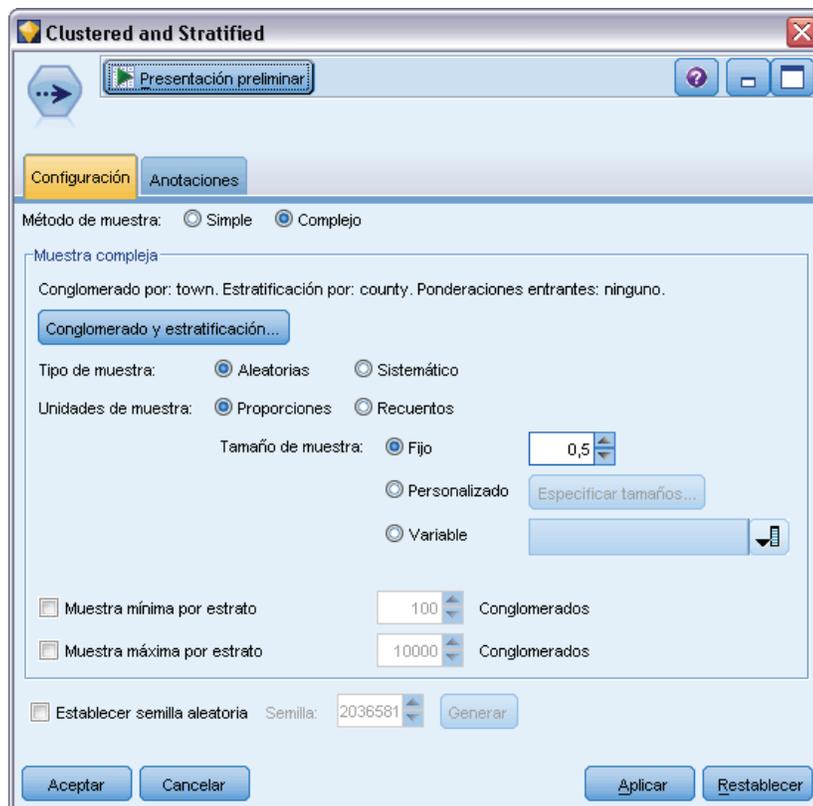
**Utilizar muestreo de nivel de bloque (sólo en base de datos).** Esta opción sólo está habilitada si selecciona el muestreo de porcentaje aleatorio al realizar minería en base de datos en una base de datos de Oracle o IBM DB2. En estas circunstancias, el muestreo de nivel de bloque puede ser más eficaz.

**Tamaño máximo de muestra.** Especifica el número máximo de registros que se incluyen en la muestra. Esta opción es redundante, por lo que se desactiva cuando se seleccionan Primeros e Incluir. Asimismo tenga en cuenta que cuando se utiliza junto con la opción Aleatorio %, esta configuración puede evitar que se seleccionen determinados registros. Por ejemplo, si tiene 10 millones de registros en la base de datos y selecciona el 50 % de ellos con un tamaño de muestra máximo de tres millones de registros, sólo se seleccionará el 50 % de los primeros seis millones de registros y los cuatro millones restantes no tendrán probabilidades de ser seleccionados. Para evitar esta limitación, seleccione el método de muestreo Complejo y solicite una muestra aleatoria de tres millones de registros sin especificar un conglomerado o variable de estratificación.

### Opciones de muestreo complejo

Las opciones de muestreo complejo permiten un control más exhaustivo de la muestra, incluidas las muestras conglomeradas, estratificadas y ponderadas junto con otras opciones.

Figura 3-4  
Opciones de muestreo complejo



**Conglomerado y estratificación.** Permite especificar campos de ponderación de entrada, estratificación y conglomerado en caso necesario. Si desea obtener más información, consulte el tema [Configuración de Conglomerado y estratificación](#) el p. 77.

**Tipo de muestra.**

- **Aleatorios.** Selecciona conglomerados o registros aleatoriamente dentro de cada estrato.
- **Systematic.** Selecciona registros en un intervalo fijo. Esta opción funciona como el método *1 de cada n*, salvo que la posición del primer registro cambia dependiendo de una semilla aleatoria. El valor de *n* se determina automáticamente en base al tamaño de la muestra o la proporción.

**Unidades de muestra.** Puede seleccionar proporciones o recuentos como unidades de muestra básicas.

**Tamaño muestral.** Puede especificar el tamaño de la muestra de varias formas:

- **Fija.** Permite especificar el tamaño general de la muestra como un recuento o proporción.
- **Personalizado.** Permite especificar el tamaño de la muestra para cada subgrupo o estrato. Esta opción sólo está disponible si se ha especificado un campo de estratificación en el subcuadro de diálogo Conglomerado y estratificación.
- **Variable.** Permite al usuario seleccionar un campo que define el tamaño de la muestra para cada subgrupo o estrato. Este campo debe tener el mismo valor para cada registro dentro de un estrato particular; por ejemplo, si la muestra se estratifica por provincia, todos los registros con *county = Surrey* deben tener el mismo valor. El campo debe ser numérico y sus valores deben coincidir con las unidades de muestra seleccionadas. Para las proporciones, los valores deben ser mayores que 0 y menor que 1; para los recuentos, el valor mínimo es 1.

**Muestra mínima por estrato.** Especifica un número mínimo de registros (o un número mínimo de conglomerados si se especifica un campo de conglomerado).

**Muestra máxima por estrato.** Especifica un número máximo de registros o conglomerados. Si selecciona esta opción sin especificar un conglomerado o campo de estratificación, se seleccionará una muestra aleatoria o sistemática del tamaño especificado.

**Establecer semilla aleatoria.** Cuando se muestran o particionan registros a partir de un porcentaje aleatorio, esta opción permite duplicar los mismos resultados en otra sesión. Al especificar el valor inicial utilizado por el generador de números aleatorios, puede garantizar que se asignan los mismos registros cada vez que se ejecuta el nodo. Introduzca el valor de semilla deseado, o pulse en el botón Generar para generar automáticamente un valor aleatorio. Si no se selecciona esta opción, se generará una muestra diferente cada vez que se ejecute el nodo.

*Nota:* cuando se utiliza la opción Establecer semilla aleatoria con registros leídos de una base de datos, puede ser necesario un nodo Ordenar, antes del muestreo con el fin de garantizar el mismo resultado cada vez que se ejecute el nodo. Esto se debe a que la semilla aleatoria depende del orden de registros, sin estar garantizado que sea el mismo en una base de datos relacional. Si desea obtener más información, consulte el tema [Nodo Ordenar](#) el p. 87.

## Configuración de Conglomerado y estratificación

El cuadro de diálogo Conglomerado y estratificación permite seleccionar campos de ponderación, estratificación y conglomerado cuando extrae una muestra compleja.

Figura 3-5  
Configuración de campo Conglomerado y estratificación



**Conglomerados.** Especifica un campo categórico empleado para agrupar registros. Los registros se muestrean en base a su pertenencia al conglomerado, incluyendo unos conglomerados y otros no. Pero si se incluye cualquier registro de un conglomerado determinado, se incluirán todos. Por ejemplo, al analizar las asociaciones de producto en los carros de la compra, podría agrupar elementos por ID de transacciones para asegurarse de que se conservan todos los elementos de las transacciones seleccionadas. En lugar de muestrear registros, que destruirá la información sobre los elementos que se venden conjuntamente, puede muestrear transacciones para asegurarse de que se conservan todos los registros de las transacciones seleccionadas.

**Estratificar por.** Especifica un campo categórico utilizado para estratificar los registros de manera que las muestras se seleccionan de forma independiente dentro de subgrupos no superpuestos de población o estratos. Si selecciona una muestra del 50 % estratificada por género, por ejemplo, se tomarán dos muestras del 50 %, una para hombres y otra para mujeres. Por ejemplo, los estratos pueden ser grupos socioeconómicos, categorías laborales, grupos de edades o grupos étnicos, permitiendo garantizar los tamaños de muestra adecuados para los subgrupos de interés. Si hay el triple de mujeres que de hombres en el conjunto de datos original, esta relación se mantendrá realizando muestreos de forma separada para cada grupo. También se pueden especificar varios campos de estratificación (por ejemplo, mediante el muestreo de líneas de producto dentro de regiones o viceversa).

*Nota:* Si estratifica por un campo que tiene valores perdidos (los valores perdidos del sistema o nulos, cadenas vacías, espacios blancos y valores perdidos definidos por el usuario o en blanco), no podrá especificar tamaños de muestra personalizados por estratos. Si desea utilizar los tamaños de muestras personalizados al estratificar por campo con valores en blanco o perdidos, tendrá que especificarlos antes.

**Utilizar ponderación de entrada.** Especifica un campo empleado para ponderar registros antes del muestreo. Por ejemplo, si el campo de ponderación tiene valores comprendidos entre 1 y 5, los registros ponderados con 5 tienen una probabilidad de cinco de ser seleccionados. Los valores

de este campo se sobrescribirán con las ponderaciones de salida final generadas por el nodo (consulte el siguiente párrafo).

**Nueva ponderación de resultados.** Especifica el nombre del campo donde se incluyen las ponderaciones finales si no se especifica ningún campo de ponderación de entrada. (Si se especifica un campo de ponderación de entrada, sus valores se sustituyen por las ponderaciones finales como se indica arriba y no se crea ningún campo de ponderación de resultados independiente.) Los valores de ponderación de resultados indican el número de registros representados por cada registro muestreado en los datos originales. La suma de los valores ponderados ofrece una estimación del tamaño de la muestra. Por ejemplo, si se toma una muestra aleatoria del 10 %, la ponderación de resultados será 10 para todos los registros, indicando que cada registro muestreado representa aproximadamente 10 registros en los datos originales. En una muestra estratificada o ponderada, los valores de ponderación de resultados pueden variar en base a la proporción de la muestra para cada estrato.

### **Comentarios**

- El muestreo conglomerado es útil si no puede obtener una lista completa de la población que desea muestrear, pero puede obtener listas completas para determinados grupos o conglomerados. También se utiliza cuando una muestra aleatoria produce una lista de sujetos de comprobación con los que sería poco práctico ponerse en contacto. Por ejemplo, sería más fácil visitar a todos los granjeros de una provincia que a una selección de granjeros de distintas provincias dentro del mismo país.
- Puede especificar tanto el conglomerado como los campos de estratificación para muestrear los conglomerados de forma independiente para cada estrato. Por ejemplo, puede muestrear valores de la propiedad estratificados por provincia y el conglomerado por ciudad dentro de cada provincia. Así, garantizará que una muestra independiente de ciudades se extrae de cada provincia. Algunas ciudades se incluirán y otras no, pero cada ciudad que se incluya, se considerarán todas las propiedades de la ciudad.
- Para seleccionar una muestra aleatoria de unidades dentro de cada conglomerado, puede encadenar dos nodos Muestrear. Por ejemplo, primero podría muestrear municipios estratificados por provincia como se ha descrito anteriormente. A continuación, se añadiría un segundo nodo Muestrear y seleccionaría *ciudad* como campo de estratificación, lo que le permitiría muestrear una proporción de registros de cada municipio.
- En los casos en los que se requiere una combinación de campos para identificar conglomerados de forma exclusiva, se puede generar un nuevo campo mediante un nodo Derivar. Por ejemplo, si varias tiendas utilizan el mismo sistema de numeración para las transacciones, podría derivar un nuevo campo que concatene la tienda y los ID de transacción.

### **Tamaños de muestra por estratos**

Cuando extraiga una muestra estratificada, la opción por defecto es muestrear la misma proporción de registros o conglomerados de cada estrato. Si un grupo supera a otro por un factor de 3, por ejemplo, normalmente querrá mantener la misma relación en la muestra. Sin embargo, si no es el caso, puede especificar el tamaño de la muestra de forma independiente para cada estrato.

Figura 3-6  
Especificación de tamaños de muestra por estratos

county	nrhood	Tamaño de muestra
1	1	0,45
1	2	0,55
1	3	0,6
1	4	0,6
1	5	Por defecto
1	6	Por defecto
1	8	Por defecto
1	9	Por defecto
1	10	Por defecto
1	11	Por defecto
1	15	Por defecto
1	16	Por defecto
1	17	Por defecto
1	18	Por defecto
1	22	Por defecto
1	23	Por defecto
1	24	Por defecto
1	25	Por defecto

Ordenar por: county  Ascendente  Descendente

Leer valores Borrar tamaños

Tamaño de muestra por defecto por estrato: 0,5 Registros

Aceptar Cancelar Ayuda

El cuadro de diálogo Tamaños de muestra por estratos enumera cada valor del campo de estratificación, permitiendo sobrescribir el valor por defecto para ese estrato. Si se seleccionan varios campos de estratificación, se enumeran todas las combinaciones de valores posibles, permitiendo especificar el tamaño de cada grupo étnico dentro de cada ciudad, por ejemplo, o cada ciudad dentro de cada provincia. Los tamaños se especifican como proporciones o recuentos, según determina la configuración actual del nodo Muestrear.

### **Especificación de tamaños de muestra por estratos**

- ▶ En el nodo Muestrear, seleccione Complejo y seleccione uno o más campos de estratificación. Si desea obtener más información, consulte el tema [Configuración de Conglomerado y estratificación](#) el p. 77.
- ▶ Seleccione Personalizado y Especificar tamaños.
- ▶ En el cuadro de diálogo Tamaños de muestra por estratos, pulse en el botón Leer valores en la parte inferior izquierda para rellenar la presentación. En caso necesario, es posible que tenga que instanciar los valores en un nodo Tipo o de origen anterior de la ruta. Si desea obtener más información, consulte el tema [¿Qué es la creación de una instancia?](#) en el capítulo 4 el p. 141.
- ▶ Pulse en cualquier fila para sobrescribir el tamaño por defecto de ese estrato.

### **Notas sobre el tamaño de la muestra**

Los tamaños de muestra personalizados pueden ser útiles si los distintos estratos tienen varianzas diferentes, por ejemplo, para hacer que los tamaños de la muestra sean proporcionales a la desviación típica. (Si los casos dentro del estrato son más variados, tendrá que muestrear más estratos para obtener una muestra representativa.) O si un estrato es pequeño, es posible que desee utilizar una proporción de muestra mayor para asegurar que se ha incluido un número mínimo de observaciones.

*Nota:* Si estratifica por un campo que tiene valores perdidos (los valores perdidos del sistema o nulos, cadenas vacías, espacios blancos y valores perdidos definidos por el usuario o en blanco), no podrá especificar tamaños de muestra personalizados por estratos. Si desea utilizar los tamaños de muestras personalizados al estratificar por campo con valores en blanco o perdidos, tendrá que especificarlos antes.

## **Nodo Equilibrar**

Los nodos Equilibrar se pueden usar para corregir los desequilibrios de los conjuntos de datos de modo que cumplan determinados criterios de comprobación. Por ejemplo, imagine que un conjunto de datos contiene sólo dos valores, *bajo* o *alto*, y que el 90% de los casos es *bajo* y sólo el 10% de ellos es *alto*. Muchas técnicas de modelado presentan problemas con estos datos sesgados, puesto que tenderán a aprender sólo el resultado *bajo* y omitirán el valor *alto*, puesto que es más inusual. Si los datos están bien balanceados con aproximadamente el mismo número de resultados de *bajo* y *alto*, los modelos tendrán más posibilidades de encontrar patrones que hagan la distinción entre los dos grupos. En este caso, un nodo Equilibrar resulta útil para la creación de una directiva de equilibrado que reduzca los casos con un resultado *bajo*.

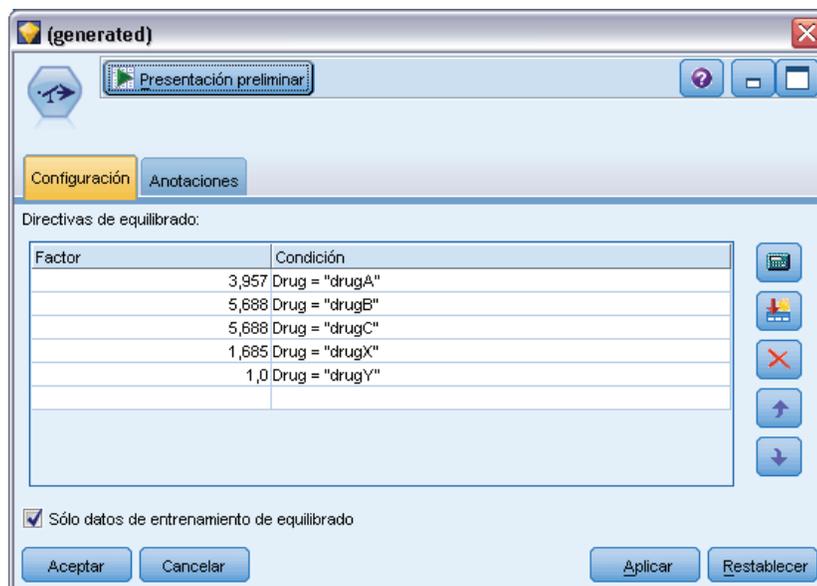
El equilibrado se lleva a cabo mediante el duplicado y posterior descarte de registros basándose en las condiciones especificadas. Los registros para los que no se establece ninguna condición siempre se pasan. Como este proceso funciona duplicando y descartando registros, la secuencia original de los datos se pierde en las operaciones efectuadas más abajo. Asegúrese de derivar cualquier valor relacionado con la secuencia antes de añadir un nodo Equilibrar a la ruta de datos.

*Nota:* Los nodos Equilibrar se pueden generar de forma automática desde histogramas y gráficos de distribución. Por ejemplo, puede equilibrar sus datos para mostrar proporciones iguales en todas las categorías de un campo categórico, como se muestra en un gráfico de distribución.

**Ejemplo.** Cuando genere una ruta de RFM para identificar clientes recientes que han respondido de forma positiva a campañas de marketing anteriores, el departamento de marketing de una empresa de ventas utiliza un nodo Equilibrar para equilibrar las diferencias entre las respuestas verdaderas y falsas de los datos.

## Opciones de configuración del nodo Equilibrar

Figura 3-7  
Configuración del nodo Equilibrar



**Directivas de equilibrado de registros.** Enumera las directivas de equilibrado actuales. Cada directiva incluye un factor y una condición que indica al software que “aumente la proporción de registros por un factor determinado si la condición es verdadera.” Un factor inferior a 1,0 significa que la proporción de registros indicada se reducirá. Por ejemplo, si desea reducir el número de registros en los que el medicamento Y es el adecuado para el tratamiento, podría crear una directiva de equilibrado con un factor 0,7 y una condición Drug = "drugY". Esta directiva significa que el número de registros en los que el medicamento Y es el apropiado para el tratamiento se reducirá al 70% en todas las operaciones siguientes.

*Nota:* los factores de equilibrado para la reducción se deben especificar con cuatro posiciones decimales. Los factores con valores inferiores a 0,0001 generarán un error, ya que los resultados no se calculan correctamente.

- **Cree condiciones** pulsando en el botón situado a la derecha del campo de texto. Así se insertará una fila vacía para introducir nuevas condiciones. Si desea crear una expresión CLEM para la condición, pulse en el botón Generador de expresiones.
- **Elimine directivas** usando el botón de eliminación rojo.
- **Ordene directivas** usando los botones de flecha arriba y abajo.

**Sólo datos de entrenamiento de equilibrado** Si un campo de partición está presente en la ruta, esta opción equilibra solamente los datos de la partición de entrenamiento. En especial, puede ser útil si genera puntuaciones ajustadas de propensión, que requiere una comprobación desequilibrada o partición de validación. Si no hay campos de partición presentes en la ruta (o si se especifican varios campos de partición), esta opción se ignora y se equilibran todos los datos.

## Nodo Agregar

La agregación es una tarea de preparación de datos que se usa con frecuencia para reducir el tamaño de un conjunto de datos. Antes de comenzar a añadir elementos, debe dedicar un tiempo a limpiar los datos, concentrándose especialmente en los valores perdidos. Una vez que haya realizado la agregación, es posible que se pierda información potencialmente útil sobre valores perdidos.

El nodo Agregar se puede usar para reemplazar una secuencia de registros de entrada con registros de salida agregados de resumen. Por ejemplo, podría tener un conjunto de registros de entradas de ventas como éste:

Edad	Sexo	Región	Rama	Ventas
23	M	S	8	4
45	M	S	16	4
37	M	S	8	5
30	M	S	5	7
44	M	N	4	9
25	M	N	2	11
29	F	S	16	6
41	F	N	4	8
23	F	N	6	2
45	F	N	4	5
33	F	N	6	10

Se pueden agregar estos registros con *Sexo* y *Región* como campos clave. A continuación puede agregar *Edad* con el modo Media y *Ventas* con el modo Suma. Seleccione Incluir recuento de registros en campo en el cuadro de diálogo del nodo Agregar para que el resultado agregado sea:

Edad (media)	Sexo	Región	Ventas (suma)	Recuento de registros
35.5	F	N	25	4
29	F	S	6	1
34.5	M	N	20	2
33.75	M	S	20	4

Del ejemplo se entiende, por ejemplo, que la edad media de las cuatro mujeres del personal de ventas en la región norte es de 35,5, y que la suma de sus ventas totales es de 25 unidades.

*Nota:* los campos como *Rama* se descartan automáticamente cuando no se especifica ningún modo de agregación.

### Opciones de configuración del nodo Agregar

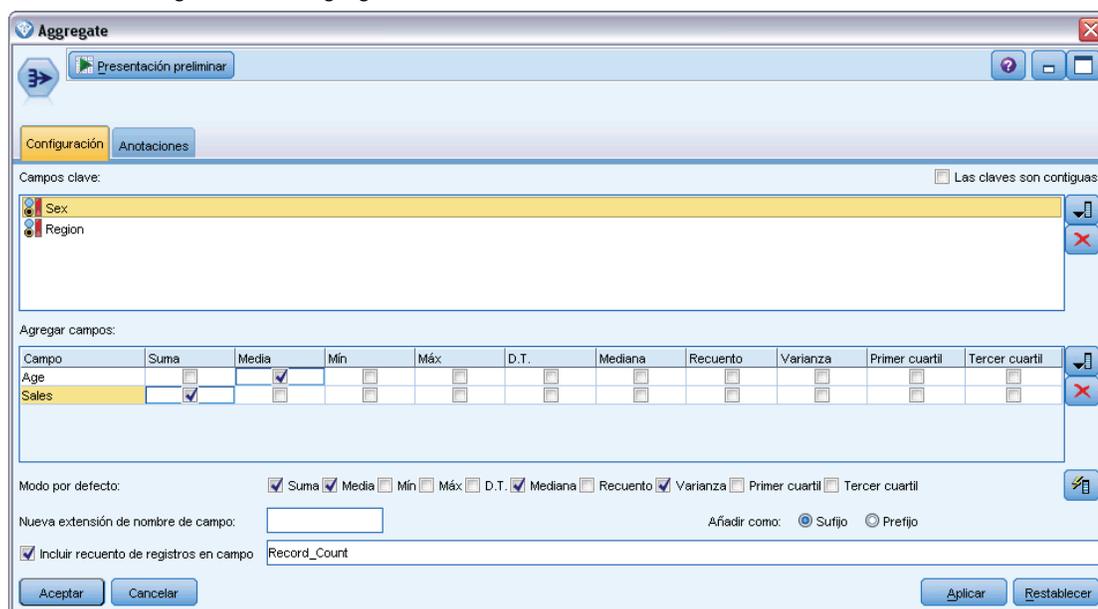
En el nodo Agregar puede especificar lo siguiente.

- Una o más claves de registro que se utilizarán como categorías para la agregación

- Uno o más campos de agregación para calcular los valores de agregación
- Uno o más modos de agregación (tipos de agregación) para obtener el resultado de cada campo de agregación

También puede especificar los modos de agregación por defecto que se utilizarán para utilizar los campos recién añadidos.

Figura 3-8  
Cuadro de diálogo del nodo Agregar



**Campos clave.** Enumera los campos que se pueden usar como categorías para la agregación. Tanto los campos (numéricos) como los campos categóricos se pueden usar como claves. Si selecciona más de un campo clave, los valores se combinarán para producir un valor clave de la agregación de registros. Se generará un registro agregado para cada campo clave único. Por ejemplo, si *Sexo* y *Región* son los campos clave, cada combinación única de *V* y *M* con las regiones *N* y *S* (cuatro combinaciones exclusivas) tendrá un registro agregado. Para añadir un campo clave, utilice el botón de selección de campos de la parte derecha de la ventana.

**Las claves son contiguas.** Seleccione esta opción si sabe que todos los registros con los mismos valores clave se agrupan en la entrada (por ejemplo, si la entrada se clasifica en los campos clave). Con ello puede mejorar el rendimiento.

**Agregar campos.** Enumera los campos cuyos valores se añadirán, así como los modos de agregación seleccionados. Para añadir campos a esta lista, utilice el botón de selección de campos de la parte derecha. Los siguientes modos de agregación están disponibles:

*Nota:* Algunos de los modos no son aplicables a campos no numéricos (por ejemplo, Sum para un campo de fecha/hora). Los modos que no se pueden utilizar con un campo de agregación seleccionada están desactivados.

- **Suma.** Seleccione esta opción para devolver los resultados sumados de cada combinación de campos clave. La suma es el total de todos los valores, a lo largo de todos los casos que no tengan valores perdidos.
- **Media.** Seleccione esta opción para devolver el valor de la media de cada combinación de campos clave. La media es una medición de la tendencia central, y es el promedio aritmético (la suma dividida por el número de casos).
- **Mín.** Seleccione esta opción para devolver los valores mínimos de cada combinación de campos clave.
- **Máx.** Seleccione esta opción para devolver los valores máximos de cada combinación de campos clave.
- **D.T.** Seleccione esta opción para devolver la desviación típica de cada combinación de campos clave. La desviación típica es una medida de la dispersión en torno a la media, igual a la raíz cuadrada de la medición de la varianza.
- **Mediana.** Seleccione esta opción para devolver el valor de la mediana de cada combinación de campos clave. La mediana es una medida de tendencia central que no es sensible a los valores atípicos (a diferencia de la media, que puede resultar afectada por unos pocos valores extremadamente altos o bajos). También se denomina percentil 50° o 2° cuartil.
- **Recuento.** Seleccione esta opción para devolver el recuento de valores no nulos de de cada combinación de campos clave.
- **Varianza.** Seleccione esta opción para devolver el valor de la varianza de cada combinación de campos clave. La varianza es una medida de dispersión en torno a la media, igual a la suma de las desviaciones al cuadrado respecto a la media, dividida por el número de casos menos 1.
- **1° cuartil.** Seleccione esta opción para devolver los valores del 1° cuartil (percentil 25°) de cada combinación de campos clave.
- **3° cuartil.** Seleccione esta opción para devolver los valores del 3° cuartil (percentil 75°) de cada combinación de campos clave.

**Modo por defecto.** Especifique el modo de agregación por defecto que se debe usar con los campos recientemente añadidos. Si utiliza con frecuencia la misma agregación, seleccione uno o varios modos aquí y use el botón Aplicar a todos de la parte derecha para que los modos seleccionados se apliquen a todos los campos de la lista.

**Nueva extensión de nombre de campo.** Seleccione esta opción para añadir un sufijo o prefijo, como “1” o “nuevo”, a los campos agregados duplicados. Por ejemplo, el resultado de una agregación de valores mínimos en el campo *Edad* producirá un nombre de campo denominado *Edad\_Mín\_1* si se seleccionó la opción de sufijo y especifica “1” como extensión. *Nota:* las extensiones de agregación como *\_Mín* o *Máx\_* se añaden automáticamente al nuevo campo, indicando el tipo de agregación realizada. Seleccione Sufijo o Prefijo para indicar el estilo de extensión que prefiera.

**Incluir recuento de registros en campo.** Seleccione esta opción para incluir un campo adicional en cada registro de salida con el nombre por defecto *Recuento\_registros*. Este campo indica la cantidad de registros de entrada añadidos para conformar cada registro añadido. Asigne un nombre personalizado a este campo escribiéndolo directamente en el cuadro de edición.

*Nota:* cuando se calculan agregados pero no se incluyen en el recuento de registros, se ejecutan valores nulos del sistema. Por otro lado, se incluyen valores vacíos en la agregación y el recuento de registros. Para excluir los valores vacíos, puede usar el nodo Rellenar, que permite sustituir

los espacios vacíos por valores nulos. También puede eliminar los espacios vacíos usando un nodo Seleccionar.

### **Rendimiento**

Las operaciones de agregación pueden mejorarse al activar el procesamiento paralelo.

## **Nodo Adición de RFM**

El nodo Adición de RFM (actualidad, frecuencia, monetario) permite tomar datos de transacciones históricas de clientes, deshacerse de los datos no utilizados y combinar todos los datos de transacciones restantes en una única fila, utilizando el ID de cliente exclusivo como clave, que enumera cuándo hizo negociaciones con los clientes por última vez (actualidad), cuántas transacciones hicieron (frecuencia) y el valor total de dichas transacciones (monetario).

Antes de comenzar a añadir elementos, debe dedicar un tiempo a limpiar los datos, concentrándose especialmente en los valores perdidos.

Una vez que haya identificado y transformado los datos mediante el nodo Adición de RFM, puede utilizar un nodo Análisis de RFM para realizar más análisis. Si desea obtener más información, consulte el tema [Nodo Análisis de RFM](#) en el capítulo 4 el p. 201.

Tenga en cuenta que una vez ejecutado el archivo de datos mediante el nodo Adición de RFM, no tendrá ningún valor objetivo, por consiguiente, antes de poder utilizarlo como entradas para análisis predictivos adicionales con nodos de modelado como C5.0 o CHAID, tendrá que fusionarlos con otros datos de cliente (por ejemplo, haciendo coincidir los ID de cliente). Si desea obtener más información, consulte el tema [Nodo Fundir](#) el p. 89.

Los nodos Adición de RFM y Análisis de RFM de IBM® SPSS® Modeler están configurados para utilizar intervalos independientes; es decir, ordenan por rango y establecen en intervalos los datos de cada medida de actividades recientes, frecuencia y valor monetario, sin considerar sus valores o las otras dos medidas.

## Opciones de configuración del nodo Adición de RFM

Figura 3-9  
Configuración de adición de RFM

**Calcular actualidad relativa a.** Especifica la fecha a partir de la que se calculará la actualidad de las transacciones. Puede ser una Fecha fija que introduzca o Fecha de hoy, definida por su sistema. Fecha de hoy se introduce por defecto y se actualiza automáticamente cuando se ejecuta el nodo.

**Los ID son contiguos.** Si los datos se han clasificado previamente de forma que todos los registros con el mismo ID aparecen en la ruta de datos, seleccione esta opción para que el procesamiento sea más rápido. Si los datos no se han clasificado previamente (o no lo sabe a ciencia cierta), no active esta opción y el nodo clasificará los datos automáticamente.

**ID.** Seleccione el campo que desea utilizar para identificar el cliente y sus transacciones. Para mostrar los campos que puede seleccionar, utilice el botón Selector de campos de la parte derecha.

**Fecha.** Seleccione el campo de fecha que se utilizará para calcular la actualidad. Para mostrar los campos que puede seleccionar, utilice el botón Selector de campos de la parte derecha.

Recuerde que para ello se necesita un campo de fecha o marca de tiempo con el formato apropiado para utilizarlo como entrada. Por ejemplo, si tiene un campo de cadena con valores como *Ene 2007*, *Feb 2007*, etc., puede convertirlo en un campo de fecha mediante el nodo

Rellenar y la función `to_date()`. Si desea obtener más información, consulte el tema [Conversión del almacenamiento mediante el nodo Rellenar](#) en el capítulo 4 el p. 179.

**Valor.** Selecciona el campo que se utilizará para calcular el valor monetario total de las transacciones del cliente. Para mostrar los campos que puede seleccionar, utilice el botón Selector de campos de la parte derecha. *Nota:* Este valor debe ser numérico.

**Nueva extensión de nombre de campo.** Seleccione esta opción para añadir un sufijo o prefijo, como “12\_month”, a los campos actualidad, frecuencia y monetario generados recientemente. Seleccione Sufijo o Prefijo para indicar el estilo de extensión que prefiera. Por ejemplo, puede ser útil cuando examine varios períodos de tiempo.

**Descartar registros con valor inferior a.** En caso necesario, puede especificar un valor mínimo por debajo del cual no se utilice ningún detalle de transacción al calcular los totales de RFM. Las unidades del valor se relacionan con el campo Valor seleccionado.

**Incluir sólo transacciones recientes.** Si analiza una base de datos grande, puede especificar que sólo se utilicen los últimos registros. Puede seleccionar utilizar los datos registrados incluso después de una determinada fecha o dentro de un período reciente:

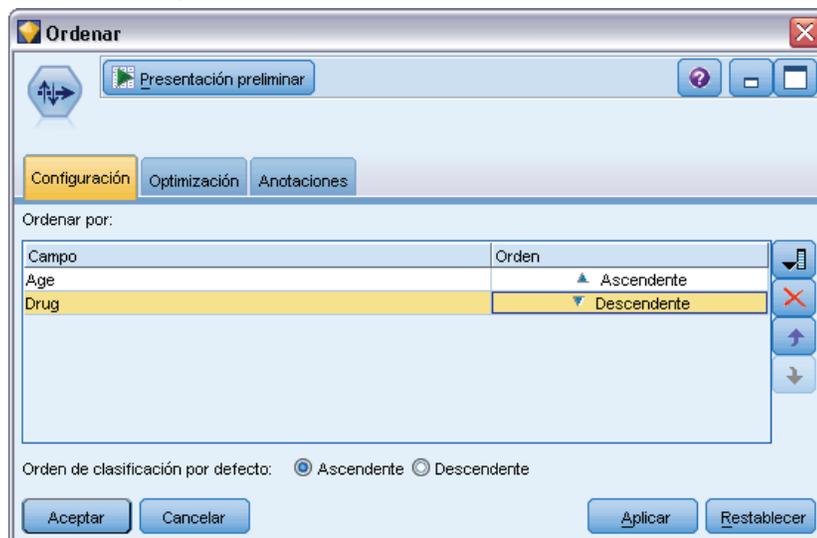
- **Fecha de transacción después de.** Especifica la fecha de transacción tras la que se incluirán los registros en su análisis.
- **Transacción dentro del último.** Especifica el número y tipo de períodos (días, semanas, meses o años) desde la fecha Calcular actualidad relativa a tras la cual se incluirán los registros en su análisis.

**Guardar fecha de la segunda transacción más reciente.** Si desea conocer la fecha de la segunda transacción más reciente para cada cliente, seleccione esta casilla. Asimismo, puede seleccionar la casilla Guardar fecha de la tercera transacción más reciente. Por ejemplo, puede ayudarle a identificar clientes que hayan podido realizar muchas transacciones hace bastante tiempo, pero sólo una transacción reciente.

## Nodo Ordenar

Los nodos Ordenar se pueden usar para organizar registros en orden ascendente o descendente atendiendo a los valores de uno o varios campos. Por ejemplo, los nodos Ordenar se usan con frecuencia para ver y seleccionar registros con los valores de datos más comunes. Generalmente, primero se añaden los datos usando el nodo Agregar y, a continuación se usa el nodo Ordenar para organizar los datos añadidos en el orden descendente del recuento de registros. Si se muestran estos resultados en una tabla, se facilita la exploración de los datos y la toma de decisiones, como la selección de registros de los 10 mejores clientes.

Figura 3-10  
Cuadro de diálogo del nodo Ordenar



**Ordenar por.** Todos los campos seleccionados como claves de ordenación se muestran en una tabla. Un campo clave funciona mejor en la ordenación si es numérico.

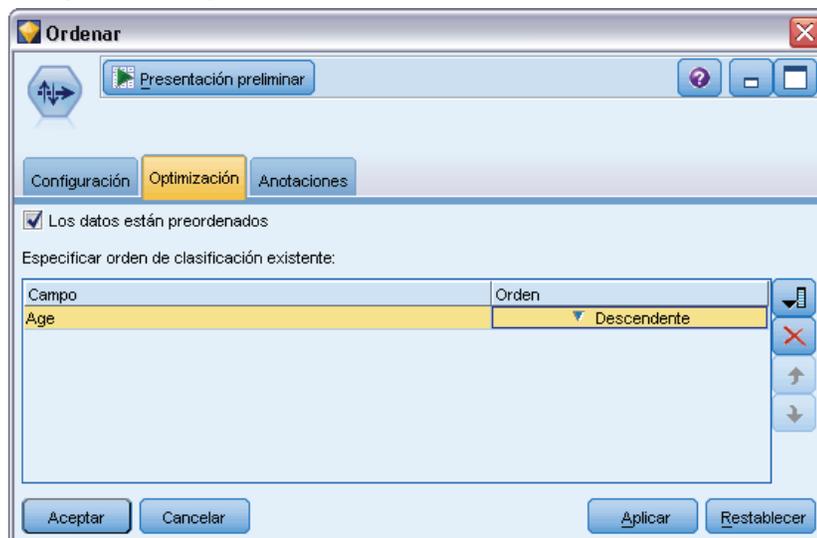
- Para **añadir campos** a esta lista, utilice el botón de selección de campos de la parte derecha.
- **Seleccione un orden** pulsando en las flechas Ascendente o Descendente de la columna *Orden* de la tabla.
- **Elimine campos** usando el botón de eliminación rojo.
- **Ordene directivas** usando los botones de flecha arriba y abajo.

**Orden de clasificación por defecto.** Seleccione Ascendente o Descendente para determinar el orden de clasificación por defecto cuando se añadan nuevos campos a la tabla.

### **Configuración de optimización de ordenación**

Si trabaja con datos que ya se han ordenado por determinados campos clave, puede especificar los campos que ya se han ordenado y permitir que el sistema ordene el resto de los datos de una manera más eficaz. Por ejemplo, desea ordenar por *Edad* (orden descendente) y *Medicamento* (orden ascendente) pero sabe que los datos ya están ordenados por *Edad* (orden descendente).

Figura 3-11  
Configuración de optimización



**Los datos están preordenados.** Determina si los datos ya se han ordenado en función de uno o más campos.

**Especificar orden de clasificación existente.** Determina los campos que ya están ordenados. Puede añadir campos a la lista mediante el cuadro de diálogo Seleccionar campos. En la columna *Orden*, especifique si cada campo se ordena en orden ascendente o descendente. Si especifica varios campos, asegúrese de que los enumera en el orden de clasificación correcto. Utilice las flechas de la derecha de la lista para organizar los campos en el orden correcto. Si comete un error al especificar el orden de clasificación existente, aparecerá un error cuando ejecute la ruta que muestra el número de registro donde la clasificación no coincide con la que especificó.

*Nota:* La velocidad de clasificación puede mejorarse al activar el procesamiento paralelo.

## Nodo Fundir

La función de un nodo Fundir es tomar varios registros de entrada para crear un registro de salida que contenga todos o algunos de los campos de entrada. Se trata de una operación útil cuando se desean fusionar datos de diferentes orígenes, como datos de clientes internos y datos demográficos adquiridos. Se pueden fusionar datos de los siguientes modos.

- **Fusionar por orden:** concatena registros correspondientes procedentes de todos los orígenes en el orden de entrada hasta vaciar el origen de datos más pequeño. Si se usa esta opción, es importante haber ordenado previamente los datos con un nodo Ordenar.

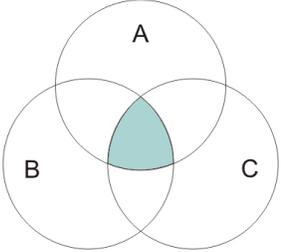
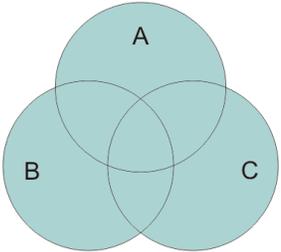
- **Fusionar usando un campo clave**, como el *ID de cliente*, para especificar cómo relacionar los registros procedentes de un origen de datos con los procedentes de otros. Hay varias posibilidades de unión, incluidas la unión interior, la exterior, la exterior parcial y la anti-unión. Si desea obtener más información, consulte el tema [Tipos de uniones](#) el p. 90.
- **Fusionar por condición** hace que se pueda especificar una condición que ha de cumplirse para que pueda realizarse la fusión. Se puede especificar la condición directamente en el nodo, o crear la condición utilizando el generador de expresiones.

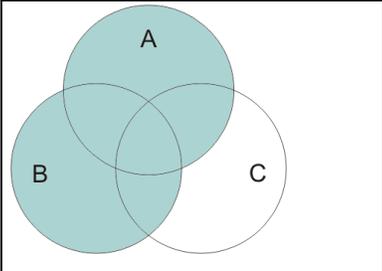
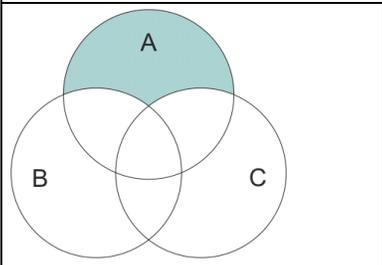
## Tipos de uniones

Cuando se utiliza un campo clave para la fusión de datos, resulta útil dedicar cierto tiempo a decidir qué registros se deben excluir y cuáles se deben incluir. Existe una gran variedad de uniones que se describen detalladamente a continuación.

Los dos tipos básicos de uniones se conocen como interior y exterior. Estos métodos suelen utilizarse para fusionar tablas de conjuntos de datos relacionados según los valores comunes de un campo clave, como *ID de cliente*. Las uniones interiores permiten la realización de fusiones limpias y conjuntos de datos de salida que incluyen únicamente registros completos. Las uniones exteriores también incluyen registros completos procedentes de los datos fusionados, pero también permiten la inclusión de datos exclusivos procedentes de una o varias tablas de entrada.

A continuación se describen con más detalle los tipos de unión permitidos.

	<p>Una <b>unión interior</b> incluye únicamente registros en los que un valor de campo clave es común a todas las tablas de entrada. Es decir, los registros no coincidentes no se incluyen en el conjunto de datos de salida.</p>
	<p>En una <b>unión exterior total</b> se incluyen todos los registros, tanto coincidentes como no coincidentes, procedentes de las tablas de entrada. Las uniones exteriores izquierda y derecha se denominan uniones exteriores parciales y se describen a continuación.</p>

	<p>En una <b>unión exterior parcial</b> se incluyen todos los registros coincidentes por un campo clave y los registros no coincidentes procedentes de las tablas especificadas. (Es decir, todos los registros de algunas etiquetas y sólo los registros coincidentes de otras.) Se pueden seleccionar tablas (como A y B en la ilustración) que incluir en la unión exterior usando el botón Seleccionar de la pestaña Fundir. Las uniones parciales también se denominan uniones exteriores izquierdas o derechas cuando sólo se fusionan dos tablas. Puesto que IBM® SPSS® Modeler permite la fusión de más de dos tablas, nos referimos a este tipo de unión como exterior parcial.</p>
	<p>Una <b>anti-unión</b> incluye únicamente registros no coincidentes de la primera tabla de entrada (la A en la ilustración). Este tipo de unión se opone a la interior y no incluye registros completos en el conjunto de datos resultante.</p>

Por ejemplo, si tiene información acerca de las granjas de un conjunto de datos y de reclamaciones del seguro relacionadas con las granjas en otro, puede establecer una correspondencia entre los registros del primer origen de datos y los del segundo mediante la opción Corresponden con.

Para determinar si un cliente de la muestra de granjas ha rellenado una reclamación del seguro, utilice la opción de unión interior para devolver una lista que muestre dónde coinciden los ID de las dos muestras.

Figura 3-12

Resultado de la muestra para una fusión de unión interior

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	claimtype	claimvalue
1	id604	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes	decomm...	281082.0...
2	id605	name605	north	1700.000	46.000	8.000	621148.000	wheat	decomm...	122006.0...
3	id620	name620	north	880.000	74.000	6.000	426988.000	rapeseed	arable_de	118885.0...

Al utilizar la opción de unión exterior completa, se obtienen tanto los registros coincidentes como los no coincidentes de las tablas de entrada. El valor perdido del sistema (\$null\$) se utilizará para los valores incompletos.

Figura 3-13

Resultado de la muestra de una fusión de unión exterior completa

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	claimtype	claimvalu
1	id601	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	decomm...	74703.1C
2	id602	name602	north	1780.000	42.000	9.000	734118.000	maize	\$null\$	\$nul
3	id604	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes	decomm...	281082.0
4	id605	name605	north	1700.000	46.000	8.000	621148.000	wheat	decomm...	122006.0
5	id606	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	arable_de	122135.0

En una unión exterior parcial se incluyen todos los registros coincidentes por un campo clave y los registros no coincidentes procedentes de las tablas especificadas. La tabla muestra todos los registros coincidentes del campo ID así como los registros coincidentes del primer conjunto de datos.

**Figura 3-14**  
*Resultado de la muestra de una fusión de unión exterior parcial*

	id	claimtype	claimvalue	name	region	farmsize	rainfall	landquality	farmincome	maincrop
1	id602	\$null\$	\$null\$	name602	north	1780.000	42.000	9.000	734118.000	maize
2	id604	decomm...	281082.0...	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes
3	id605	decomm...	122006.0...	name605	north	1700.000	46.000	8.000	621148.000	wheat
4	id607	\$null\$	\$null\$	name607	southeast	1820.000	29.000	6.000	211605.000	maize
5	id608	\$null\$	\$null\$	name608	southeast	1640.000	108.0...	7.000	1167040.0...	maize
6	id609	\$null\$	\$null\$	name609	southwest	1600.000	101.0...	5.000	756755.000	wheat
7	id615	\$null\$	\$null\$	name615	midlands	920.000	86.000	6.000	442554.000	potatoes
8	id618	\$null\$	\$null\$	name618	southeast	1180.000	98.000	3.000	368646.000	maize

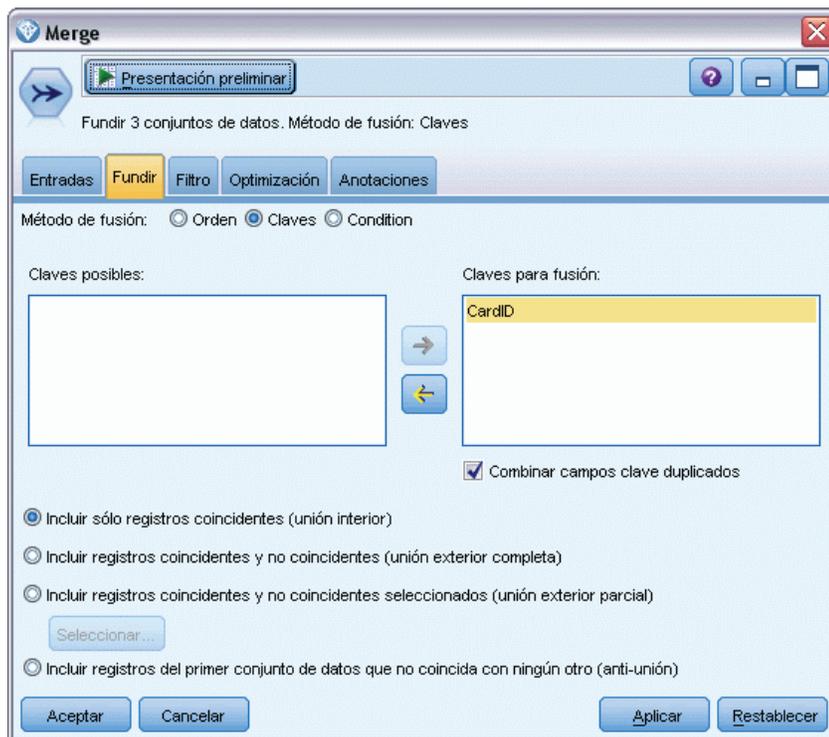
Si utiliza la opción anti-unión, la tabla devolverá sólo los registros no coincidentes para la primera tabla de entrada.

**Figura 3-15**  
*Resultado de la muestra para una fusión anti-unión*

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop
1	id602	name602	north	1780.000	42.000	9.000	734118.000	maize
2	id607	name607	southeast	1820.000	29.000	6.000	211605.000	maize
3	id608	name608	southeast	1640.000	108.0...	7.000	1167040.0...	maize
4	id609	name609	southwest	1600.000	101.0...	5.000	756755.000	wheat
5	id615	name615	midlands	920.000	86.000	6.000	442554.000	potatoes
6	id618	name618	southeast	1180.000	98.000	3.000	368646.000	maize
7	id619	name619	north	840.000	64.000	8.000	457552.000	potatoes

## Especificación de un método de fusión y claves

**Figura 3-16**  
*Uso de la pestaña Fundir para configurar opciones del método de fusión*



**Método de fusión.** Seleccione Orden o Claves para especificar el método de fusión de registros. Si selecciona Claves, se activa la mitad inferior del cuadro de diálogo.

- **Ordenar.** Fusiona los registros de tal manera que el *enésimo* registro de cada entrada se fusione para producir el *enésimo* registro de salida. Cuando algún registro se queda sin registro de entrada correspondiente, no se producen más registros de salida. Esto quiere decir que el número de registros creados equivale al de registros en el conjunto de datos de menor tamaño.
- **Claves.** Usa un campo clave, como el *ID de transacción*, para fusionar registros que compartan el mismo valor en este campo. Equivale a una “equi-uniión” de base de datos. Si un valor clave tiene lugar más de una vez, se devuelven todas las combinaciones posibles. Por ejemplo, si registros con el mismo valor de campo clave *A* contienen valores diferentes *B*, *C* y *D* en otros campos, los campos fusionados producirán un registro independiente para cada combinación de *A* con el valor *B*, *A* con el valor *C* y *A* con el valor *D*.

*Nota:* los valores nulos no se consideran idénticos en el método de fusión por claves, por lo que no se unen.

- **Condición.** Utilice esta opción para especificar una condición para la fusión. Si desea obtener más información, consulte el tema [Especificación de condiciones para una fusión](#) el p. 94.

**Claves posibles.** Sólo enumera los campos con nombres de campo que coincidan exactamente en todos los orígenes de datos de entrada. Seleccione un campo de esta lista y use el botón de flecha para añadirlo como campo clave utilizado para la fusión de registros. Es posible usar más de un campo clave. Puede cambiar el nombre de los campos de entrada no coincidentes mediante un nodo Filtro o la pestaña Filtro de un nodo de origen.

**Claves para fusión.** Enumera todos los campos usados para fusionar registros desde todos los orígenes de datos de entrada basándose en valores de los campos clave. Para quitar una clave de la lista, selecciónela y use el botón de flecha para devolverla a la lista Claves posibles. Cuando se seleccionan varios campos clave, la opción de la parte inferior se desactiva.

**Combinar campos clave duplicados.** Cuando se seleccionan arriba varios campos clave, esta opción garantiza que sólo haya un campo de salida con dicho nombre. Esta opción está habilitada por defecto, salvo en el caso en que las rutas se hayan importado de versiones anteriores de IBM® SPSS® Modeler. Cuando la opción está desactivada, es preciso cambiar el nombre de los campos clave duplicados o excluirlos mediante la pestaña Filtro, en el cuadro de diálogo del nodo Fusionar.

**Incluir sólo registros coincidentes (uniión interior).** Seleccione esta opción para fusionar sólo registros completos.

**Incluir registros coincidentes y no coincidentes (uniión exterior completa).** Seleccione esta opción para llevar a cabo una “uniión exterior completa”. Esto quiere decir que si los valores del campo clave no están presentes en todas las tablas de entrada, los registros incompletos se conservarán. El valor no definido (\$null\$) se añade al campo clave y se incluye en el registro de salida.

**Incluir registros coincidentes y no coincidentes seleccionados (uniión exterior parcial).** Seleccione esta opción para llevar a cabo una “uniión exterior parcial” de las tablas que seleccione en un cuadro de diálogo secundario. Pulse en Seleccionar para especificar tablas de las que conservar registros incompletos en la fusión.

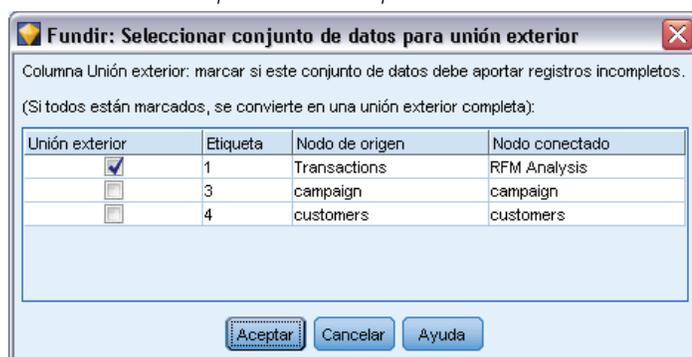
**Incluir registros del primer conjunto de datos que no coincida con ningún otro (anti-uniión).** Seleccione esta opción para llevar a cabo un tipo de “anti-uniión” donde sólo se pasen a la parte de abajo los registros no coincidentes del primer conjunto de datos. Puede especificar el orden de los conjuntos

de datos de entrada mediante las flechas de la pestaña Entradas. Este tipo de unión no incluye registros completos en el conjunto de datos de salida. Si desea obtener más información, consulte el tema [Tipos de uniones](#) el p. 90.

### **Selección de datos para uniones parciales**

Para llevar a cabo una unión exterior parcial, debe seleccionar las tablas para las que se deben retener registros incompletos. Por ejemplo, puede que desee conservar todos los registros de una tabla de clientes y sólo los registros de la tabla de préstamos hipotecarios.

Figura 3-17  
Selección de datos para una unión parcial o exterior

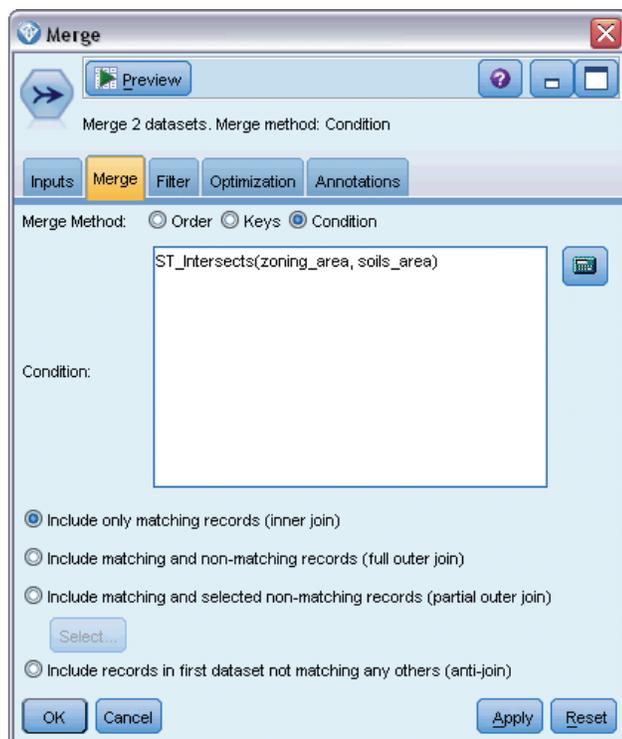


**Columna Unión exterior.** En la columna *Unión exterior*, seleccione los conjuntos de datos que desee incluir íntegramente. Si desea realizar una unión parcial, se conservarán los registros superpuestos, así como los registros incompletos de los conjuntos de datos aquí seleccionados. Si desea obtener más información, consulte el tema [Tipos de uniones](#) el p. 90.

### **Especificación de condiciones para una fusión**

Al establecer el método de fusión en Condición, se pueden especificar una o más condiciones que se deben cumplir para que se pueda realizar la fusión.

Figura 3-18  
Establecimiento de condiciones para una fusión

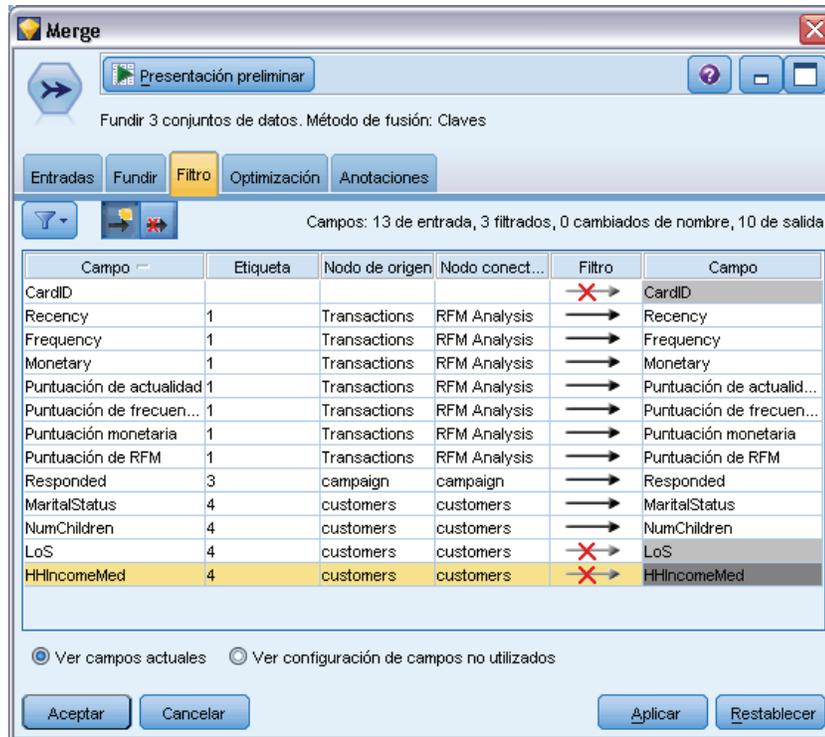


Se pueden introducir las condiciones directamente en el campo Condición, o crearlas con la ayuda del generador de expresiones, pulsando en el símbolo de calculadora a la derecha del campo.

### ***Filtrados de campos del nodo Fundir***

Los nodos Fundir incorporan una forma cómoda de filtrar o cambiar el nombre a los campos resultantes de una fusión con varios orígenes de datos. Pulse en la pestaña Filtro del cuadro de diálogo para seleccionar las opciones de filtrado.

Figura 3-19  
Filtrado desde el nodo Fundir



Las opciones que aquí se presentan son casi idénticas a las del nodo Filtrar. Existen, por tanto, opciones adicionales que aquí no se detallan y que están disponibles en el nodo Filtrar. Si desea obtener más información, consulte el tema [Cambio de nombre o filtrado de campos](#) en el capítulo 4 el p. 154.

**Campo.** Muestra los campos de entrada de los orígenes de datos conectados actualmente.

**Etiqueta.** Enumera los nombres (o números) de etiqueta asociados con el enlace del origen de datos. Pulse en la pestaña Entradas para modificar los enlaces activos a este nodo Fundir.

**Nodo de origen.** Muestra el nodo de origen cuyos datos se están fusionando.

**Nodo conectado.** Muestra el nombre del nodo que está conectado al nodo Fundir. Con frecuencia, la minería de datos complejos requiere la realización de varias fusiones o de operaciones adicionales que pueden incluir el mismo nodo de origen. El nombre del nodo conectado proporciona una forma de diferenciarlas.

**Filtro.** Muestra las conexiones actuales entre el campo de entrada y salida. Las conexiones activas muestran una flecha no interrumpida. Las conexiones con una X roja indican los campos filtrados.

**Campo.** Enumera los campos de salida tras la fusión o adición. Los campos duplicados se muestran en rojo. Pulse en el campo Filtro para desactivar los campos duplicados.

**Ver campos actuales.** Seleccione esta opción para ver información sobre los campos seleccionados para ser campos clave.

**Ver configuración de campos no utilizados.** Seleccione esta opción para ver información sobre los campos que no están en uso actualmente.

## Configuración del orden de entrada y etiquetado

El uso de la pestaña Entradas de los cuadros de diálogo de los nodos Fundir y Añadir permite especificar el orden de los orígenes de datos de entrada y efectuar modificaciones a los nombres de etiqueta de cada origen.

Figura 3-20

Uso de la pestaña Entradas para especificar etiquetas y orden de entrada



**Etiquetas y orden de los conjuntos de datos de entrada.** Seleccione esta opción para fusionar o añadir sólo registros completos.

- Etiqueta.** Enumera los nombres de etiqueta actuales de cada origen de datos de entrada. Los nombres de etiqueta, o **etiquetas**, constituyen un modo exclusivo de identificar los enlaces de los datos en la operación de fusionar o añadir. Por ejemplo, imagine una masa de agua procedente de varias tuberías que se encuentran en un punto para continuar por una sola tubería. Los datos fluyen del mismo modo en IBM® SPSS® Modeler; el punto de fusión es a veces una compleja interacción entre varios orígenes de datos. Las etiquetas ofrecen una forma de administrar las entradas (“tuberías”) en un nodo Fundir o Añadir de tal modo que, si éste se guarda o desconecta, los enlaces permanecen y son fácilmente identificables.

Cuando conecta datos adicionales en un nodo Fundir o Añadir, se crean automáticamente etiquetas por defecto a partir de números que representan el orden en que los nodos se conectaron. Este orden no está relacionado con el orden de los campos en los conjuntos de datos de entrada o

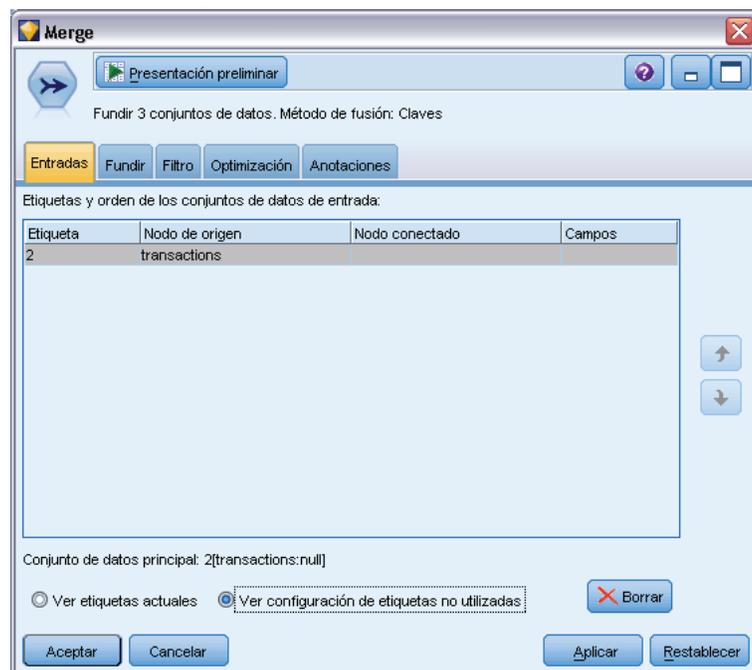
salida. Se puede cambiar el orden por defecto de las etiquetas especificando un nuevo nombre en la columna *Etiqueta*.

- **Nodo de origen.** Muestra el nodo de origen cuyos datos se están combinando.
- **Nodo conectado.** Muestra el nombre del nodo que está conectado al nodo Fundir o Añadir. Con frecuencia, la minería de datos complejos requiere la realización de varias operaciones de fusión que pueden incluir el mismo nodo de origen. El nombre del nodo conectado proporciona una forma de diferenciarlas.
- **Campos.** Detalla el número de campos de cada origen de datos.

**Ver etiquetas actuales.** Seleccione esta opción para ver las etiquetas que están usando de forma activa los nodos Fundir o Añadir. En otras palabras, las etiquetas actuales identifican enlaces al nodo que contienen los datos que fluyen. Volviendo a la metáfora de las tuberías, las etiquetas actuales son las tuberías con flujo de agua activo.

**Ver configuración de etiquetas no utilizadas.** Seleccione esta opción para ver etiquetas o enlaces previamente usados para la conexión con los nodos Fundir o Añadir pero que actualmente no están conectados a ningún origen de datos. Serían las tuberías vacías aún intactas de un sistema de fontanería. Puede optar por conectar estas “tuberías” a un nuevo origen o quitarlas. Para quitar etiquetas no utilizadas del nodo, pulse en Borrar. Al hacerlo, se borrarán todas las etiquetas no utilizadas de una vez.

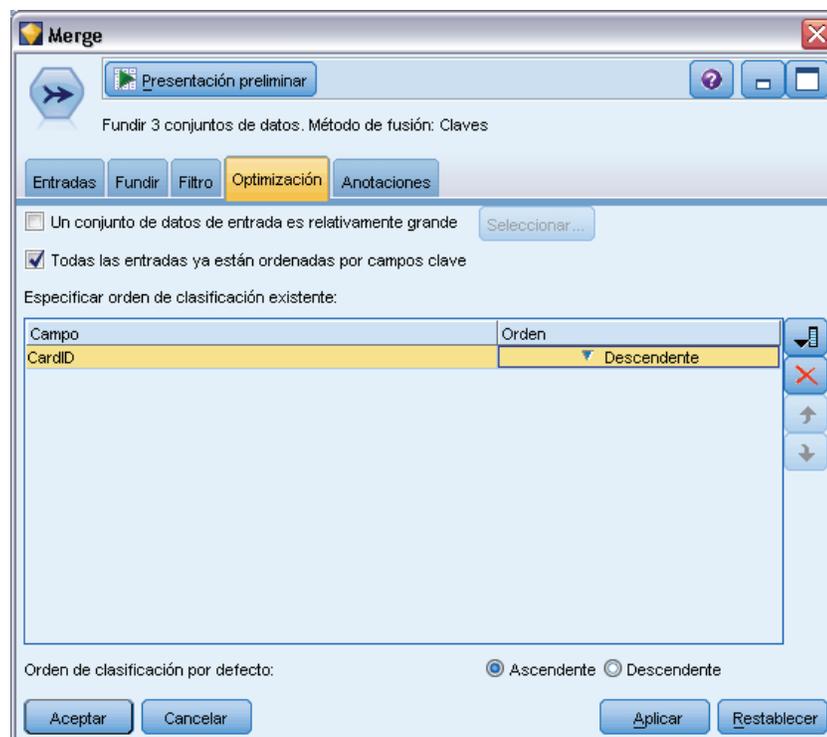
Figura 3-21  
Eliminación de etiquetas no utilizadas del nodo Fundir



## Configuración de optimización de fusión

El sistema ofrece dos opciones que pueden ayudar a fusionar datos de una manera más eficaz en ciertas situaciones. Estas opciones permiten optimizar la fusión cuando el conjunto de datos de entrada es significativamente mayor que los otros conjuntos de datos o cuando los datos ya se han ordenado por todos o algunos de los campos clave que utiliza para la fusión.

Figura 3-22  
Configuración de optimización



**Un conjunto de datos de entrada es relativamente grande.** Seleccione esta opción para indicar que uno de los conjuntos de datos de entrada es mucho mayor que los otros. El sistema almacenará en la caché los conjuntos de datos de menor tamaño de la memoria y, a continuación, realizará la fusión al procesar el conjunto de datos grande sin almacenarla en caché ni ordenarla. Este tipo de unión se suele utilizar con los datos que se diseñan mediante un esquema de estrella o de diseño similar, donde existe una tabla central de gran tamaño de datos compartidos (por ejemplo, en los datos transaccionales). Si selecciona esta opción, pulse en *Seleccionar* para especificar el conjunto de datos grande. Tenga en cuenta que sólo puede seleccionar *un* conjunto de datos grande. La siguiente tabla resume las uniones que se pueden optimizar con este método.

Tipo de unión	¿Se puede optimizar para un conjunto de datos de entrada grande?
Interior	Sí
Parcial	Sí, si no hay ningún registro incompleto en el conjunto de datos grande.
Completa	No
Anti-unión	Sí, si el conjunto de datos grande es la primera entrada.

**Todas las entradas ya están ordenadas por campos clave.** Seleccione esta opción para indicar que los datos de entrada ya están ordenados por uno o más de los campos clave que está utilizando para la fusión. Asegúrese de que *todos* los conjuntos de datos de entrada están ordenados.

**Especificar orden de clasificación existente.** Determina los campos que ya están ordenados. Puede añadir campos a la lista mediante el cuadro de diálogo Seleccionar campos. Sólo puede realizar la selección de los campos clave que se utilizan para la fusión (especificados en la pestaña Fundir). En la columna *Orden*, especifique si cada campo se ordena en orden ascendente o descendente. Si especifica varios campos, asegúrese de que los enumera en el orden de clasificación correcto. Utilice las flechas de la derecha de la lista para organizar los campos en el orden correcto. Si comete un error al especificar el orden de clasificación existente, aparecerá un error cuando ejecute la ruta que muestra el número de registro donde la clasificación no coincide con la que especificó.

Dependiendo de la sensibilidad de caracteres del método de ordenación utilizado por la base de datos, puede que la optimización no funcione correctamente cuando la base de datos ordene una o más entradas. Por ejemplo, si tiene dos entradas de las que una distingue entre mayúsculas y minúsculas y la otra no, los resultados de la ordenación podrían ser diferentes. La optimización de fusión hace que los registros se procesen mediante su ordenación. Como resultado, si las entradas se ordena utilizando métodos de ordenación diferentes, el nodo Fundir indica un error y muestra el número de registro cuando la ordenación no es coherente. Cuando todas las entradas son de un origen o se ordenan mediante ordenaciones inclusivas mutuamente, los registros se pueden combinar correctamente.

*Nota:* La velocidad de fusión puede mejorarse al activar el procesamiento paralelo.

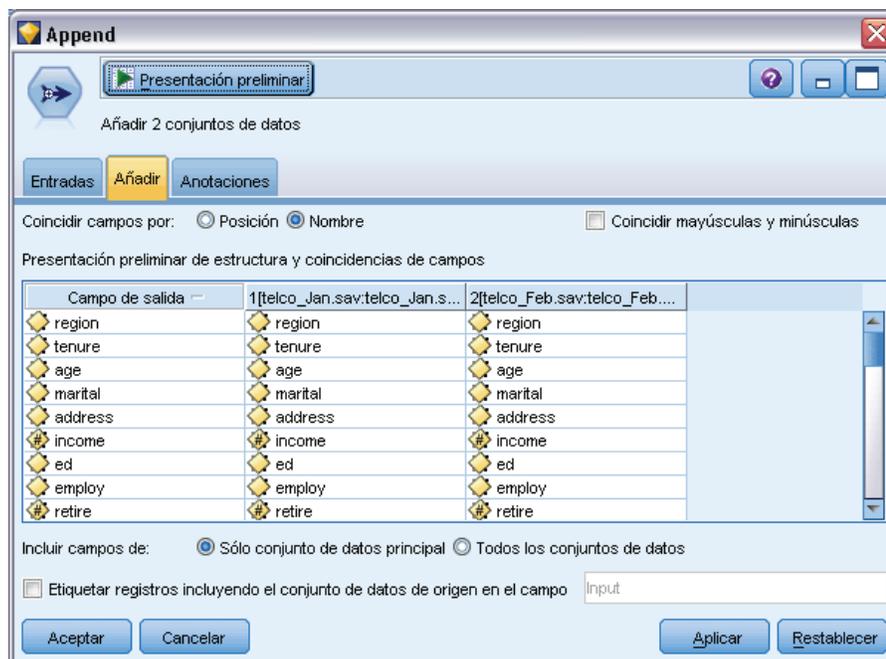
## ***Nodo Añadir***

Los nodos Añadir se pueden usar para concatenar conjuntos de registros. A diferencia con los nodos Fundir, que une registros de diferentes orígenes, los nodos Añadir leen y pasan a la parte de abajo todos los registros de un único origen hasta que no quede ninguno. A continuación, los registros procedentes del siguiente origen se leen usando la misma estructura de datos (número de registros, número de campos, etc.) que la entrada primera (o primaria). Cuando el origen primario contiene más campos que otro registro de entrada, se usa la cadena de valor nulo del sistema (\$null\$) para los valores incompletos.

Los nodos Añadir son útiles para combinar conjuntos de datos con estructuras similares pero datos diferentes. Por ejemplo, podría tener datos de transacción almacenados en diferentes archivos para diferentes períodos (un archivo de datos de venta para el mes de marzo y otro para el mes de abril, por ejemplo). Suponiendo que tengan la misma estructura (los mismos campos en el mismo orden), el nodo Añadir los une en un archivo de gran tamaño que se puede analizar.

*Nota:* Con el fin de añadir los archivos, los niveles de medición de campo deben coincidir. Por ejemplo, un campo *Nominal* no puede añadirse a un campo cuyo nivel de medición sea *Continuo*.

Figura 3-23  
Cuadro de diálogo del nodo Añadir con una correspondencia de campos por nombre.



## Opciones de configuración de Añadir

**Coincidir campos por.** Seleccione un método que utilizar para asignar los campos que añadir.

- **Posición.** Seleccione esta opción para añadir los conjuntos de datos basándose en la posición que tienen los campos en el origen de datos principal. Cuando use este método, los datos deben estar ordenados para garantizar una correcta adición.
- **Nombre.** Seleccione esta opción para añadir los conjuntos de datos basándose en el nombre que tienen los campos en los conjuntos de datos de entrada. Seleccione también Coincidir mayúsculas y minúsculas para activar la sensibilidad de caracteres en la coincidencia de nombres de campos.

**Campo de salida.** Enumera los nodos de origen conectados al nodo Añadir. El primero de la lista es el origen de entrada primario. Los campos de la lista se pueden ordenar pulsando en el encabezado de columna. Esta organización no ordena realmente los campos del conjunto de datos.

**Incluir campos de.** Seleccione Sólo conjunto de datos principal para producir campos de salida basándose en los campos del conjunto de datos principal. El conjunto de datos principal es la primera entrada, especificada en la pestaña Entradas. Seleccione Todos los conjuntos de datos para producir campos de salida para todos los campos en todos los conjuntos de datos, independientemente de si existe un campo coincidente en todos los conjuntos de datos de entrada.

**Etiquetar registros incluyendo el conjunto de datos de origen en el campo.** Seleccione esta opción para añadir un campo de información adicional al archivo de resultados cuyos valores indican el conjunto de datos de origen de cada registro. Especifique un nombre en el cuadro de texto. El nombre por defecto es *Entrada*.

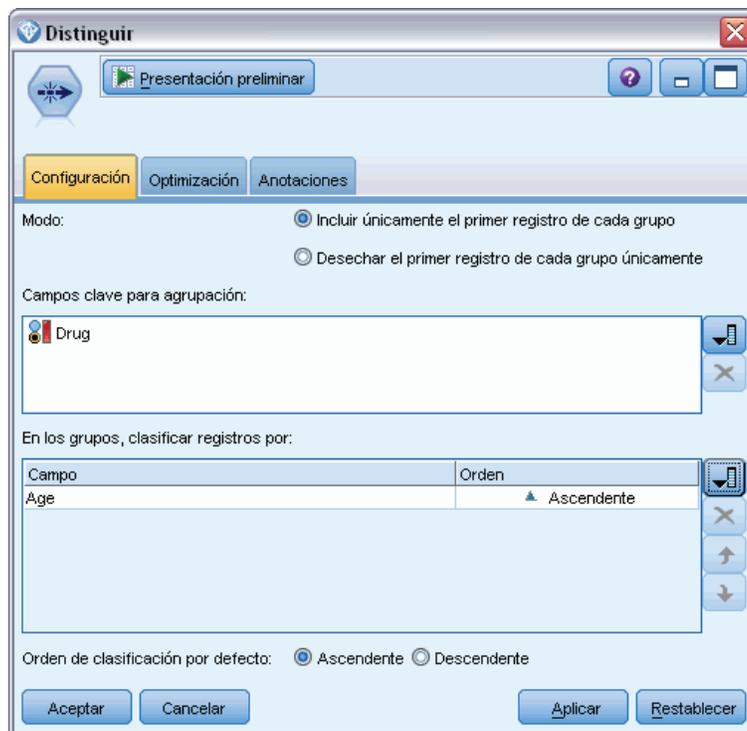
## Nodo Distinguir

Los registros duplicados en una base de datos deben eliminarse antes de comenzar con la minería de datos. Por ejemplo, en una base de datos de marketing, las personas pueden aparecer muchas veces con diferentes direcciones o diferente información de empresas. Puede utilizar el nodo Distinguir para encontrar, eliminar o duplicar registros en su conjunto de datos.

Utilizando el nodo Distinguir, puede eliminar registros duplicados pasando el primero de los registros distintos a la ruta de datos o encontrar registros duplicados descartando el primer registro y pasando cualquier duplicado a la ruta de datos en su lugar.

Además, puede definir un orden de clasificación en cada valor clave distinto para los resultados devueltos. Si desea que se devuelva una fila concreta para cada clave distinta, debe clasificar los registros en el nodo Distinguir en lugar de utilizar un nodo Clasificar anterior (consulte “Ordenación de registros dentro del nodo Distinguir” posterior).

Figura 3-24  
Cuadro de diálogo del nodo Distinguir



**Moda.** Especifique si desea incluir o excluir (descartar) el primer registro.

- **Incluir sólo el primer registro en cada grupo.** Incluye el primer registro distinto en la ruta de datos y elimina cualquier duplicado.
- **Descartar sólo el primer registro en cada grupo.** Descarta el primer registro distinto detectado y pasa cualquier registro duplicado a la ruta de datos en su lugar. Esta opción resulta útil para *buscar* duplicados en los datos con el fin de examinarlos posteriormente en la ruta.

**Campos clave para el agrupamiento.** Enumera el campo o campos utilizados para determinar si los registros son idénticos. Tiene la posibilidad de:

- Para añadir campos a esta lista, utilice el botón selector de campos de la parte derecha.
- Elimine campos de la lista utilizando el botón X rojo (eliminar).

**En grupos, ordenar registros por.** Enumera los campos utilizados para especificar el modo en que se ordenan los registros dentro de cada valor clave distinto, y si se ordenan de forma ascendente o descendente. Tiene la posibilidad de:

- Para añadir campos a esta lista, utilice el botón selector de campos de la parte derecha.
- Elimine campos de la lista utilizando el botón X rojo (eliminar).
- Mueva campos mediante los botones de arriba o abajo si está ordenando por más de un campo.

**Orden de clasificación por defecto.** Especifique si, por defecto, los registros se ordenan de forma Ascendente o Descendente.

### ***Ordenación de registros dentro del nodo Distinguir***

Utilizando el En grupos, ordenar registros por del nodo Distinguir, puede devolver a una fila específica para cada clave diferente y no tendrá la necesidad de utilizar un nodo Ordenar anterior. Por ejemplo, supongamos que tenemos los siguientes datos sobre las edades de los usuarios de medicamentos con receta.

<b>Edad</b>	<b>Medicamento</b>
50	Medicamento A
71	Medicamento B
44	Medicamento A
65	Medicamento X
39	Medicamento A
75	Medicamento C
72	Medicamento Y
57	Medicamento X
79	Medicamento Y
69	Medicamento C
74	Medicamento B
85	Medicamento Y
69	Medicamento X

Para encontrar el usuario de más edad de cada medicamento, defina el modo para incluir sólo el primer registro en cada grupo, utilice Medicamento como campo clave y Edad como campo de clasificación en orden descendente. El orden de entrada no afecta al resultado porque la selección de clasificación especifica cuáles de las diversas filas de un Medicamento concreto se van a devolver y el resultado de datos final sería el siguiente:

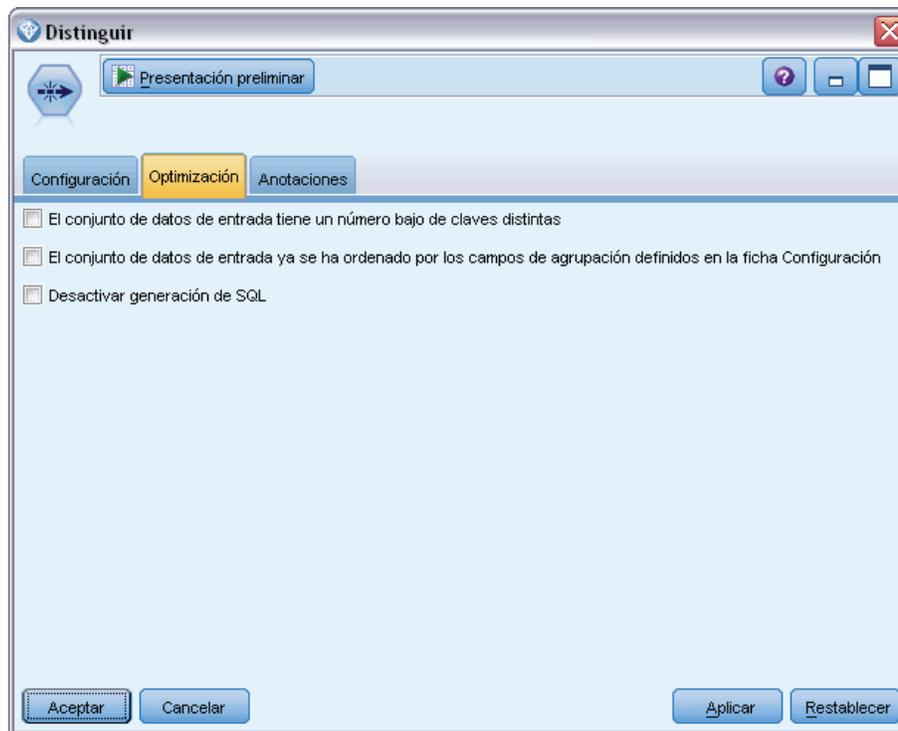
<b>Edad</b>	<b>Medicamento</b>
50	Medicamento A
74	Medicamento B
75	Medicamento C
69	Medicamento X
85	Medicamento Y

### ***Configuración de optimización de Distinguir***

Si los datos en los que trabaja sólo tienen un pequeño número de registros o ya se han ordenado, puede optimizar la forma en la que se gestionan para que IBM® SPSS® Modeler procese los datos de forma más eficiente.

*Nota:* Si selecciona El conjunto de datos de entrada tiene un pequeño número de teclas diferentes o la generación SQL para el nodo, se puede devolver cualquier fila dentro del valor de clave distinta; para controlar la fila que se devuelve, tiene que especificar el orden de clasificación mediante En grupos, ordenar registros por de la pestaña Configuración. Las opciones de optimización no afectan a los resultados ofrecidos por el nodo Distinguir siempre que haya especificado un orden de clasificación en la pestaña Configuración.

Figura 3-25  
Configuración de optimización



**El conjunto de datos de entrada tiene un pequeño número de teclas diferentes.** Seleccione esta opción si sólo tiene un pequeño número de registros, un pequeño número de valores únicos del campo(s) clave o ambos. Con ello puede mejorar el rendimiento.

**El conjunto de datos de entrada ya está ordenado por los campos de agrupación y los campos de clasificación en la pestaña Configuración.** Seleccione esta opción únicamente si los datos ya están ordenados según los campos enumerados en En grupos, ordenar registros por en la pestaña Configuración, y si el orden de clasificación ascendente o descendente de los datos es el mismo. Con ello puede mejorar el rendimiento.

**Desactivar generación de SQL.** Seleccione esta opción para desactivar la generación de SQL para el nodo.

# Nodos de operaciones con campos

## Conceptos básicos de las operaciones con campos

Una vez que haya realizado una exploración de datos inicial, es posible que tenga que seleccionar, limpiar o construir datos para preparar el análisis. La paleta Operaciones con campos contiene muchos nodos útiles para esta transformación y preparación.

Por ejemplo, con un nodo Derivar se puede crear un atributo que no se representa en los datos en la actualidad. También se puede utilizar un nodo Intervalos para recodificar automáticamente valores de campos en análisis objetivos. Con frecuencia utilizará los nodos Tipo, ya que permiten asignar valores, un nivel de medición y un papel de modelado a cada campo del conjunto de datos. Estas operaciones son útiles para gestionar valores perdidos y modelado posterior en la ruta.

La paleta Operaciones con campos contiene los siguientes nodos:



El nodo de preparación automática de datos (ADP) puede analizar sus datos e identificar los valores fijos, filtrar los campos problemáticos o que no serán útiles y derivar nuevos atributos cuando sea necesario y mejorar el rendimiento mediante técnicas de filtrado y muestreo inteligente. Puede utilizar el nodo de forma totalmente automática, permitiendo que el nodo seleccione y aplique valores fijos, o bien puede tener una vista previa de los cambios antes de que se apliquen y aceptarlos o rechazarlos. Si desea obtener más información, consulte el tema [Preparación automática de datos](#) el p. 108.



El nodo Tipo especifica propiedades y metadatos de campo. Por ejemplo, puede especificar un nivel de medición (continuo, nominal, ordinal o marca) para cada campo, establecer las opciones para gestionar valores perdidos y nulos del sistema, establecer el papel de un campo con fines de modelado, especificar las etiquetas de valor y campo y especificar los valores de un campo. Si desea obtener más información, consulte el tema [Nodo Tipo](#) el p. 135.



El nodo Filtro filtra (descarta) campos, vuelve a nombrar campos y establece asociaciones de un nodo de origen a otro. Si desea obtener más información, consulte el tema [Cambio de nombre o filtrado de campos](#) el p. 154.



El nodo Derivar modifica los valores de datos o crea campos nuevos desde uno o más campos existentes. Crea campos del tipo fórmula, marca, nominal, estado, recuento y condicional. Si desea obtener más información, consulte el tema [Nodo Derivar](#) el p. 165.



El nodo Conjunto combina dos o más nugget de modelo para obtener pronósticos más precisos que pueden conseguirse de cualquier modelo. Si desea obtener más información, consulte el tema [Nodo Conjunto](#) el p. 161.



El nodo Rellenar sustituye valores de campos y cambia el almacenamiento. Puede sustituir los valores en función de una condición CLEM, como `@BLANK(@FIELD)`. También puede sustituir todos los espacios vacíos o valores nulos por un valor específico. Un nodo Rellenar suelen utilizarse junto con un nodo Tipo para sustituir valores perdidos. Si desea obtener más información, consulte el tema [Nodo Rellenar](#) el p. 177.



El nodo Anonimizar transforma la manera en que se representan los nombres y los valores de los campos a partir de ese punto de la ruta, lo que permite disfrazar los datos originales. Puede resultar útil si desea permitir que otros usuarios generen modelos utilizando datos confidenciales, como los nombres de los clientes u otros detalles. Si desea obtener más información, consulte el tema [Nodo Anonimizar](#) el p. 180.



El nodo Reclasificar transforma un conjunto de valores categóricos en otro. La reclasificación es útil para contraer categorías o reagrupar datos para su análisis. Si desea obtener más información, consulte el tema [Nodo Reclasificar](#) el p. 185.



El nodo Intervalos crea automáticamente nuevos campos nominales (conjunto) en función de los valores de uno o más campos continuos (rango numérico) existentes. Por ejemplo, puede transformar un campo de ingresos continuo en un campo categórico nuevo que contenga grupos de ingresos como desviaciones desde la media. Una vez creados los intervalos para el campo nuevo, puede generar un nodo Derivar en función de los puntos de corte. Si desea obtener más información, consulte el tema [Nodo Intervalos](#) el p. 189.



El nodo Análisis de RFM (actualidad, frecuencia, monetario) permite determinar cuantitativamente qué clientes son los mejores examinando cuándo ha sido la compra más reciente de un cliente (actualidad), cuántas veces suelen comprar (frecuencia) y cuánto gasta el cliente en todas las transacciones (valor monetario). Si desea obtener más información, consulte el tema [Nodo Análisis de RFM](#) el p. 201.



El nodo Partición genera un campo de partición, que divide los datos en subconjuntos diferentes para las fases de entrenamiento, comprobación y validación en la generación del modelo. Si desea obtener más información, consulte el tema [Nodo Partición](#) el p. 205.



El nodo Marcas deriva varios campos de marcas en función de los valores categóricos definidos para uno o más campos nominales. Si desea obtener más información, consulte el tema [Nodo Marcas](#) el p. 207.



El nodo Reestructurar convierte un campo nominal o marca en un grupo de campos que se puede rellenar con los valores todavía de otro campo. Por ejemplo, para un campo determinado llamado *tipo de pago*, con valores de *crédito*, *efectivo*, y *débito*, se crearían tres campos nuevos (*crédito*, *efectivo*, *débito*), que contendría cada uno el valor del pago real realizado. Si desea obtener más información, consulte el tema [Nodo Reestructurar](#) el p. 209.



El nodo Transponer intercambia los datos en filas y columnas de manera que los registros se conviertan en campos y los campos en registros. Si desea obtener más información, consulte el tema [Nodo Transponer](#) el p. 211.



El nodo Intervalos de tiempo especifica los intervalos y genera etiquetas (si es necesario) para modelar los datos de series temporales. Si los valores no están espaciados de manera uniforme, el nodo puede rellenar o agregar valores según sea necesario para crear un intervalo uniforme entre registros. Si desea obtener más información, consulte el tema [Nodo de intervalos de tiempo](#) el p. 216.



El nodo Histórico se utiliza para crear campos nuevos que contienen datos de los campos de registros anteriores. Los nodos Histórico se suelen utilizar para los datos secuenciales, como los datos de series temporales. Antes de utilizar un nodo Histórico, puede desear ordenar los datos utilizando un nodo Ordenar. Si desea obtener más información, consulte el tema [Nodo Histórico](#) el p. 237.



El nodo Reorg. campos define el orden natural utilizado para mostrar los campos en la parte posterior de la ruta. Este orden afecta a la visualización de los campos en diversas ubicaciones, como las tablas, las listas y el selector de campos. Esta operación resulta útil al trabajar con conjuntos de datos amplios que hacen más visibles los campos de interés. Si desea obtener más información, consulte el tema [Nodo Reorg. campos](#) el p. 239.

Algunos de estos nodos pueden generarse directamente desde el informe de auditoría creado por el nodo Auditar datos. Si desea obtener más información, consulte el tema [Generación de otros nodos para la preparación de datos](#) en el capítulo 6 el p. 426.

## ***Preparación automática de datos***

La preparación de los datos para su análisis es uno de los pasos más importantes en cualquier proyecto y, tradicionalmente, uno de los que más tiempo requieren. Preparación automática de datos (ADP) controla las tareas automáticamente, analizando los datos e identificando problemas, filtrando campos problemáticos o sin posibilidades de ser útiles, derivando nuevos atributos cuando sea necesario y mejorando el rendimiento mediante técnicas de filtrado inteligente. Puede utilizar el algoritmo de una forma totalmente **automática**, permitiendo seleccionar y aplicar soluciones; o de forma **interactiva**, previendo los cambios antes de que se realicen y aceptarlos o rechazarlos según sea necesario.

ADP permite hacer que sus datos estén listos para la generación de modelos de forma rápida y fácil, sin necesidad de tener conocimientos previos de los conceptos previos implicados. Los modelos tienden a crearse y puntuarse con mayor rapidez; además, el uso de ADP mejora la solidez de los procesos de modelado automatizados, como actualización de modelos y análisis Champion/Challenger.

*Nota:* cuando el ADP prepara un campo para su análisis, crea un nuevo campo con los ajustes o transformaciones, en vez de reemplazar los valores y propiedades existentes del campo anterior. El campo anterior no se usa en más análisis, su papel se define como Ninguno.

**Ejemplo.** Una correduría de seguros con recursos limitados para investigar las reclamaciones de seguros de los asegurados desea crear un modelo para etiquetar las reclamaciones sospechosas y potencialmente fraudulentas. Antes de construir el modelo, leerán los datos para el modelado mediante la preparación automática de datos. Como desean revisar las transformaciones propuestas antes de que se apliquen las transformaciones, utilizarán la preparación automática de datos en modo interactivo.

Un grupo del sector del automóvil desea realizar un seguimiento de las ventas de diversos vehículos a motor. Para poder identificar los modelos como mejor y peor rendimiento, desean establecer una relación entre las ventas de vehículos y las características de los vehículos. Utilizarán la preparación automática de datos para preparar los datos para el análisis y crearán modelos utilizando la preparación “anterior” y “posterior” de datos para ver cómo difieren los resultados.

**Figura 4-1**  
*Pestaña Objetivo de Preparación automática de datos*

La preparación automática de datos puede recomendar pasos de preparación de datos que acelerarán la creación de modelos y mejorarán el poder predictivo. Puede incluir la transformación, creación y selección de características. El objetivo también se puede transformar.

¿Cuál es su objetivo?

**Equilibrar velocidad y precisión**  
Transformar los datos poniendo énfasis en la construcción de modelos con un equilibrio entre velocidad y precisión.

**Optimizar para velocidad**  
Transformar los datos poniendo énfasis en la construcción de modelos tan rápido como sea posible.

**Optimizar para precisión**  
Transformar los datos poniendo énfasis en la construcción de modelos con la mayor potencia predictiva.

**Análisis personalizado**  
Seleccione esta opción para afinar el algoritmo de la ficha Configuración.

**¿Cuál es su objetivo?** Preparación automática de datos recomienda ejecutar pasos para la preparación de datos que afectan a la velocidad con la que el resto de algoritmos pueden generar modelos y mejorar el potencial predictivo de esos modelos. Pueden incluir la transformación, construcción y selección de característica. El destino también puede transformarse. Puede especificar las prioridades de generación de modelos en las que se deben centrar el proceso de preparación de datos.

- **Equilibrar velocidad y precisión.** Esta opción prepara los datos para dar igual prioridad a la velocidad con la que se procesan los datos por algoritmos de creación de modelos y la precisión de los pronósticos.
- **Optimizar velocidad.** Esta opción prepara los datos para dar prioridad a la velocidad con la que se procesan los datos por los algoritmos de construcción de modelos. Si trabaja con conjuntos de datos muy grandes o busca una respuesta rápida, seleccione esta opción.
- **Optimizar precisión.** Esta opción prepara los datos para dar prioridad a la precisión de los pronósticos producidas por los algoritmos de construcción de modelos.
- **Análisis personalizado.** Seleccione esta opción si desea cambiar manualmente el algoritmo de la pestaña Configuración. Tenga en cuenta que esta configuración se selecciona automáticamente si realiza cambios posteriores a muchas opciones de la pestaña Configuración que sean incompatibles con los de otros objetivos.

### **Formación del nodo**

El nodo ADP se implementa como nodo de proceso y funciona de forma parecida al nodo Tipo; el **entrenamiento** del nodo ADP se corresponde con la creación de una instancia en el nodo Tipo. Una vez realizado el análisis, las transformaciones especificadas se aplican a los datos sin realizar más análisis siempre que el modelo de datos anteriores no cambie. Al igual que los nodos Tipo y Filtro, si el nodo ADP está desconectado recuerda el modelo de datos y las transformaciones,

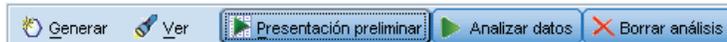
de modo que si vuelve a conectarse no hay que volver a entrenarlo; esto le permite entrenarlo en un subconjunto de datos normales y a continuación copiarlo o distribuirlo para utilizarlo en datos activos con la frecuencia necesaria.

### **Uso de la barra de herramientas**

La barra de herramientas le permite ejecutar y actualizar la visualización del análisis de datos y generar nodos que puede utilizar junto con los datos originales.

Figura 4-2

*Preparación automática de datos: barra de herramientas*



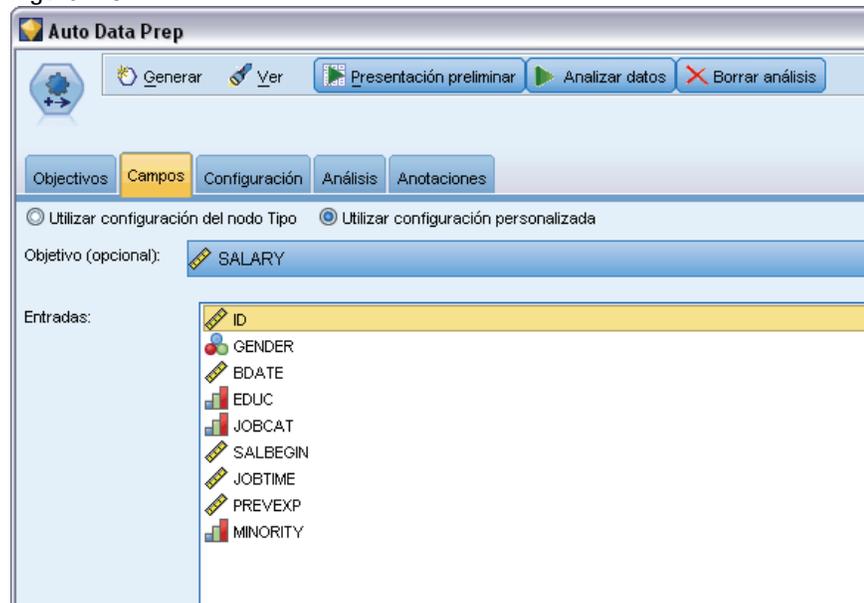
- **Generar** Desde este menú puede generar un nodo Filtro o Derivar. Tenga en cuenta que este menú sólo está disponible cuando se muestra un análisis en la pestaña Análisis.  
El nodo Filtro elimina los campos de entrada transformados. Si configura el nodo ADP para que deje los campos de entrada originales en el conjunto de datos, se restaurará el conjunto original de entradas que le permitirán interpretar el campo de puntuación en cuanto a las entradas. Por ejemplo, puede ser de gran utilidad si desea producir un gráfico del campo de puntuación con respecto a diversas entradas.  
El nodo Derivar puede restaurar el conjunto de datos original y las unidades destino. Sólo puede generar un nodo Derivar si el nodo ADP contiene un análisis que vuelve a trazar un destino de rango (es decir, el cambio de escala Box-Cox está seleccionado en el panel Preparar entradas y objetivo). No puede generar un nodo Derivar si el destino no es un rango o si el cambio de escala Box-Cox no está seleccionado. Si desea obtener más información, consulte el tema [Generación de un nodo Derivar](#) el p. 133.
- **Ver** Contiene opciones que controlan lo que se muestra en la pestaña Análisis. Se incluyen los controles de edición de gráficos y las selecciones de visualización para el panel principal y las vistas vinculadas.
- **Presentación preliminar** Muestra las transformaciones que se aplicarán a los datos de entrada.
- **Analizar datos** Inicia un análisis mediante la configuración actual y muestra los resultados en la pestaña Análisis.
- **Borrar análisis** Elimina el análisis existente (sólo está disponible cuando existe un análisis actual).

### **Estado del nodo**

El estado del nodo ADP en el lienzo de IBM® SPSS® Modeler se indica mediante una flecha o marca en el icono que muestra si el análisis se ha realizado o no.

## Pestaña Campos

Figura 4-3



Para generar un modelo, primero se deben especificar los campos que se desea usar como objetivos y como entradas. Salvo algunas excepciones, todos los nodos de modelado usarán la información de los campos procedente de un nodo Tipo anterior en la ruta. Si utiliza un nodo Tipo para seleccionar campos de entrada y objetivo, no es necesario cambiar nada en esta pestaña.

**Utilizar configuración del nodo Tipo.** Esta opción permite indicar al nodo que use la información de campo de un nodo Tipo situado en un punto anterior de la ruta. Este es el método por defecto.

**Utilizar configuración personalizada.** Esta opción permite indicar al nodo que use la información de campo especificada aquí en lugar de la proporcionada en nodos Tipo situados en cualquier punto anterior de la ruta. Después de seleccionar esta opción, especifique los campos siguientes si es necesario.

**Destino.** En los modelos que requieran uno o varios campos objetivo, selecciónelos. Se trata de una acción similar a establecer el papel del campo en *Objetivo* en un nodo Tipo.

**Entradas.** Seleccione el campo(s) de entrada. Se trata de una acción similar a establecer el papel del campo en *Entrada* en un nodo Tipo.

## Pestaña Configuración

La pestaña Configuración contiene diferentes grupos de ajustes que puede modificar para ajustar con precisión la forma en que el algoritmo procesa sus datos. Si realiza algún cambio en la configuración por defecto que sea incompatible con el resto de objetivos, la pestaña Objetivo se actualiza automáticamente para seleccionar la opción Personalizar análisis.

## Configuración de campos

Figura 4-4

Preparación automática de datos - Configuración de campos

La configuración de campos no se verá afectada si cambia su objetivo.

Utilizar campo de frecuencia

Utilizar campo de ponderación

Cómo gestionar los campos excluidos del modelado:

Filtrar campos no utilizados

Definir la dirección de campos no utilizados a "Ninguna"

Si los campos entrantes no generan los análisis existentes:

Detener la ejecución y mantener el análisis existente

Borrar el análisis existente y analizar los nuevos datos

**Utilizar campo de frecuencia** Esta opción le permite seleccionar un campo como ponderación de frecuencia. Utilícela si los registros de sus datos de formación representan más de una unidad cada uno, por ejemplo, si está utilizando datos agregados. Los valores de campo deben corresponder al número de unidades que representa cada registro.

**Utilizar campo de ponderación** Esta opción le permite seleccionar un campo como ponderación de caso. Las ponderaciones de caso se utilizan para contabilizar las diferencias existentes en la varianza entre los niveles del campo de salida.

**Cómo gestionar los campos excluidos del modelado** Especifica lo que ocurre a los campos excluidos; puede optar por filtrarlos de los datos o simplemente definir su *Papel* como Ninguno.

**Si los campos entrantes no coinciden con el análisis existente** Especifique lo que ocurre si uno o más campos de entrada obligatorios faltan del conjunto de datos entrante cuando ejecuta un nodo ADP entrenado.

- **Detener la ejecución y mantener el análisis existente** Esto detiene el proceso de ejecución, conserva la información del análisis actual y muestra un error.
- **Borrar el análisis existente y analizar los nuevos datos** Esto borra el análisis existente, analiza los datos entrantes y aplica las transformaciones recomendadas a dichos datos.

## Preparar fechas y horas

Figura 4-5  
Preparación automática de datos: Configuración de fecha y hora

Preparar fechas y horas para el modelado

**Calcular duración**

Calcular el tiempo transcurrido hasta la fecha de referencia

Fecha de referencia

Fecha de hoy

Fecha fija

Fecha: 2009-05-22

Unidades de duración de fecha

Automática

Unidades fijas

Unidad: Meses

Calcular el tiempo transcurrido hasta la hora de referencia

Hora de referencia

Hora actual

Hora fija

Hora: 11:42:01

Unidades de duración de tiempo

Automática

Unidades fijas

Unidad: Horas

**Extraer elementos temporales cíclicos**

Extraer de fechas:

Año  Mes  Día

Extraer de horas:

Hora  Minuto  Segundo

Muchos algoritmos no pueden tratar directamente los detalles de fecha y hora; estas configuraciones permiten derivar nuevos datos de duración que pueden utilizarse como entradas de modelo de fechas y horas de sus datos existentes. Los campos que contienen las fechas y las horas se deben predefinir con los tipos de almacenamiento de fecha u hora. Los campos de fecha y hora originales no se recomiendan como entradas de modelo posteriores a la preparación automática de datos.

**Preparar fechas y horas para el modelado.** Si cancela la selección de esta opción se desactivan todos los demás controles de Preparar fechas y horas mientras se mantienen las selecciones.

**Calcular tiempo transcurrido hasta fecha de referencia.** Esto produce el número de años/meses/días desde una fecha de referencia para cada variable que contenga fechas.

- **Fecha de referencia.** Especifique la fecha desde la que se calculará la duración en lo relativo a la información de fecha de los datos de entrada. Si selecciona Fecha de hoy, la fecha actual del sistema se utilizará siempre que se ejecute el nodo ADP. Para utilizar una fecha específica, seleccione Fecha fija e introduzca la fecha obligatoria. La fecha actual se introduce automáticamente en el campo Fecha fija, con formato de fecha de ruta, cuando se crea el nodo por primera vez.
- **Unidades de duración de fecha.** Especifique si el nodo debería decidir automáticamente sobre la unidad de duraciones de fecha o establezca Unidades fijas como Años, Meses o Días.

**Calcular tiempo transcurrido hasta hora de referencia.** Esto produce el número de horas/minutos/segundos desde una hora de referencia para cada variable que contenga horas.

- **Hora de referencia.** Especifique la hora desde la que se calculará la duración en lo relativo a la información de hora de los datos de entrada. Si selecciona Hora actual, la hora actual del sistema se utilizará siempre que se ejecute el nodo ADP. Para utilizar una hora específica, seleccione Hora fija e introduzca los detalles obligatorios. La hora actual se introduce automáticamente en el campo Hora fija, con formato de fecha de ruta, cuando se crea el nodo por primera vez.
- **Unidades de duración de tiempo.** Especifique si el nodo debería decidir automáticamente sobre la unidad de duraciones de hora o establezca Unidades fijas como Horas, Minutos o Segundos.

**Extraer elementos temporales cíclicos.** Utilice esta configuración para dividir un único campo de fecha o de hora en uno o más campos. Por ejemplo, si selecciona las tres casillas de verificación de fecha, el campo de fecha de entrada “1954-05-23” se dividirá en tres campos: 1954, 5 y 23, cada uno con el sufijo definido en el panel Nombres de campos y el campo de fecha original se ignorará.

- **Extraer de fechas.** Para cualquier entrada de fecha, especifique si desea extraer años, meses, días o cualquier combinación.
- **Extraer de horas.** Para cualquier entrada de hora, especifique si desea extraer horas, minutos, segundos o cualquier combinación.

## Excluir campos

Figura 4-6  
Configuración de Excluir campos de preparación automática de datos

Los campos constantes siempre se excluirán.

Excluir campos de entrada de baja calidad

Excluir campos de entrada

Excluir campos con demasiados valores ausentes

Porcentaje máximo de valores ausentes:  %

Excluir campos nominales con demasiadas categorías únicas

Número máximo de categorías:

Excluir campos categóricos con demasiados valores en una única categoría

Porcentaje máximo en una única categoría:  %

Los datos de mala calidad pueden afectar a la precisión de sus predicciones; por lo tanto, puede especificar el nivel de calidad aceptable de las características de entrada. Todos los campos que no sean constantes o les falte el 100% de los valores se excluirán automáticamente.

**Excluir campos de entrada de baja calidad.** Si cancela la selección de esta opción se desactivan todos los demás controles de Excluir campos mientras se mantienen las selecciones.

**Excluir campos con demasiados valores perdidos.** Los campos con un porcentaje de valores perdidos mayor que el porcentaje especificado se eliminan de análisis posteriores. Especifique un valor superior o igual a 0 (que equivale a cancelar la selección de esta opción) y menor o igual a 100, aunque los campos que tienen valores que faltan se excluyen automáticamente. El valor por defecto es 50.

**Excluir campos nominales con demasiadas categorías únicas.** Los campos nominales con un número de categorías superior al especificado se eliminarán de análisis posteriores. Especifique un número entero positivo. El valor predeterminado es 100. Esto resulta útil para eliminar automáticamente campos que contengan información única de registros para el modelado, como ID, dirección o nombre.

**Excluir campos categóricos con demasiados valores en una única categoría.** Los campos nominales y ordinales con una categoría con un porcentaje de registros superior al especificado se eliminarán de análisis posteriores. Especifique un valor superior o igual a 0 (que equivale a cancelar la selección de esta opción) y menor o igual a 100, aunque los campos constantes se excluyen automáticamente. El valor por defecto es 95.

## Preparación de entradas y destinos

Ya que los datos nunca están en el estado perfecto para su procesamiento, puede que desee ajustar algunas configuraciones antes de ejecutar un análisis. Por ejemplo, esto puede incluir la eliminación de valores atípicos, la especificación de cómo tratar valores ausentes o el ajuste del tipo.

*Nota:* Si cambia los valores de este panel, la pestaña Objetivos se actualiza automáticamente para seleccionar la opción Análisis personalizado.

Figura 4-7

Preparación automática de datos: Configuración de entrada y destino

Preparar los campos de entrada y objetivo para el modelado

Ajustar el tipo y mejorar la calidad de los datos

Entradas Objetivo

- Ajustar el tipo de campos numéricos (ordinales y continuos)
- Reordenar campos nominales para tener la categoría en primer lugar y la mayor en último lugar
- Sustituir valores extremos en campos continuos (recomendado para campos de entrada si se utilizarán en una escala común)
- Campos continuos: sustituir valores ausentes con media
- Campos nominales: sustituir valores ausentes con modo
- Campos ordinales: sustituir valores ausentes con mediana

Número máximo de valores de campos ordinales:

Número mínimo de valores de campos ordinales:

Valor de corte atípico:  (desviaciones típicas)

Método de sustitución de valores atípicos:  Sustituir con valor de corte  Eliminar valor

Transformar campo continuo

- Poner todos los campos de entrada continuos en una escala común (muy recomendado si se ejecutará la función de construcción)

Método de cambio de escala:  Media final:  Desviación típica final:

- Cambiar la escala de objetivo continuo con transformación de Box-Cox para reducir la asimetría

Media final:  Desviación típica final:

**Preparar los campos de entrada y destino para el modelado.** Activa y desactiva todos los campos del panel.

**Ajustar el tipo y mejorar la calidad de los datos.** Para las entradas y el destino puede especificar varias transformaciones de datos independientemente; esto es así porque tal vez no desee cambiar los valores del destino. Por ejemplo, una predicción de los ingresos en dólares tiene más sentido que una predicción medida en  $\log(\text{dollars})$ . Además, si el destino tiene valores ausentes, no hay ganancias predictivas para rellenar valores ausentes, mientras que al rellenar valores ausentes de entradas se puede permitir que algunos algoritmos procesen información que de otro modo se perdería.

Las configuraciones adicionales para estas transformaciones, como el valor de corte atípico, son comunes para el destino y las entradas.

Puede seleccionar la siguiente configuración para las entradas, para el destino o para ambos:

- **Ajustar el tipo de campos numéricos.** Seleccione esta opción para determinar si los campos numéricos con un nivel de medición *Ordinal* puede convertirse a *Continuo* o viceversa. Puede especificar los valores de umbral mínimo y máximo para controlar la conversión.
- **Reordenar campos nominales.** Seleccione esta opción para ordenar los campos nominales (conjunto) de menor a mayor según su categoría.
- **Sustituir valores extremos en campos continuos.** Especifique si desea sustituir los valores atípicos; utilice esta opción junto con la opción Método de sustitución de valores atípicos que aparece a continuación.
- **Campos continuos: reemplazar los valores perdidos con la media.** Seleccione esta opción para sustituir los valores ausentes de características (rango) continuos.
- **Campos nominales: reemplazar los valores perdidos con el modo** Seleccione esta opción para sustituir los valores ausentes de características (conjunto) nominales.
- **Campos ordinales: reemplazar los valores perdidos con la mediana.** Seleccione esta opción para sustituir los valores ausentes de características (conjunto ordenado) ordinales.

**Número máximo de valores de campos ordinales.** Especifique el umbral para redefinir los campos ordinales (conjunto ordenado) como continuo (rango). El valor predeterminado es 10, por lo que si un campo ordinal tiene más de 10 categorías, se redefine como continuo (rango).

**Número mínimo de valores de campos continuos.** Especifique el umbral para redefinir los campos de escalas o continuo (rango) como ordinal (conjunto ordenado). El valor predeterminado es 5, por lo que si un campo continuo tiene menos de 5 valores, se redefine como ordinal (conjunto ordenado).

**Valor de corte atípico.** Especifique el criterio de corte atípico, medido en desviaciones típicas; el valor por defecto es 3.

**Método de sustitución de valores atípicos.** Seleccione si los valores atípicos deben sustituirse mediante el recorte (forzar) con el valor de corte o si deben eliminarse y establecerse como valores ausentes. Todos los valores atípicos establecidos como valores ausentes siguen la configuración de gestión de valores ausentes seleccionada anteriormente.

**Poner todos los campos de entrada continuos en una escala común.** Para normalizar los campos de entrada continuos, seleccione esta casilla de verificación y seleccione el método de normalización. El valor por defecto es Transformación de puntuaciones  $z$ , donde puede especificar la Media final, que tiene un valor por defecto de 0, y la Desviación típica final, que tiene un valor por defecto de 1. También puede decidir utilizar Transformación mínima/máxima y especificar los valores mínimo y máximo, que tienen valores por defecto de 0 y 100, respectivamente.

Este campo es especialmente útil cuando selecciona Realizar creación de características en el panel Crear y seleccionar características.

**Cambiar la escala de un destino continuo con una transformación de Box-Cox.** Para normalizar un campo destino continuo (escala o rango), seleccione esta casilla de verificación. La transformación de Box-Cox tiene valores por defecto de 0 para la Media final y de 1 para la Desviación típica final.

*Nota:* Si decide normalizar el destino, la dimensión del destino se transformará. En este caso, puede que necesite generar un nodo Derivar para aplicar una transformación inversa para volver a cambiar las unidades transformadas a un formato reconocible para procesarlo con posterioridad. Si desea obtener más información, consulte el tema [Generación de un nodo Derivar](#) el p. 133.

## Creación y selección de características

Para mejorar el poder predictivo de sus datos, puede transformar los campos de entrada o crear otros nuevos basados en los campos existentes.

*Nota:* Si cambia los valores de este panel, la pestaña Objetivos se actualiza automáticamente para seleccionar la opción Análisis personalizado.

Figura 4-8

Preparación automática de datos: Configuración de transformación, creación y selección

Transformar, crear y seleccionar campos de entrada para mejorar el poder predictivo

**Campos de entrada categóricos**

Combinar categorías dispersas para aumentar al máximo la asociación con el objetivo Valor de probabilidad: 0,05

Se excluirán los campos de entrada que sólo tienen una categoría después de la fusión supervisada.

Si no hay un objetivo, combine las categorías dispersas basados en recuentos

Características ordinales  Características nominales % mínimo de casos en cualquier categoría: 10

**Campos de entrada continuos**

Desechar campos discontinuos al preservar el poder predictivo (disponible en objetivo categórico únicamente)

Valor de probabilidad: 0,05

Se excluirán los campos de entrada que sólo tienen una categoría después de los intervalos.

**Selección y creación de características**

Realizar selección de características

Valor de probabilidad: 0,05

La selección de características se aplica a campos de entrada continuos si el objetivo es continuo, y a entradas categóricas.

Realizar creación de características

La creación de características se aplica a entradas continuas si el objetivo es continuo o si no hay objetivo.

**Transformar, crear y seleccionar campos de entrada para mejorar el poder predictivo.** Activa y desactiva todos los campos del panel.

**Combinar categorías dispersas para aprovechar al máximo la asociación con el destino.** Seleccione esta opción para realizar un modelo más parsimonioso reduciendo el número de variables que deben procesarse junto con el destino. Si es necesario, cambie el valor de probabilidad del valor por defecto de 0,05.

Tenga en cuenta que si todas las categorías se combinan en una, las versiones original y derivada del campo se excluyen porque no tienen ningún valor como predictor.

**Si no hay ningún destino, combine las categorías dispersas según los recuentos.** Si está tratando con datos que no tienen destino, puede fusionar las categorías dispersas de características ordinales (conjunto ordenado), nominales (conjunto) o ambas. Especifique el porcentaje mínimo de casos o registros de los datos que identifica las categorías que deben fusionarse; el valor por defecto es 10.

Las categorías se fusionan mediante las reglas siguientes:

- La fusión no se realiza en campos binarios.
- Si sólo hay dos categorías durante la fusión, ésta se detiene.
- Si no hay ninguna categoría original ni ninguna categoría creada durante la fusión con menos del porcentaje mínimo especificado de casos, la fusión se detiene.

**Desechar campos discontinuos al preservar el poder predictivo.** Si el conjunto de datos incluye un destino categórico, puede crear un intervalo para entradas continuas con asociaciones fuertes para mejorar el rendimiento del procesamiento. Si es necesario, cambie el valor de probabilidad para los subconjuntos homogéneos del valor por defecto de 0,05.

Si la operación de creación de intervalos da como resultado un único intervalo para un campo específico, las versiones original y con intervalos del campo se excluyen porque no tienen ningún valor como predictor.

*Nota:* Los intervalos en ADP son diferentes de intervalos óptimos utilizados en otras partes de IBM® SPSS® Modeler. Intervalos óptimos utiliza entropía de información para convertir una variable continua en una variable categórica; necesita ordenar los datos y almacenarlo todo en memoria. ADP utiliza subconjuntos homogéneos para agrupar una variable continua, significa que intervalos ADP no necesita ordenar los datos ni almacenar los datos en memoria. El uso del método de subconjunto homogéneo para agrupar una variable continua significa que el número de categorías después de la agrupación es siempre menor o igual que el número de categorías del destino.

**Realizar selección de características.** Seleccione esta opción para eliminar características con un coeficiente de correlación bajo. Si es necesario, cambie el valor de probabilidad del valor por defecto de 0,05.

Esta opción sólo se aplica a características de entrada continuas en las que el destino sea continuo y a características de entrada categóricas.

**Realizar construcción de características.** Seleccione esta opción para derivar nuevas características de una combinación de varias características existentes (que luego se descartan del modelado).

Esta opción sólo es aplicable a características de entrada continuas en las que el destino es continuo o en las que no hay destino.

## Nombres de campos

Figura 4-9  
Configuración de Nombrar campos de preparación automática de datos

**Campos transformados y construidos**

Extensión de nombre de campos objetivo transformado:

Extensión de nombre de campos de entrada transformada:

Nombre raíz de características creadas:

**Duraciones calculadas de fechas y horas**

Extensiones de nombres de duraciones calculadas de fechas

Años:     Meses:     Días:

Extensiones de nombres de duraciones calculadas de horas

Horas:     Minutos:     Segundos:

**Elementos cíclicos extraídos de fechas y horas**

Extensiones de nombres para elementos cíclicos extraídos de fechas

Año:     Mes:     Día:

Extensiones de nombres para elementos cíclicos extraídos de horas

Hora:     Minuto:     Segundo:

Para identificar fácilmente las características nuevas y transformadas, ADP crea y aplica nombres, prefijos o sufijos básicos nuevos. Puede modificar estos nombres para que sean más relevantes para sus propias necesidades y datos. Si desea especificar otras etiquetas, deberá realizarlo en un nodo Tipo posterior.

**Campos transformados y construidos.** Especifique las extensiones de nombre que se aplicarán a campos de entrada y de destino transformado.

Tenga en cuenta que en el nodo ADP, si establece que los campos de cadena no contengan nada puede provocar un error dependiendo de cómo decida gestionar los campos sin utilizar. Si *Cómo gestionar los campos excluidos del modelado* se establece como *Filtrar campos no utilizados* en el panel *Configuración de campos* de la pestaña *Configuración*, las extensiones de nombres de entradas y el destino pueden establecerse como nada. Los campos originales se filtran y los campos transformados se guardan sobre ellos; en este caso, los nuevos campos transformados tendrán el mismo nombre que el original.

Sin embargo si selecciona *Definir la dirección de campos no utilizados* a 'Ninguna', las extensiones de nombre vacías o nulas para el destino y las entradas provocarán un error porque estará intentando crear nombres de campo duplicados.

Además, especifique el nombre de prefijo que se aplicará a todas las características que se creen mediante la configuración de *Crear y seleccionar*. El nuevo nombre se crea adjuntando un sufijo numérico a este nombre de raíz de prefijo. El formato del número depende de cuántas nuevas características se deriven, por ejemplo:

- 1-9 características creadas se denominarán: característica1 a característica9.

- 10-99 características creadas se denominarán: característica01 a característica99.
- 100-999 características creadas se denominarán: característica001 a característica999, etcétera.

De esta forma se garantiza que las características creadas se ordenen de forma adecuada independientemente de cuántas sean.

**Duraciones calculadas de fechas y horas.** Especifique las extensiones de nombre que se aplicarán a duraciones calculadas a partir de fechas y horas.

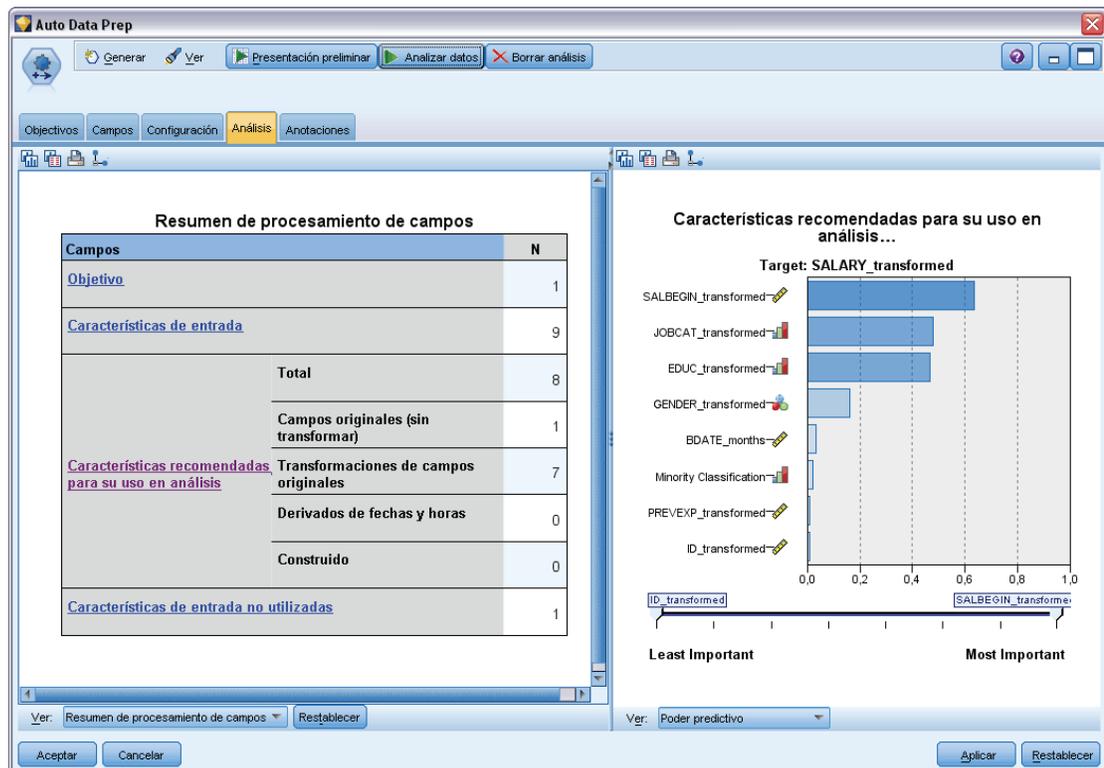
**Elementos cíclicos extraídos de fechas y horas.** Especifique las extensiones de nombre que se aplicarán a elementos cíclicos extraídos de fechas y horas.

## Pestaña análisis

- Cuando haya terminado con la configuración del nodo ADP, incluyendo las modificaciones realizadas en las pestañas Objetivos, Campos y Configuración, pulse Analizar datos; el algoritmo aplica la configuración a las entradas de datos y muestra los resultados en la pestaña Análisis.

La pestaña Análisis contiene resultados tabulares y gráficos que resumen el procesamiento de sus datos y muestra recomendaciones acerca de cómo los datos se pueden modificar o mejorar para establecer la puntuación. Puede revisar y aceptar o rechazar esas recomendaciones.

Figura 4-10  
Pestaña análisis de preparación automática de datos



La pestaña Análisis se compone de dos paneles, la vista principal en la parte izquierda y la vista relacionada o auxiliar de la derecha. Hay tres vistas principales:

- Resumen de procesamiento de campos (la configuración por defecto). Si desea obtener más información, consulte el tema [Resumen de procesamiento de campo](#) el p. 122.
- Campos. Si desea obtener más información, consulte el tema [Campos](#) el p. 123.
- Resumen de acciones. Si desea obtener más información, consulte el tema [Resumen de acciones](#) el p. 125.

Hay cuatro vistas relacionadas/auxiliares:

- Poder predictivo (la configuración por defecto). Si desea obtener más información, consulte el tema [Poder predictivo](#) el p. 126.
- Tabla de campos. Si desea obtener más información, consulte el tema [Tabla de campos](#) el p. 127.
- Detalles de campo. Si desea obtener más información, consulte el tema [Detalles de campo](#) el p. 128.
- Detalles de acción. Si desea obtener más información, consulte el tema [Detalles de acción](#) el p. 130.

### ***Enlaces entre vistas***

En la vista principal, el texto subrayado de las tablas controla la visualización en la vista vinculada. Si pulsa el texto podrá obtener detalles de un campo concreto, conjunto de campos o paso de procesamiento. El enlace que ha seleccionado aparece en color más oscuro; de esta forma podrá identificar la conexión entre el contenido de los dos paneles de vista.

### ***Restablecimiento de las vistas***

Para volver a mostrar las recomendaciones de análisis originales y abandonar los cambios que haya realizado en las vistas de análisis, pulse Restablecer en la parte inferior del panel de vista principal.

## Resumen de procesamiento de campo

Figura 4-11  
Resumen de procesamiento de campos

Resumen de procesamiento de campos		N
<u>Objetivo</u>		1
<u>Características de entrada</u>		9
	<b>Total</b>	8
	<b>Campos originales (sin transformar)</b>	1
<u>Características recomendadas para su uso en análisis</u>	<b>Transformaciones de campos originales</b>	7
	<b>Derivados de fechas y horas</b>	0
	<b>Construido</b>	0
<u>Características de entrada no utilizadas</u>		1

La tabla Resumen de procesamiento de campos proporciona una instantánea del impacto total previsto de procesamiento, incluyendo los cambios en el estado y el número de características creadas.

Tenga en cuenta que no se crea un modelo realmente, por lo que no existe una medida ni un gráfico del cambio con el poder predictivo total antes y después de la preparación de los datos. Por contra, puede visualizar los gráficos de poder predictivo de los predictores individuales recomendados.

La tabla muestra la siguiente información:

- El número de campos de destino.
- El número de predictores (de entrada) originales.
- Los predictores recomendados para su uso en el análisis y modelado. Incluye el número total de campos recomendados; el número de campos originales sin transformar recomendados; el número de campos transformados recomendados (excluyendo las versiones intermedias de campos, campos derivados de predictores de fecha y hora y predictores creados); el número de campos recomendados derivados de los campos de fecha/hora; y el número de predictores creados recomendados.
- El número de predictores de entrada no recomendados para su uso en cualquier formulario, ya sea en su formato original, como campo derivado o como entrada en un predictor creado.

Si cualquiera de la información de los Campos está subrayada, pulse para visualizar más detalles en una vista vinculada. Los detalles de Destino, Características de entrada y Características de entrada no utilizadas se muestran en la vista vinculada Tabla de campos. Si desea obtener más

información, consulte el tema [Tabla de campos](#) el p. 127. Las características recomendadas para su uso en el análisis se muestran en la vista vinculada Poder predictivo. Si desea obtener más información, consulte el tema [Poder predictivo](#) el p. 126.

## Campos

Figura 4-12  
Fields

**Campos**

**Objetivo**

Nombre	Tipo
<a href="#">SALARY</a>	

**Funciones**  Incluir campos no recomendados en la tabla

Versión de uso	Nombre	Tipo	Poder predictivo
Transformados	<a href="#">SALBEGIN</a>		0,64
Transformados	<a href="#">JOB CAT</a>		0,48
Transformados	<a href="#">EDUC</a>		0,47
Transformados	<a href="#">GENDER</a>		0,16
Transformados	<a href="#">BDATE_Duration Months</a>		0,03
Original	<a href="#">MINORITY</a>		0,02
Transformados	<a href="#">PREVEXP</a>		0,01

La vista principal Campos muestra los campos procesados y si el modo ADP recomienda su uso en modelos posteriores. Puede omitir la recomendación de cualquier campo; por ejemplo, para excluir las características creadas o incluir características que el nodo ADP recomienda excluir. Si un campo se ha transformado, puede decidir si acepta la transformación sugerida o utiliza la versión original.

La vista Campos tiene dos tablas, una para el destino y otra para los predictores procesados o creados.

### **Tabla Destino**

La tabla Destino sólo se muestra si se ha definido un destino en los datos.

La tabla contiene dos columnas:

- **Nombre.** Es el nombre de la etiqueta o del campo de destino; el nombre del original se utiliza siempre, incluso si el campo se ha transformado.
- **Nivel de medida.** Muestra el icono que representa el nivel de medición; pase el ratón por encima del icono para mostrar una etiqueta (continuo, ordinal, nominal, etc.) que describe los datos. Si el destino se ha transformado, la columna Nivel de medida refleja la versión final transformada. *Nota:* no puede desactivar las transformaciones del destino.

### **Tabla Predictores**

La tabla Predictores se muestra siempre. Cada fila de la tabla representa un campo. Por defecto, las filas se clasifican en orden descendente de potencia predictiva.

En características ordinarias, el nombre original siempre se utiliza como el nombre de la fila. Las versiones original y derivada de los campos de fecha/hora aparecen en la tabla (en filas separadas); la tabla también incluye los predictores creados.

Tenga en cuenta que las versiones transformadas de los campos que aparecen en la tabla siempre representan las versiones finales.

Por defecto sólo se muestran los campos recomendados en la tabla Predictores. Para mostrar el resto de campos, seleccione el cuadro Incluir campos no recomendados en la tabla encima de la tabla; estos campos se mostrarán en la parte inferior de la tabla.

La tabla muestra las siguientes columnas:

- **Versión de uso.** Muestra una lista desplegable que controla si un campo se utilizará posteriormente y si se utilizarán las transformaciones sugeridas. Por defecto, la lista desplegable refleja las recomendaciones.

Para los predictores comunes que se han transformado, la lista desplegable tiene tres opciones: Transformada, Original y No utilizar.

Para los predictores comunes sin transformar, las opciones son: Original y No utilizar.

Para los campos derivados de fecha/hora y los predictores creados, las opciones son: Transformada y No utilizar.

Para los campos de fecha originales, la lista desplegable está desactivada y definida a No utilizar.

*Nota:* Para los predictores con versiones originales y transformadas, si cambia entre las versiones Original y Transformadas, se actualiza automáticamente la configuración de Nivel de medida y Poder predictivo de esas características.

- **Nombre.** Cada nombre de campo es un enlace. Pulse en un nombre para ver más información acerca del campo en la vista vinculada. Si desea obtener más información, consulte el tema [Detalles de campo](#) el p. 128.

- **Nivel de medida.** Muestra el icono que representa el tipo de datos; pase el ratón por encima del icono para mostrar una etiqueta (continuo, ordinal, nominal, etc.) que describe los datos.
- **Poder predictivo.** El poder predictivo sólo se muestra en los campos que ADP recomienda. Esta columna no se muestra si no hay un destino definido. La potencia predictiva varía de 0 a 1, siendo los mayores valores los que indican “mejores” predictores. En general, la potencia predictiva resulta útil para comparar predictores con un análisis ADP, aunque los valores de potencia predictiva no deben compararse en distintos análisis.

## Resumen de acciones

Figura 4-13  
Resumen de acciones

### Resumen de acción

Acción
Campos de texto
<a href="#">Características de fecha y hora</a>
Inspección de características
<a href="#">Tipo de comprobación</a>
Valores atípicos
Valores perdidos
<a href="#">Objetivo</a>
<a href="#">Características categóricas</a>
<a href="#">Características continuas</a>

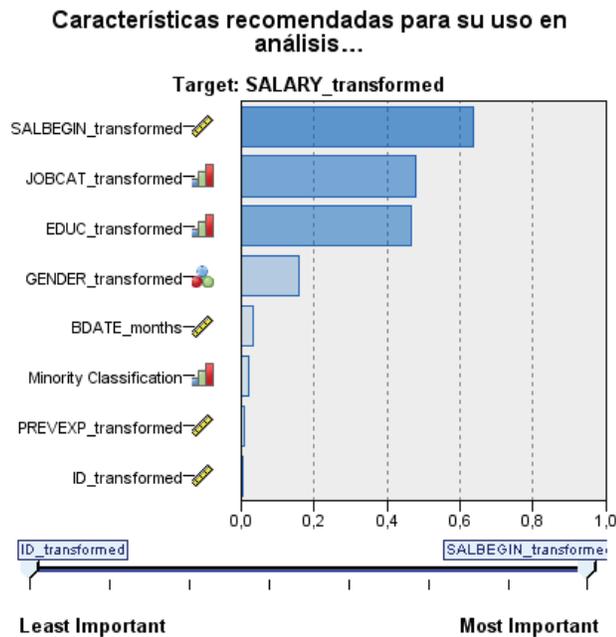
En cada acción realizada por la preparación automática de datos, los predictores de entrada se transforman y/o se filtran; los campos que sobreviven una acción se utilizarán en la acción siguiente. Los campos que sobreviven hasta el último paso se recomiendan para su uso en modelado, mientras que los predictores creados y transformados se filtran.

El Resumen de acciones es una sencilla tabla que enumera las acciones de procesamiento realizadas por ADP. Si alguna Acción está subrayada, pulse para ver más detalles en una vista vinculada sobre las acciones que se realizan. Si desea obtener más información, consulte el tema [Detalles de acción](#) el p. 130.

*Nota:* Sólo se muestran las versiones transformadas originales y finales de cada campo, no las versiones intermedias utilizadas durante el análisis.

## Poder predictivo

Figura 4-14  
Poder predictivo



Se muestra por defecto cuando el análisis se ejecuta por primera vez o cuando selecciona Predictores recomendados para su uso en análisis en la vista principal Resumen del procesamiento de campos, el gráfico muestra el poder predictivo de los predictores recomendados. Los campos se ordenan según el poder predictivo, con el campo de mayor valor en la parte superior.

En versiones transformadas de predictores comunes, el nombre del campo refleja su elección del sufijo en el panel Nombres de campos de la pestaña Configuración, por ejemplo: *\_transformadas*.

Los iconos de Nivel de medida se muestran después de los nombres de campo individuales.

El poder predictivo de cada predictor recomendado se calcula a partir de una regresión lineal o un modelo de naïve Bayes, dependiendo de si el destino es continuo o categórico.

## Tabla de campos

Figura 4-15  
Tabla de campos

**Características de entrada**

Nombre	Tipo
ID	 Continuo
GENDER	 Establecer
BDATE	 Continuo
EDUC	 Conjunto ordenado
JOBCAT	 Conjunto ordenado
SALBEGIN	 Continuo
JOBTIME	 Continuo
PREVEXP	 Continuo
MINORITY	 Conjunto ordenado

La vista Tabla de campos aparece cuando pulsa en Destino, Predictores o Predictores no utilizados en la vista principal Resumen del procesamiento de campos y muestra una tabla simple con las características relevantes.

La tabla contiene dos columnas:

- **Nombre.** Nombre del predictor.

En los destinos se utiliza el nombre original o la etiqueta del campo, incluso si el destino se ha transformado.

En versiones transformadas de predictores comunes, el nombre refleja su elección del sufijo en el panel Nombres de campos de la pestaña Configuración, por ejemplo: *\_transformadas*.

En los campos derivados de las fechas y horas se utiliza el nombre de la versión final transformada; por ejemplo: *fnacimiento\_años*.

En los predictores creados, se utiliza el nombre del predictor creado, por ejemplo: *Predictor1*.

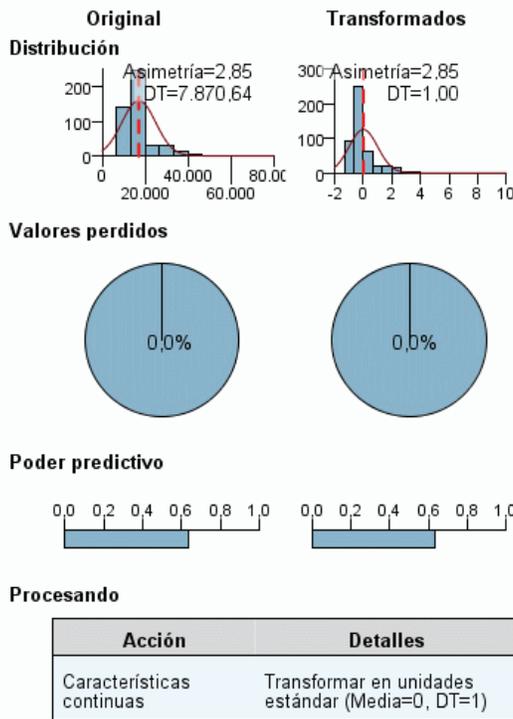
- **Nivel de medida.** Muestra el icono que representa el tipo de datos.

En Destino, el Nivel de medida siempre refleja la versión transformada (si el destino se ha transformado); por ejemplo, si se ha cambiado de ordinal (conjunto ordenado) a continuo (rango, escala) o viceversa.

## Detalles de campo

Figura 4-16  
 Buscar información detallada

### Información detallada para Beginning Salary



Nombre del campo transformado: SALBEGIN\_transformed

Se muestra cuando pulsa cualquier Nombre en la vista principal Campos, la vista Detalles de campo contiene distribución, valores perdidos y gráficos de poder predictivo (si procede) del campo seleccionado. Además, también se muestran el historial de procesamiento del campo y el nombre del campo transformado (si procede).

En cada conjunto de gráficos, las dos versiones se muestran juntas para comparar el campo con y sin las transformaciones aplicadas; si no existe una versión transformada del campo, se muestra un gráfico de la versión original únicamente. En los campos de fecha u hora derivados y en los predictores creados, los gráficos sólo se muestran para el nuevo predictor.

*Nota:* Si se excluye un campo porque tiene demasiadas categorías, solo se muestra el historial de procesamiento.

### Gráfico Distribución

La distribución de campos continuos se muestra como una curva normal superpuesta y una línea de referencia vertical para el valor principal; los campos categóricos se muestran como un gráfico de barras.

Los histogramas se etiquetan y muestran la desviación y asimetría típica; sin embargo, la asimetría no se muestra si el número de valores es 2 o menos si la varianza del campo original es inferior a 10-20.

Pase el ratón por encima del gráfico para mostrar la media de los histogramas o el número y el porcentaje del número total de registros para las categorías en gráficos de barras.

### **Gráfico de valor perdido**

Los gráficos de sectores comparan el porcentaje de valores perdidos con y sin transformaciones aplicadas; las etiquetas de gráficos muestran el porcentaje.

Si el nodo ADP ha ejecutado la gestión de valores perdidos, el gráfico de sectores posterior a la transformación también incluye el valor de sustitución como una etiqueta, es decir, el valor que se utiliza en lugar de los valores perdidos.

Pase el ratón por encima del gráfico para mostrar el valor perdido y el porcentaje del número total de registros.

### **Gráfico de poder predictivo**

En los campos recomendados, los gráficos de barras muestran el poder predictivo antes y después de la transformación. Si el destino se ha transformado, el poder predictivo se calcula con respecto al destino transformado.

*Nota:* Los gráficos de poder predictivo no se muestran si se define el destino o si el pulsa el destino en el panel de vista principal.

Pase el ratón por encima del gráfico para mostrar el valor del poder predictivo.

### **Tabla Historial de procesamiento**

La tabla muestra cómo se ha derivado la versión transformada de un campo. Las acciones que realiza ADP aparecen en el orden en que se ejecutan; sin embargo, en algunos pasos se han realizado varias acciones en un campo concreto.

*Nota:* Esta tabla no se muestra para los campos que no se han transformado.

La información de la tabla se divide en dos o tres columnas:

- **Acción.** Nombre de la acción. Por ejemplo, Predictores continuos. Si desea obtener más información, consulte el tema [Detalles de acción](#) el p. 130.
- **Detalles.** La lista de procesos ejecutados. Por ejemplo, Transformar a unidades estándar.
- **Función.** Sólo se muestra para predictores creados y se muestra la combinación lineal de campos de entrada, por ejemplo,  $0,06 * \text{edad} + 1,21 * \text{altura}$ .

## Detalles de acción

Figura 4-17  
Análisis ADP: Detalles de acción

### Paso 9: Características continuas

Transformación	Número de características	Criterios	
		Media	SD
Transformar en unidades estándar	5	0	1

Construcción de espacio de características	N
Características construidas	0
Características excluidas debido a una asociación baja con el destino	1
Características excluidas por ser constantes tras la agrupación	0

La vista vinculada Detalles de acción aparece cuando selecciona cualquier Acción subrayada en la vista principal Resumen de acciones y muestra los datos comunes y específicos de cada paso de procesamiento realizado; los detalles específicos de la acción se muestran primero.

En cada acción, se utiliza la descripción como título en la parte superior de la vista vinculada. Los detalles específicos de la acción se muestran bajo el título y pueden incluir detalles sobre el número de predictores derivados, reestructuración de campos, transformaciones de destinos, categorías fusionadas o reordenadas y predictores creados o excluidos.

A medida que se procesa cada acción, puede cambiar el número de predictores utilizados en el procesamiento, por ejemplo a medida que se excluyen o fusionan los predictores.

*Nota:* Si se ha desactivado una acción o si no se ha especificado un destino, aparece un mensaje de error en lugar de los detalles de la acción cuando pulsa en la acción en la vista principal Resumen de acciones.

Existen nueve acciones disponibles; sin embargo, no todas están necesariamente activas para cada análisis.

### Tabla Campos de texto

La tabla muestra el número de:

- Valores en blanco posteriores recortados.
- Predictores excluidos del análisis.

**Tabla Predictores de fecha y hora**

La tabla muestra el número de:

- Duraciones derivadas de los predictores de fecha y hora.
- Elementos de fecha y hora.
- Predictores derivados de fecha y hora, en total.

La fecha u hora de referencia se muestra como nota al pie si se han calculado algunas de las duraciones de fecha.

**Tabla Cribado de predictores**

La tabla muestra el número de los siguientes predictores excluidos del procesamiento:

- Constantes.
- Predictores con demasiados valores perdidos.
- Predictores con demasiados casos en una única categoría.
- Campos nominales (conjuntos) con demasiadas categorías.
- Predictores cribados, en total.

**Tabla Comprobación de nivel de medida**

La tabla muestra los números de reestructuración de campos, que se dividen en:

- Reestructuración de campos ordinales (conjuntos ordenados) como campos continuos.
- Reestructuración de campos continuos como campos ordinales.
- Reestructuración de números totales.

Si los campos de entrada (destinos o predictores) no eran continuos u ordinales, se muestra como nota al pie.

**Tabla Valores atípicos**

La tabla muestra cómo se han tratado los valores atípicos.

- El número de campos continuos donde se han encontrado y suprimido valores atípicos o el número de campos continuos donde se han encontrado valores atípicos y se han definido como perdidos, dependiendo de su configuración en el panel Preparar entradas y destino en la pestaña Configuración.
- Se ha excluido el número de campos continuos porque eran constantes, después del tratamiento de los valores atípicos.

Una nota al pie muestra el valor de corte atípico; mientras que se muestra otra nota al pie si no hay campos de entrada continuos (destino o predictores).

**Tabla Valores perdidos**

La tabla muestra el número de campos con valores perdidos sustituidos y desglosados en:

- Destino. Esta fila no se muestra si no se han especificado destinos.
- Predictores. Pueden desglosarse por el número de nominales (conjunto), ordinales (conjunto ordenado) y continuas.
- El número total de valores perdidos sustituidos.

**Tabla Destino**

La tabla muestra si se ha transformado el destino, que se muestra como:

- Transformación de Box-Cox a normalidad. Se desglosa a su vez en columnas que muestran los criterios especificados (media y la desviación típica) y Lambda.
- Categorías de destino reordenadas para mejorar la estabilidad.

**Tabla Predictores categóricos**

La tabla muestra el número de predictores categóricos:

- Cuyas categorías se han reordenado de menor a mayor para mejorar la estabilidad.
- Características cuyas categorías se han fusionado para aumentar al máximo su asociación con el destino.
- Características cuyas categorías se han fusionado para tratar categorías dispersas.
- Características cuyas categorías se han excluido por su asociación baja con el destino.
- Características cuyas categorías se han excluido porque eran constantes después de la fusión.

Se muestra una nota al pie si no se han introducido predictores categóricos.

**Tabla Predictores continuos**

Hay dos tablas. La primera muestra uno de los siguientes números de transformaciones:

- Valores de predictores transformados a unidades estándar. Además, muestra el número de predictores transformados, la media especificada y la desviación estándar.
- Valores de predictores asignados a un rango común. Además, muestra el número de predictores transformados utilizando una transformación mínima-máxima, así como los valores mínimo y máximo especificados.
- Valores de predictores en intervalos y el número de predictores en intervalos.

La segunda tabla muestra los detalles de creación de espacio predictor, mostrados como el número de predictores:

- Construido.
- Características cuyas categorías se han excluido por su asociación baja con el destino.

- Características cuyas categorías se han excluido porque eran constantes después de la agrupación.
- Excluido por ser constante tras la construcción.

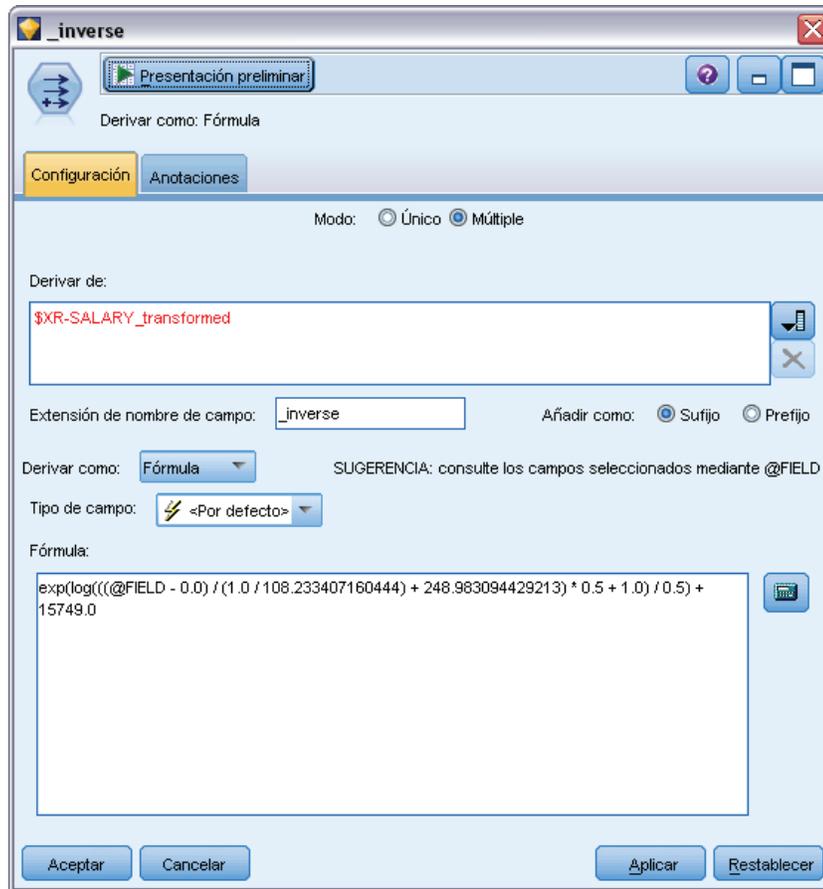
Se muestra una nota al pie si no se han introducido predictores continuos.

## ***Generación de un nodo Derivar***

Cuando genera un nodo Derivar, aplica la transformación de destino inversa al campo de puntuación. Por defecto, el nodo introduce el nombre del campo de puntuación que debería producir un modelador automático (como Clasificador automático o Autonumérico) o el nodo Conjunto. Si se ha transformado un destino (rango) de escala, el campo de puntuación se muestra en unidades transformadas; por ejemplo, log(\$) en lugar de \$. Para interpretar y utilizar los resultados, debe convertir el valor pronosticado a la escala original.

*Nota:* Sólo puede generar un nodo Derivar si el nodo ADP contiene un análisis que vuelve a trazar un destino de rango (es decir, el cambio de escala Box-Cox está seleccionado en el panel Preparar entradas y objetivo). No puede generar un nodo Derivar si el destino no es un rango o si el cambio de escala Box-Cox no está seleccionado.

Figura 4-18  
 Nodo Derivar generado desde el nodo Preparación automática de datos



El nodo Derivar se crea en modo Múltiple y utiliza @FIELD en la expresión para que pueda agregar el destino transformado si fuera necesario. Por ejemplo, utilizando los siguientes detalles:

- Nombre de campo de destino: respuesta
- Nombre de campo de destino transformado: respuesta\_transformada
- Nombre de campo de puntuación: \$XR-respuesta\_transformada

El nodo Derivar crearía un nuevo campo: \$XR-respuesta\_transformada\_inversa.

*Nota:* Si no utiliza un nodo de modelado automático o Conjunto, necesitará modificar el nodo Derivar para transformar el campo de puntuación correcto de su modelo.

### ***Destinos continuos normalizados***

Por defecto, si selecciona la casilla de verificación Cambiar la escala de destino continuo con transformación de Box-Cox en el panel Preparar entradas y objetivo, transforma el objetivo y crea un nuevo campo, será el destino de su construcción de modelos. Por ejemplo, si su destino

original era *respuesta*, el nuevo destino será *respuesta\_transformada*; los modelos posteriores al nodo ADP seleccionarán este nuevo destino automáticamente.

Sin embargo, puede causar algunos problemas, dependiendo del destino original. Por ejemplo, si el destino era *Edad*, los valores del nuevo destino no serán *Años*, sino una versión transformada de *Años*. Significa que no puede ver los resultados e interpretarlos ya que no están en unidades reconocibles. En este caso puede aplicar una transformación inversa que convertirá sus unidades transformadas en lo que eran originalmente. Para ello:

- ▶ Después de hacer pulsar Analizar datos para ejecutar el análisis ADP, seleccione *Nodo Derivar* en el menú *Generar*.
- ▶ Coloque el nodo Derivar después del nugget en el lienzo del modelo.

El nodo Derivar restaurará el campo de puntuación a las dimensiones original para que la predicción esté en los valores originales de *Años*.

Por defecto, el nodo Derivar transforma el campo de puntuación generado por un modelo de conjunto o de modelado automático. Si está creando un modelo individual, necesita modificar el nodo Derivar para derivarlo de su campo de puntuación real. Si desea evaluar su modelo, debe añadir el destino transformado a Derivar de en el nodo Derivar. Aplica la misma transformación inversa al destino y cualquier nodo Evaluación o Análisis posterior los utilizará correctamente siempre que los modifique para utilizar los nombres de campos en lugar de los metadatos.

Si también desea restaurar el nombre original, puede utilizar un nodo Filtrar para eliminar el campo de destino original si aún sigue allí y cambia el nombre del destino y los campos de puntuación.

## Nodo Tipo

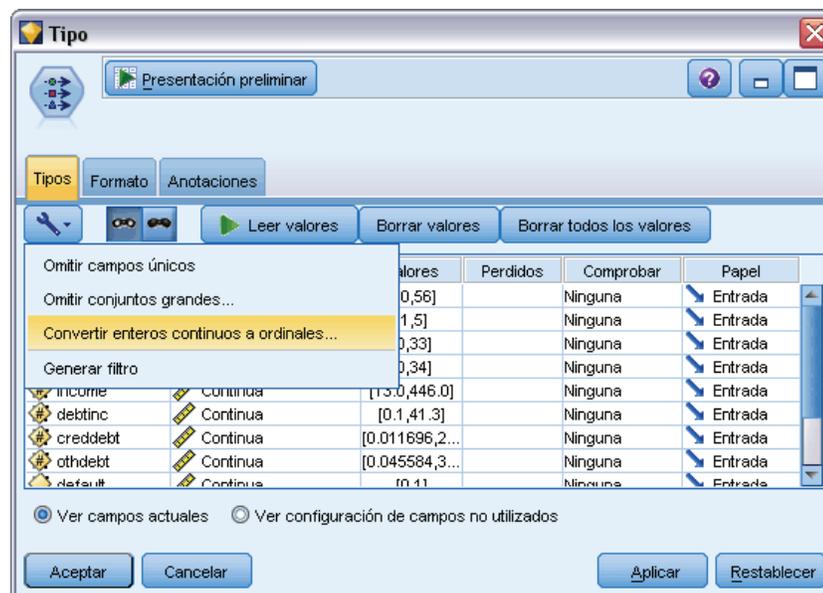
Las propiedades de campo se pueden especificar en un nodo de origen o en un nodo Tipo independiente. La funcionalidad es similar en ambos nodos. Se trata de las siguientes propiedades:

- **Campo.** Pulse dos veces en cualquier nombre de campo para especificar etiquetas de valor y de campo para los datos de IBM® SPSS® Modeler. Por ejemplo, los metadatos de campo importados de IBM® SPSS® Statistics se pueden ver o modificar aquí. Del mismo modo, puede crear nuevas etiquetas para campos y sus valores. Las etiquetas que se especifican aquí se muestran en SPSS Modeler dependiendo de las selecciones realizadas en el cuadro de diálogo de propiedades de la ruta.
- **Medida.** Este es el nivel de medición que se usa para describir características de los datos en un campo determinado. Si se conocen todos los detalles de un campo, éste se denomina **completamente instanciado**. Si desea obtener más información, consulte el tema [Niveles de medida](#) el p. 137.

*Nota:* El nivel de medición de un campo difiere del tipo de almacenamiento de un campo, lo cual indica si los datos están almacenados como cadenas, números enteros, números reales, fechas, horas o marcas de tiempo.

- **Valores.** Esta columna le permite especificar las opciones para leer los valores de datos desde el conjunto de datos, o utilizar la opción Especificar para especificar los niveles de medición y los valores en un cuadro de diálogo separado. También puede optar por pasar los campos sin leer sus valores. Si desea obtener más información, consulte el tema [Valores de campos](#) el p. 142.
- **Perdidos.** Se utiliza para especificar cómo se tratan los valores perdidos de un campo. Si desea obtener más información, consulte el tema [Definición de valores perdidos](#) el p. 147.
- **Comprobar.** En esta columna puede configurar opciones que garanticen que los valores de campo se ajusten a los valores o rangos especificados. Si desea obtener más información, consulte el tema [Comprobación de valores de tipo](#) el p. 148.
- **Papel.** Se usa para indicar a los nodos de modelado si los campos serán Entrada (campos predictores) o Objetivo (campos pronosticados) para un proceso de aprendizaje automático. Ambas y Ninguna también son papeles disponibles, junto con Partición, que indica un campo utilizado para crear particiones de registros en muestras independientes para entrenamiento, comprobación y validación. El valor Split especifica que se generarán modelos diferentes para cada valor posible del campo. Si desea obtener más información, consulte el tema [Definición del papel de campos](#) el p. 149.

Figura 4-19  
Opciones del nodo Tipo



Se pueden especificar otras opciones con la ventana del nodo Tipo:

- Con el botón de menú de herramientas puede elegir Omitir campos únicos una vez instanciado un nodo Tipo (bien a través de las especificaciones, la lectura de los valores o la ejecución de la ruta). Al omitir los campos únicos, se omiten automáticamente los campos con un solo valor.
- Con el botón de menú de herramientas puede elegir Omitir conjuntos grandes una vez instanciado un nodo Tipo. Al omitir conjuntos grandes, se omiten automáticamente los conjuntos con un número elevado de miembros.

- Con el botón de menú de herramientas puede elegir Convertir enteros continuos a ordinal una vez instanciado un nodo Tipo. Si desea obtener más información, consulte el tema [Conversión de datos continuos](#) el p. 140.
- Con el botón de menú de herramientas puede generar un nodo Filtro para descartar los campos seleccionados.
- Con los botones de gafas puede establecer los valores por defecto de todos los campos en Leer o Pasar. La pestaña Tipos del nodo de fuente pasa los campos por defecto, mientras que el propio nodo Tipo lee los valores por defecto.
- Con el botón Borrar valores puede borrar los cambios de los valores del campo realizados en este nodo (valores no heredados) y volver a leer valores de las operaciones anteriores a la ruta. Esta opción es útil para restablecer los cambios que pueda haber realizado a los campos específicos anteriores en la ruta.
- Con el botón Borrar todos los valores puede restablecer los valores de **todos** los campos que se leen en el nodo. Esta opción establece de forma eficaz la columna *Valores* en **Leer** para todos los campos. También resulta útil para restablecer los valores de todos los campos y volver a leer los valores y los tipos de las operaciones anteriores en la ruta.
- En el menú contextual puede elegir Copiar los atributos de un campo a otro. Si desea obtener más información, consulte el tema [Copia de atributos de tipo](#) el p. 150.
- Con la opción Ver configuración de campos no utilizados puede ver la configuración de los tipos que ya no están presentes en los datos o que estuvieron conectados alguna vez a este nodo Tipo. Esta opción es útil al reutilizar un nodo Tipo para conjuntos de datos que han cambiado.

## Niveles de medida

El nivel de medición (antes conocido como “tipo de datos” o “tipo de uso”) describe el uso de los campos de datos en IBM® SPSS® Modeler. El nivel de medición puede especificarse en la pestaña Tipos de un nodo de fuente o Tipo. Por ejemplo, puede que desee definir el nivel de medición para un campo de enteros con valores de 1 y 0 para una *Marca*. Esto suele indicar que 1=*Verdadero* y 0=*Falso*.

**Almacenamiento frente a medición.** Tenga en cuenta que el nivel de medición de un campo difiere de su tipo de almacenamiento, que indica si los datos están almacenados como una cadena, un número entero, un número real, una fecha, una hora o una marca de tiempo. Los tipos de datos se pueden modificar en cualquier punto de una ruta mediante un nodo Tipo, mientras que el almacenamiento ha de definirse en la fuente, al leer los datos en SPSS Modeler (si bien posteriormente puede cambiarse a través de una función de conversión). Si desea obtener más información, consulte el tema [Configuración de almacenamiento de campos y formato](#) en el capítulo 2 el p. 32.

Algunos nodos de modelado indican los tipos de nivel de medición permitidos para sus campos de entrada y objetivo por iconos en su pestaña Campos.

**Iconos de nivel de medición**

Icono	Nivel de medida
	Por defecto
	Continuo
	Categórico
	Marca
	Nominal
	Ordinal
	Sin tipo

Los siguientes niveles de medición están disponibles:

- **Por defecto.** Los datos cuyo tipo de almacenamiento y valores son desconocidos (por ejemplo, porque aún no se han leído) se muestran como <Por defecto>.
- **Continuo.** Se usa para describir valores numéricos, como el rango de 0 a 100 o de 0,75 a 1,25. Un valor continuo pueden ser un número entero, un número real o la fecha/hora.
- **Categóricas.** Se utiliza en el caso de los valores de cadena, cuando se desconoce un número exacto de valores distintos. Se trata de un tipo de datos **sin instanciar**, lo que significa que toda la información posible acerca del almacenamiento y utilización de los datos aún no se conoce. Una vez leídos los datos, el nivel de medición será *Marca*, *Nominal*, o *Sin tipo*, dependiendo del número máximo de campos nominales especificado en el cuadro de diálogo Propiedades de la ruta.
- **Marca.** Se utiliza para datos con dos valores distintos que indican la presencia o falta de una característica, como *true* y *false*, *Yes* y *No* o *0* y *1*. Los valores utilizados pueden variar, pero uno siempre debe designarse como el valor verdadero y el otro como el valor falso. Los datos pueden representarse como texto, número entero, número real, fecha, hora o marca de tiempo.
- **Nominal.** Se utiliza para describir datos con varios valores distintos, cada uno tratado como un miembro de un conjunto, como *small/medium/large*. Los datos nominales pueden tener cualquier almacenamiento: numérico, cadena o fecha/hora. Tenga en cuenta que, al definir el nivel de medición a *Nominal*, no se cambian automáticamente los valores a almacenamiento de cadena.
- **Ordinal.** Se usa para describir datos con múltiples valores distintos que tienen un orden inherente. Por ejemplo, las categorías salariales o los rangos de satisfacción pueden escribirse como datos ordinales. El orden se define a través del orden de clasificación natural de los elementos de datos. Por ejemplo, *1, 3, 5* es el orden de clasificación predeterminado de un conjunto de números enteros, mientras que *HIGH, LOW, NORMAL* (orden alfabético ascendente) es el orden de un conjunto de cadenas. El nivel de medición ordinal le permite definir un conjunto de datos categóricos como datos ordinales para la visualización, generación de modelos y la exportación a otras aplicaciones (como IBM® SPSS® Statistics), que reconoce los datos ordinales como un tipo distinto. Puede utilizar un campo ordinal en

cualquier punto donde pueda utilizarse un campo nominal. Además, los campos de cualquier tipo de almacenamiento (real, entero, cadena, fecha, hora, etc.) pueden definirse como ordinal.

- **Sin tipo.** Se utiliza para datos que no se ajustan a ninguno de los tipos anteriores, para campos con un único valor, o para datos nominales donde el conjunto tiene más miembros que el máximo definido. Resulta también útil con los casos en los que, de lo contrario, el nivel de medición sería un conjunto con demasiados miembros (como un número de cuenta). Cuando selecciona Sin tipo para un campo, el papel se establece automáticamente en Ninguno, con ID de registro como única alternativa. El tamaño máximo por defecto de los conjuntos es de 250 valores únicos. Este número puede ajustarse o desactivarse en la pestaña Opciones del cuadro de diálogo Propiedades de la ruta, que puede abrirse desde el menú Herramientas.

Puede especificar manualmente los niveles de medición o bien dejar que el software lea los datos y determine el nivel de medición según los valores que lee.

Asimismo, si tiene varios campos de datos continuos que deberían tratarse como datos categóricos, puede seleccionar una opción para convertirlos. Si desea obtener más información, consulte el tema [Conversión de datos continuos](#) el p. 140.

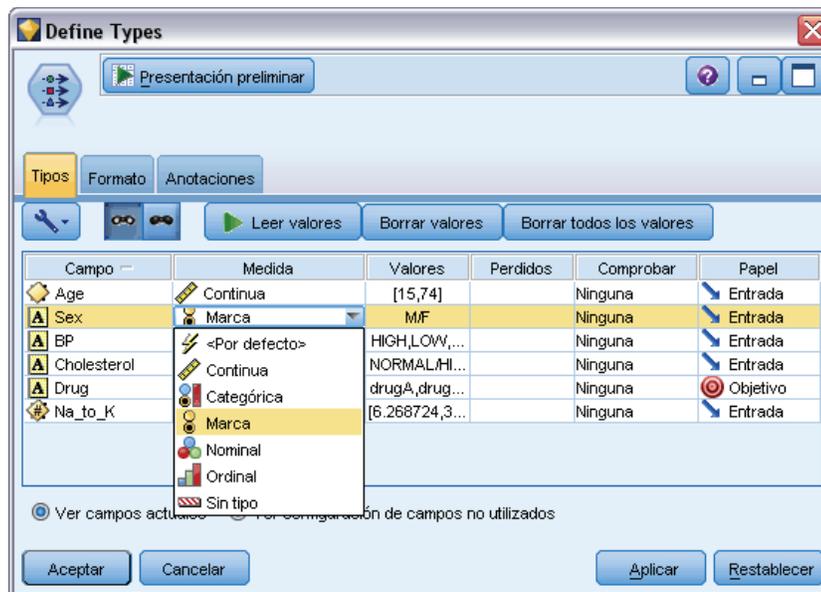
#### ***Para utilizar la definición automática de tipos***

- ▶ En el nodo Tipo o en la pestaña Tipos de un nodo de fuente, establezca la columna *Valores* en <Leer> para los campos deseados. Así se pondrán los metadatos a disposición de todos los nodos que se encuentran a continuación en la ruta. Puede establecer rápidamente todos los campos en <Leer> o <Pasarse> usando los botones de gafas del cuadro de diálogo.
- ▶ Pulse en Leer valores para leer inmediatamente los valores de la fuente de datos.

#### ***Para establecer manualmente el nivel de medición de un campo***

- ▶ Seleccione un campo de la tabla.
- ▶ En la lista desplegable de la columna *Medición* seleccione un nivel de medición para el campo.
- ▶ Si lo prefiere, puede pulsar Ctrl-A o Ctrl mientras pulsa en un campo para seleccionar varios campos antes de utilizar la lista desplegable con objeto de seleccionar un nivel de medición.

Figura 4-20  
Configuración manual de los niveles de medición

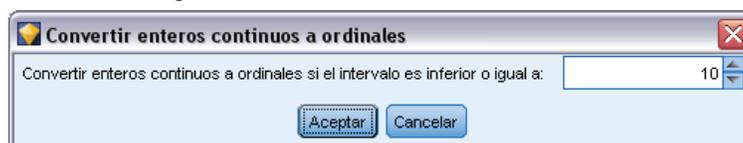


### Conversión de datos continuos

Si se trata un tipo de datos categóricos como continuos, la calidad de un modelo puede verse muy afectada, especialmente si se trata del campo de destino; por ejemplo, podría producirse un modelo de regresión en lugar de un modelo binario. Para evitarlo, puede convertir rangos enteros en tipos categóricos como *Ordinal* o *Marca*.

- ▶ Desde el botón Operaciones y menú Generar (con el símbolo de herramienta), seleccione Convertir enteros continuos a ordinal. Aparecerá el cuadro de diálogo Valores de conversión.

Figura 4-21  
Cuadro de diálogo Valores de conversión



- ▶ Especifique el tamaño del rango que se convertirá automáticamente; esto es aplicable a cualquier rango inferior e igual al tamaño que introduzca.
- ▶ Pulse en Aceptar. Los rangos afectados se convierten a *Marca* u *Ordinal* y se muestran en la pestaña Tipos del nodo Tipo.

### **Resultados de la conversión**

- Cuando un campo *Continuo* con almacenamiento entero cambia a *Ordinal*, los valores inferior y superior se amplían para incluir todos los valores enteros desde el valor inferior hasta el valor superior. Por ejemplo, si el rango es 1, 5, el conjunto de valores es 1, 2, 3, 4, 5.
- Cuando el campo *Continuo* cambia a *Marca*, los valores inferior y superior se convierten en los valores false y true del campo Marca.

## **¿Qué es la creación de una instancia?**

**La creación de una instancia** es el proceso de leer o especificar información, como los valores y el tipo de almacenamiento de un campo de datos. Para optimizar los recursos del sistema, la creación de instancias es un proceso dirigido por el usuario, que le indica al software que lea los valores especificando opciones en la pestaña Tipos de un nodo de fuente o ejecutando los datos a través de un nodo Tipo.

- Los datos con tipos desconocidos también se denominan **sin instanciar**. Los datos cuyos valores y tipo de almacenamiento son desconocidos se muestran en la columna *Medición* de la pestaña Tipos como <Por defecto>.
- Cuando se dispone de alguna información acerca del almacenamiento de un campo, como cadena o numérico, los datos se denominan **parcialmente instanciados**. Categórico o Continuo son niveles de medición parcialmente instanciados. Por ejemplo, Categórico especifica que el campo es simbólico, pero no se conoce si es nominal, ordinal o marca.
- Cuando se conocen todos los detalles de un tipo, incluidos los valores, se muestra un nivel de medición (nominal, ordinal, marca o continuo) **completamente instanciado** en esta columna. *Nota:* el tipo *Continuo* se utiliza tanto para los campos de datos parcialmente instanciados como para los campos de datos completamente instanciados. Los datos continuos pueden ser números enteros o reales.

Durante la ejecución de una ruta de datos con un nodo Tipo, los tipos sin instanciar se vuelven parcialmente instanciados, en función de los valores de datos iniciales. Una vez que todos los datos pasan por el nodo, se vuelven completamente instanciados a menos que los valores estuvieran establecidos en <Pasar>. Si la ejecución se interrumpe, los datos permanecerán como parcialmente instanciados. Una vez instanciada la pestaña Tipos, los valores de un campo se vuelven estáticos en ese punto de la ruta. Esto significa que cualquier cambio anterior a la ruta no afectará a los valores de un campo determinado, incluso si se vuelve a ejecutar la ruta. Para cambiar o actualizar los valores en función de los nuevos datos o manipulaciones añadidas, debe editarlos en la propia pestaña Tipos o establecer el valor de un campo en <Leer> o <Leer +>.

### **Cuándo se puede crear una instancia**

Normalmente, si el conjunto de datos no es muy grande y no tiene intención de añadir más adelante ningún campo a la ruta, la creación de una instancia en el nodo de fuente es el método más conveniente. Sin embargo, la creación de una instancia en un nodo Tipo independiente resulta útil cuando:

- El conjunto de datos es grande y la ruta filtra un subconjunto anterior al nodo Tipo.
- Los datos se han filtrado en la ruta.

- Los datos se han fusionado o añadido en la ruta.
- Los campos de datos nuevos se derivan durante el procesamiento.

## Valores de campos

Utilizando la columna *Valores* de la pestaña Tipos, se pueden leer automáticamente los valores desde los datos o especificar niveles de medición y valores en un cuadro de diálogo independiente.

Figura 4-22  
Selección de métodos para leer, pasar o especificar valores de datos



Las opciones disponibles de esta lista desplegable proporcionan las siguientes instrucciones para la definición automática de tipos:

Opción	Función
<Leer>	Los datos se leerán cuando el nodo se ejecute.
<Leer+>	Los datos se leerán y se añadirán a los datos actuales (si existe alguno).
<Pasar>	No se lee ningún dato.
<Actual>	Mantiene los valores de datos actuales.
Especificar...	Se ejecuta un cuadro de diálogo independiente en el que especificar los valores y las opciones de nivel de medición.

Al ejecutar un nodo Tipo o pulsar en Leer valores, los valores se leerán y se escribirán automáticamente desde la fuente de datos en función de la selección. Estos valores también se pueden especificar manualmente con la opción Especificar o al pulsar dos veces en una casilla de la columna *Campo*.

Una vez realizados los cambios a los campos en el nodo Tipo, puede restablecer la información de los valores con los siguientes botones de la barra de herramientas del cuadro de diálogo.

- Con el botón **Borrar valores** puede borrar los cambios de los valores del campo realizados en este nodo (valores no heredados) y volver a leer valores de las operaciones anteriores a la ruta. Esta opción es útil para restablecer los cambios que pueda haber realizado a los campos específicos anteriores en la ruta.
- Con el botón **Borrar todos los valores** puede restablecer los valores de **todos** los campos que se leen en el nodo. Esta opción establece de forma eficaz la columna *Valores* en **Leer** para todos los campos. También resulta útil para restablecer los valores de todos los campos y volver a leer los valores y los niveles de medición de las operaciones anteriores en la ruta.

### Utilización del cuadro de diálogo de valores

Si pulsa en la columna *Valores* o *Perdidos* de la pestaña *Tipos*, aparecerá una lista desplegable de valores predefinidos. Si selecciona la opción *Especificar* en la lista, se abre un cuadro de diálogo independiente donde puede establecer opciones para leer, especificar, etiquetar y gestionar valores para el campo seleccionado.

Figura 4-23  
Opciones de configuración de los valores de los datos

**Valores de Drug**

Medida: Nominal Almacenamiento: Cadena Campo de modelo...

Valores:  Leer de datos  Pasar  Especificar valores y etiquetas

Valores	Etiquetas
drugA	
drugB	
drugC	
drugX	
drugY	

Extender valores a partir de los datos Longitud máxima de cadena:

Comprobar valores: Ninguna

Definir vacíos

Valores perdidos

Rango  a:

Nulo  Espacio en blanco

Descripción:

Aceptar Cancelar Ayuda

Muchos de los controles son comunes a todos los tipos de datos. Aquí se describen estos controles comunes.

**Medida.** Muestra el nivel de medición seleccionado actualmente. Puede cambiar la definición para reflejar la forma en la que piensa utilizar los datos. Por ejemplo, si un campo llamado *día\_de\_la\_semana* contiene números que representan días individuales, es posible que desee

cambiar este tipo a datos nominales para crear un nodo de distribución que examine cada categoría individualmente.

**Almacenamiento.** Muestra el tipo de almacenamiento si se conoce. Los tipos de almacenamiento no se ven afectados por el nivel de medición que seleccione. Para alterar el tipo de almacenamiento, puede utilizar la pestaña Datos de los nodos de fuente Archivo fijo y Archivo var. o una función de conversión en un nodo Rellenar.

**Campo de modelo** Para campos generados como resultado de puntuar un nugget de modelo, también se pueden visualizar los detalles del campo de modelo. Incluyen el nombre del campo objetivo así como la función del campo de modelado (ya sea un valor pronosticado, probabilidad, propensión, etc.).

**Valores.** Permite seleccionar un método para determinar los valores para el campo seleccionado. Las selecciones que realice aquí omiten cualquier selección efectuada anteriormente en la columna *Valores* del cuadro de diálogo del nodo Tipo. Entre las opciones para leer valores se incluyen:

- **Leer de datos.** Seleccione esta opción para leer los valores cuando el nodo se ejecuta. Esta opción es la misma que <Leer>.
- **Pasar.** Omite la lectura de los datos del campo actual. Esta opción es la misma que <Pasar>.
- **Especificar valores.** Estas opciones se utilizan para especificar los valores y las etiquetas del campo seleccionado. Utilizada junto con la comprobación de valor, esta opción permite especificar valores en función de sus conocimientos del campo actual. Esta opción activa controles únicos para cada tipo de campo. Las opciones de los valores y las etiquetas se tratan individualmente en los temas posteriores. *Nota:* No se pueden especificar valores o etiquetas para un campo cuyo nivel de medición sea *Sin tipo* o <Por defecto>.
- **Extender valores a partir de los datos.** Seleccione esta opción para añadir los datos actuales con los valores que se introducen aquí. Por ejemplo, si *campo\_1* tiene un rango desde (0,10) e introduce un rango de valores desde (8,16), el rango se extiende al añadir el 16, sin eliminar el mínimo original. El nuevo rango sería (0,16). Al seleccionar esta opción, se establece automáticamente la opción de definición automática de tipos en <Leer+>.

**Comprobar valores.** Permite seleccionar un método de forzar valores para ajustarse a los valores especificados continuos, marca o nominales. Esta opción corresponde a la columna *Comprobar* del cuadro de diálogo del nodo Tipo y la configuración aquí realizada omite la del cuadro de diálogo. Utilizada junto con la opción de especificar valores, la comprobación de valor permite ajustar los valores de los datos a los valores esperados. Por ejemplo, si se especifican valores como 1, 0 y, a continuación, se utiliza la opción Descartar, puede descartar todos los registros con valores distintos de 1 o 0.

**Definir vacíos.** Seleccione esta opción para activar los controles que aparecen a continuación y que permiten declarar valores perdidos o vacíos en los datos.

- **Tabla Valores perdidos.** Permite definir valores específicos (como 99 o 0) como vacíos. El valor debe ser adecuado al tipo de almacenamiento del campo.
- **Rango.** Se usa para especificar un rango de valores perdidos, como edades de 1 a 17 o mayores de 65. Si un valor límite se deja en blanco, el rango no tendrá límites; por ejemplo, si se especifica un límite inferior a 100 sin límite superior, todos los valores mayores o iguales que 100 se definirán como perdidos. Los valores límite son inclusivos; por ejemplo, un rango con un límite inferior a 5 y un límite superior a 10 incluye el 5 y el 10 en la definición de rango. El

rango de valores perdidos puede definirse para cualquier tipo de almacenamiento, incluida la fecha/hora y la cadena (en cuyo caso se utilizará el orden alfabético para determinar si un valor está dentro del rango).

- **Espacio en blanco/nulo.** También puede especificar **espacios nulos** del sistema (que se muestran en los datos como \$null\$) y **espacios en blanco** (valores de cadena sin caracteres visibles) como espacios vacíos. Tenga en cuenta que el nodo Tipo también trata cadenas vacías como espacios en blanco con fines analíticos, si bien internamente éstas se almacenan de forma distinta y posiblemente se deban tratarse de otra manera en determinados casos.

*Nota:* para codificar espacios vacíos como no definidos o \$null\$, se debe utilizar el nodo Rellenar.

**Descripción.** Utilice este cuadro de texto para especificar una etiqueta de campo. Estas etiquetas aparecen en distintas ubicaciones (gráficos, tablas, resultados y exploradores de modelos), dependiendo de las selecciones realizadas en el cuadro de diálogo de propiedades de la ruta.

### **Especificación de valores y etiquetas para datos continuos**

El nivel de medición *Continuo* se utiliza para campos numéricos. Existen tres tipos de almacenamiento para datos continuos:

- Real
- Entero
- Fecha/hora

Se utiliza el mismo cuadro de diálogo para editar todos los campos continuos; solamente se muestra el tipo de almacenamiento como referencia.

Figura 4-24

*Opciones para la especificación de valores continuos y sus etiquetas*

### **Especificación de valores**

Los siguientes controles son exclusivos de los campos continuos y se utilizan para especificar un rango de valores:

**Inferior.** Especifica un límite inferior para el rango de valores.

**Superior.** Especifica un límite superior para el rango de valores.

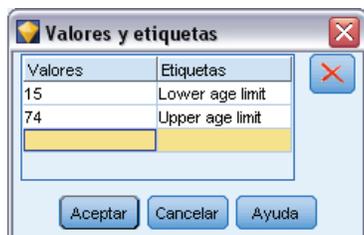
### Especificación de etiquetas

Puede especificar etiquetas para cualquier valor de un campo de rango. Pulse en el botón Etiquetas para abrir un cuadro de diálogo independiente con el fin de especificar las etiquetas de valores.

### Subcuadro de diálogo Valores y etiquetas

Al pulsar en la opción Etiquetas del cuadro de diálogo Valores de un campo de rango, se abre un cuadro de diálogo nuevo en el que pueden especificar etiquetas para cualquier valor del rango.

Figura 4-25  
Suministro de etiquetas (opcional) para valores de rango



Puede utilizar las columnas *Valores* y *Etiquetas* de esta tabla para definir pares de etiquetas y valores. A continuación se muestran los pares definidos actualmente. Puede añadir pares de etiquetas nuevos pulsando en una casilla vacía e introduciendo un valor y su correspondiente etiqueta. *Nota:* al añadir pares de valor/valor-etiqueta a esta tabla no se añade ningún valor nuevo al campo. En su lugar, simplemente se crean metadatos para el valor del campo.

Las etiquetas que se especifican en el nodo Tipo se muestran en muchos lugares (como Información sobre herramientas, etiquetas de resultados, etc.) dependiendo de las selecciones efectuadas en el cuadro de diálogo de propiedades de la ruta.

### Especificación de valores y etiquetas para datos nominales y ordinales

Los niveles de medición nominales (conjunto) y ordinales (conjunto ordenado) indican que los valores de datos se utilizan de forma discreta como miembros de un conjunto. Los tipos de almacenamiento de un conjunto pueden ser una cadena, un número entero, un número real o la fecha/hora.

Figura 4-26  
Opciones para la especificación de valores y etiquetas nominales



Los siguientes controles son exclusivos de los campos nominales y ordinales y se utilizan para especificar valores y etiquetas:

**Valores.** La columna *Valores* de la tabla permite especificar valores en función del conocimiento del campo actual. Mediante esta tabla, puede introducir valores esperados para el campo y comprobar la conformidad del conjunto de datos para estos valores utilizando la lista desplegable Comprobar valores. Puede modificar valores existentes y reordenar o eliminar valores con los botones de flecha o de eliminación.

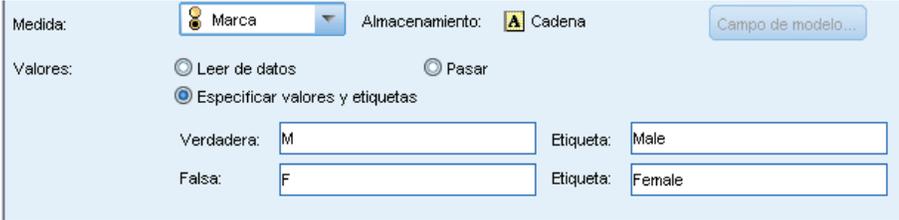
**Etiquetas** La columna *Etiquetas* permite especificar etiquetas para cada valor del conjunto. Estas etiquetas aparecen en distintas ubicaciones (gráficos, tablas, resultados y exploradores de modelos), dependiendo de las selecciones realizadas en el cuadro de diálogo de propiedades de la ruta.

### Especificación de valores para una marca

Los campos de marca se utilizan para mostrar datos que tienen dos valores distintos. Los tipos de almacenamiento de las marcas pueden ser una cadena, un número entero, un número real o la fecha/hora.

Figura 4-27

Opciones para especificar valores de campo de marcas



Medida:  Almacenamiento:

Valores:  Leer de datos  Pasar  Especificar valores y etiquetas

Verdadera:  Etiqueta:

Falsa:  Etiqueta:

**Verdadero.** Permite especificar un valor de marca para el campo cuando se cumple la condición.

**Falso.** Permite especificar un valor de marca para el campo cuando no se cumple la condición.

**Etiquetas** Permite especificar etiquetas para cada valor del campo de marcas. Estas etiquetas aparecen en distintas ubicaciones (gráficos, tablas, resultados y exploradores de modelos), dependiendo de las selecciones realizadas en el cuadro de diálogo de propiedades de la ruta.

### Definición de valores perdidos

La columna Missing de la pestaña Tipos indica si se ha definido la gestión de valores perdidos para un campo. Los ajustes posibles son:

**On (\*).** Indica que se ha definido la gestión de los valores perdidos para este campo. Puede hacerse mediante un nodo Rellenar posterior o mediante una especificación explícita utilizando la opción Specify.

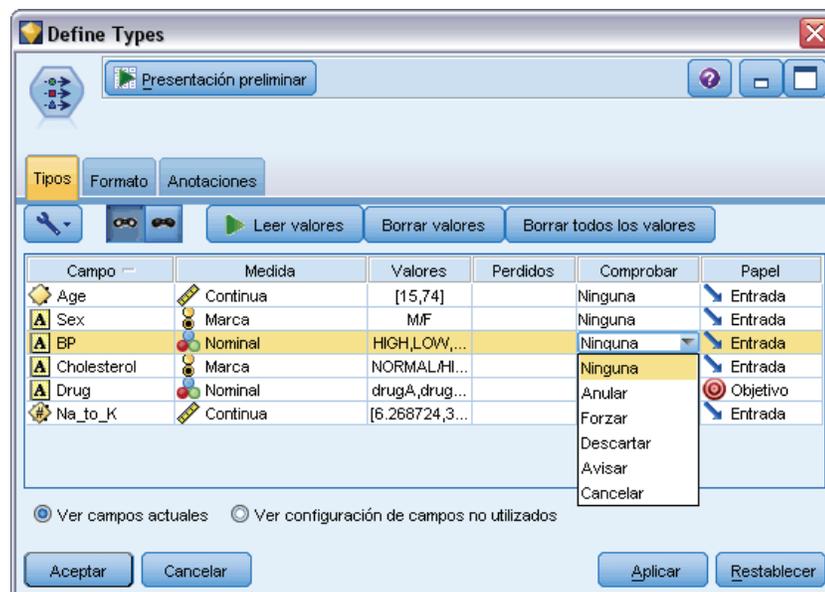
**Off.** Este campo no tiene definida la gestión de valores perdidos.

**Especificar.** Seleccione esta opción para abrir un cuadro de diálogo donde puede declarar que los valores explícitos se consideren como valores perdidos para este campo.

## Comprobación de valores de tipo

Al activar la opción Comprobar en cada campo, se examinan todos los valores de los campos para determinar si se ajustan a la configuración de tipos actual o a los valores especificados en el cuadro de diálogo Especificar valores. Esta opción es útil para limpiar los conjuntos de datos y reducir el tamaño de un conjunto de datos en una sola operación.

Figura 4-28  
Selección de opciones de comprobación para el campo seleccionado



La configuración de la columna *Comprobar* del cuadro de diálogo del nodo Tipo determina lo que ocurre si se descubre un valor que se encuentre fuera de los límites del tipo. Para cambiar la configuración de *Comprobar* en un campo, utilice la lista desplegable de ese campo en la columna *Comprobar*. Para establecer la configuración de *Comprobar* para todos los campos, pulse en la columna *Campo* y pulse Ctrl-A. A continuación, utilice la lista desplegable de cualquier campo de la columna *Comprobar*.

Éstas son las configuraciones de *Comprobar* disponibles:

**Ninguno.** Los valores pasan sin comprobación. Ésta es el ajuste por defecto.

**Anular.** Permite cambiar los valores que se encuentran fuera de los límites del valor nulo del sistema (\$null\$).

**Forzar.** Los campos cuyos niveles de medición están completamente instanciados se comprueban para determinar los valores que están fuera de los rangos especificados. Los valores no especificados se convertirán en un valor válido para este nivel de medición según las siguientes reglas:

- En el caso de las marcas, cualquier valor distinto de verdadero y falso se convierte en el valor falso.
- En el caso de los conjuntos (nominales u ordinales), cualquier valor desconocido se convierte en el primer miembro de los valores del conjunto.

- Los números que superen el límite superior de un rango se sustituyen por el límite superior.
- Los números que estén por debajo del límite inferior de un rango se sustituyen por el límite inferior.
- Los valores nulos de un rango toman el valor del punto medio de ese rango.

**Descartar.** Cuando se encuentran valores no válidos, el registro entero se descarta.

**Avisar.** El número de elementos no válidos se cuenta y se indica en el cuadro de diálogo de propiedades de la ruta tras leer todos los datos.

**Cancelar.** El primer valor no válido que se encuentra detiene la ejecución de la ruta. Se informa del error en el cuadro de diálogo de propiedades de la ruta.

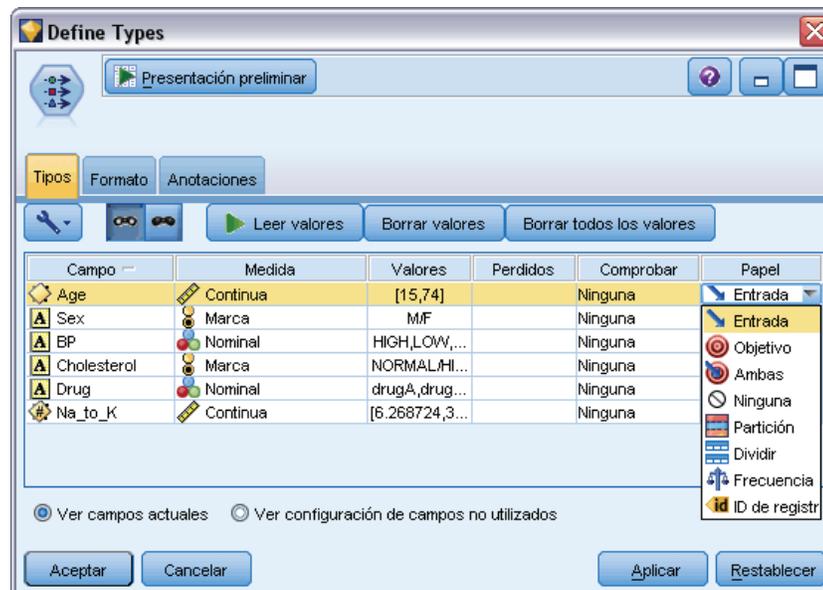
## Definición del papel de campos

El papel de un campo especifica cómo se utiliza en la generación de modelos; por ejemplo, si un campo es de entrada o si es un campo objetivo (lo que se desea pronosticar).

*Nota:* Los papeles Partición, Frecuencia e ID de registro pueden aplicarse por separado a un único campo.

Figura 4-29

Opciones de configuración para el papel del nodo Tipo



Éstos son los papeles disponibles:

**Entrada.** El campo se utiliza como una entrada en el aprendizaje de las máquinas (un campo predictor).

**Objetivo.** El campo se utiliza como un campo de salida o un campo objetivo para el aprendizaje de las máquinas (uno de los campos que el modelo intentará pronosticar).

**Ambas.** El campo se utilizará como campo de entrada y de salida por el nodo A priori. El resto de los nodos de modelado omiten el campo.

**Ninguno.** El aprendizaje de las máquinas omitirá el campo. Los campos cuyo nivel de medición se haya definido a Sin tipo se establecen automáticamente en Ninguna en la columna *Papel*.

**Partición.** Indica un campo utilizado para dividir los datos en muestras independientes para entrenamiento, comprobación y validación (opcional). El campo debe ser un tipo de conjunto instanciado con dos o más valores positivos (tal y como se define en el cuadro de diálogo Valores de campos). El primer valor representa la muestra de entrenamiento, el segundo representa la muestra de comprobación y el tercero (si está presente) representa la muestra de validación. Cualquier valor adicional se omite y los campos de marcas no se pueden usar. Tenga en cuenta que para utilizar la partición en un análisis, ésta debe estar activada en la pestaña Opciones de modelo en el nodo Análisis o la generación de modelos adecuada. Los registros con valores nulos para el campo de partición se excluyen del análisis cuando la partición está activada. Si se han definido varios campos de partición en la ruta, se deberá especificar un campo de partición simple en la pestaña Campos en todos los nodos de modelado correspondientes. Si no existe ningún campo apropiado en los datos, puede crear uno usando un nodo Partición o un nodo Derivar. Si desea obtener más información, consulte el tema [Nodo Partición](#) el p. 205.

**Segmentar.** (Campos nominales, ordinales o marca únicamente) Especifica que se creará un modelo para cada valor posible del campo.

**Frecuencia.** (Campos numéricos únicamente) La configuración de este papel permite que se pueda utilizar el valor del campo como un factor de ponderación de frecuencia para el registro. Esta función es admitida únicamente por los modelos Árbol C&, CHAID, QUEST y lineales; el resto de los nodos ignoran este papel. La ponderación de frecuencias está activada por la opción Utilizar ponderaciones de frecuencia en la pestaña Campos de los nodos de modelado que admiten la función.

**ID de registro.** El campo se utilizará como el identificador de registros exclusivo. La mayoría de los nodos ignoran esta función; sin embargo es compatible con los modelos lineales y es necesaria para los nodos de minería interna de la base de datos IBM Netezza.

## ***Copia de atributos de tipo***

Puede copiar con facilidad los atributos de un tipo (como los valores, las opciones de comprobación y los valores perdidos) desde un campo a otro.

- ▶ Pulse con el botón derecho en el campo cuyos atributos desea copiar.
- ▶ En el menú contextual, seleccione Copiar.
- ▶ Pulse con el botón derecho en los campos cuyos atributos desea cambiar.
- ▶ En el menú contextual, seleccione Pegado especial. *Nota:* puede seleccionar varios campos manteniendo pulsada la tecla Ctrl mientras pulsa en un campo o con la opción Seleccionar campos del menú contextual.

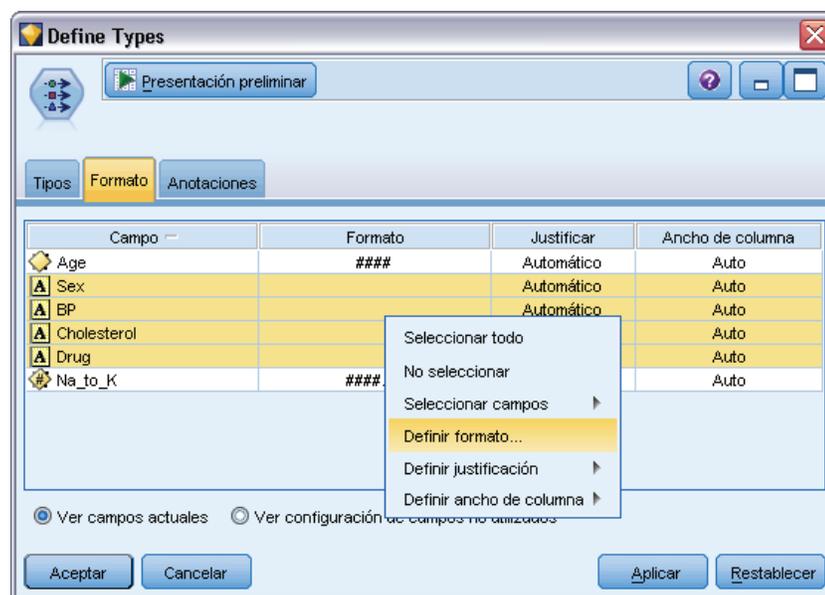
Se abrirá un cuadro de diálogo nuevo desde el que puede seleccionar los atributos específicos que desea pegar. Si pega los datos en varios campos, las opciones que seleccione aquí se aplicarán en todos los campos objetivo.

**Pegar los siguientes atributos.** Elija un elemento de la lista que aparece a continuación para pegar atributos de un campo a otro.

- **Tipo.** Seleccione para pegar el nivel de medición.
- **Valores.** Seleccione esta opción para pegar los valores de campo.
- **Perdidos.** Seleccione esta opción para pegar los valores perdidos.
- **Comprobar.** Seleccione esta opción para pegar las opciones de comprobación de valor.
- **Papel.** Seleccione esta opción para pegar el papel de un campo.

## Pestaña Configuración de Formato de campo

Figura 4-30  
nodo Tipo, pestaña Formato



La pestaña Formato de los nodos Tabla y Tipo muestra una lista de campos actuales o no utilizados así como opciones de formato para cada campo. A continuación se muestra una descripción de cada columna de la tabla de formato de campo:

**Campo.** Muestra el nombre del campo seleccionado.

**Formato.** Al pulsar dos veces en una casilla de esta columna, puede especificar el formato de los campos de forma individual con el cuadro de diálogo que se abre. Si desea obtener más información, consulte el tema [Opciones de configuración de Formato de campo](#) el p. 152. El formato especificado aquí omite el especificado en las propiedades de ruta globales.

*Nota:* los nodos Exportar Statistics y Resultados de Statistics exportan archivos *.sav* que incluyen formatos por campos en sus metadatos. Si se especifica un formato por campos que no sea compatible con el formato de archivo IBM® SPSS® Statistics *.sav*, el nodo utilizará el formato por defecto de SPSS Statistics.

**Justificar.** Utilice esta columna para especificar la forma en que los valores deben justificarse en la columna Tabla. La configuración por defecto es Automático, que justifica a la izquierda los valores simbólicos y a la derecha los numéricos. Puede omitir los valores por defecto seleccionando Izquierda, Derecha o Centro.

**Ancho de columna.** Por defecto, los anchos de columna se calculan automáticamente en función de los valores del campo. Para omitir el cálculo de ancho automático, pulse en una casilla de la tabla y utilice la lista desplegable para seleccionar un ancho nuevo. Para introducir un ancho personalizado que no aparezca en la lista, abra el subcuadro de diálogo de formatos de campo pulsando dos veces en una casilla de la tabla de la columna *Campo* o *Formato*. También puede pulsar con el botón derecho en una casilla y seleccionar Definir formato.

**Ver campos actuales.** Por defecto, el cuadro de diálogo muestra la lista de los campos que están activos actualmente. Para ver la lista de campos no utilizados, seleccione Ver configuración de campos no utilizados.

**Menú contextual.** El menú contextual de esta pestaña proporciona distintas opciones de actualización de selección y configuración.

- **Seleccionar todo.** Selecciona todos los campos.
- **No seleccionar.** Borra la selección.
- **Seleccionar campos.** Selecciona los campos en función del tipo o las características de almacenamiento. Las opciones son Seleccionar categórico, Seleccionar continuo (numérico), Seleccionar sin tipo, Seleccionar cadenas, Seleccionar números o Seleccionar fecha y hora. Si desea obtener más información, consulte el tema [Niveles de medida](#) el p. 137.
- **Definir formato.** Abre un subcuadro de diálogo para especificar las opciones de fecha, hora y decimales por campos.
- **Definir justificación.** Establece la justificación de los campos seleccionados. Las opciones son Automático, Centro, Izquierda o Derecha.
- **Definir ancho de columna.** Establece el ancho de campo de los campos seleccionados. Especifique Automático para leer el ancho de los datos. También puede establecer el ancho del campo en 5, 10, 20, 30, 50, 100 o 200.

### ***Opciones de configuración de Formato de campo***

El formato de campo se especifica en un subcuadro de diálogo disponible en la pestaña Formato de los nodos Tipo y Tabla. Si ha seleccionado más de un campo antes de abrir este cuadro de diálogo, la configuración del primer campo de la selección se utiliza para todos los campos. Al pulsar en Aceptar después de realizar las especificaciones, se aplicará esta configuración a todos los campos seleccionados en la pestaña Formato.

Figura 4-31  
Opciones de configuración de formato para uno o más campos.

The dialog box 'Formato de campo' includes the following settings:

- Formato de fecha: DDMMAA
- Formato de hora: HHMMSS
- Formato de presentación de los números: Estándar (###,###)
- Símbolo decimal: Punto (.)
- Símbolo de agrupación de los números: Valor por defecto de la ruta
- Mostrar cifras decimales:  Valor por defecto de la ruta  Especificar: 3
- Cifras decimales de científica:  Valor por defecto de la ruta  Especificar: 3
- Cifras decimales de moneda:  Valor por defecto de la ruta  Especificar: 2
- Exportar cifras decimales:  Valor por defecto de la ruta  Especificar: 6
- Justificar: Automático
- Ancho de columna:  Automático  Especificar: 10

Buttons: Aceptar, Cancelar, Ayuda

Las siguientes opciones están disponibles por campos. Muchas de estas configuraciones también se pueden especificar en el cuadro de diálogo de propiedades de la ruta. Cualquier configuración realizada en el nivel de campo omite la configuración por defecto especificada para la ruta.

**Formato de fecha.** Seleccione un formato de fecha que usar en los campos de almacenamiento de fechas o cuando las funciones de fecha de CLEM interpreten las cadenas como fechas.

**Formato de hora.** Seleccione un formato de hora que utilizar en los campos de almacenamiento de horas o cuando las funciones de hora de CLEM interpreten las cadenas como horas.

**Formato de presentación de los números.** Se puede elegir entre los formatos de presentación estándar (#####,###), científico (#,###E+##) o de moneda (\$###,##).

**Símbolo decimal.** Permite seleccionar una coma (,) o un punto (.) como separador decimal.

**Símbolo de agrupación.** Para los formatos de presentación de los números, seleccione el símbolo utilizado para agrupar valores (por ejemplo, el punto en 3.000,00). Entre las opciones se incluyen la ausencia de puntuación, el punto, la coma, el espacio y la configuración regional definida (caso en el que se utilice lo establecido por defecto en la configuración regional).

**Cifras decimales (estándar, científico, moneda, exportación).** En el caso de los formatos de presentación de los números, permite especificar el número de cifras decimales que se utilizan cuando se presentan, imprimen o exportan números reales. Esta opción se especifica de forma independiente en cada formato de presentación. El formato de exportación solamente se aplica a los campos con un almacenamiento real.

**Justificar.** Especifica la forma en que los valores deben justificarse en la columna Tabla. La configuración por defecto es Automático, que justifica a la izquierda los valores simbólicos y a la derecha los numéricos. Puede omitir los valores por defecto seleccionando Izquierda, Derecha o Centro.

**Ancho de columna.** Por defecto, los anchos de columna se calculan automáticamente en función de los valores del campo. Puede especificar un ancho personalizado en intervalos de cinco usando las flechas situadas a la derecha del cuadro de lista.

## ***Cambio de nombre o filtrado de campos***

Puede cambiar el nombre o excluir campos en cualquier punto de una ruta. Por ejemplo, como investigador médico, es posible que no esté interesado en el nivel de potasio (datos de nivel de campo) de los pacientes (datos de nivel de registro); por ello, puede filtrar el campo *K* (potasio). Se puede hacer utilizando un nodo Filtro diferente o utilizando la pestaña Filtro en un nodo de fuente o resultado. Esta funcionalidad es la misma con independencia del nodo del que se accede.

- En los nodos de fuente, como Archivo variable, Archivo fijo, Archivo Statistics y XML, puede cambiar el nombre de los campos o filtrarlos a medida que los datos se leen en IBM® SPSS® Modeler.
- Al utilizar un nodo Filtro, puede cambiar el nombre de los campos o filtrarlos en cualquier punto de la ruta.
- Desde los nodos Exportar Statistics, Transformación de Statistics, Modelo Statistics y Resultados de Statistics puede filtrar los campos o cambiarles el nombre para que cumplan los estándares de nomenclatura de IBM® SPSS® Statistics. Si desea obtener más información, consulte el tema [Cambio de nombre o filtrado de campos para IBM SPSS Statistics](#) en el capítulo 8 el p. 503.
- Puede utilizar el nodo Filtro en cualquiera de los nodos anteriores para definir o editar conjuntos de respuestas múltiples. Si desea obtener más información, consulte el tema [Edición de conjuntos de respuestas múltiples](#) el p. 158.
- Finalmente, puede utilizar un nodo Filtro para asignar campos de un nodo de fuente a otro.

Figura 4-32  
Opciones de configuración del nodo Filtro



### Opciones de configuración de filtrado

La tabla utilizada en la pestaña Filtro muestra el nombre de cada campo tal y como entra en el nodo, así como el nombre de cada uno de ellos tal y como sale. Puede utilizar las opciones de esta tabla para filtrar o cambiar el nombre de los campos que están duplicados o no son necesarios para posteriores operaciones.

- **Campo.** Muestra los campos de entrada de las fuentes de datos conectadas actualmente.
- **Filtro.** Muestra el estado del filtro de todos los campos de entrada. Los campos filtrados incluyen una X roja en la columna, indicativo de que este campo no se pasará a la parte posterior de la ruta. Pulse en la columna *Filtro* del campo seleccionado para activar o desactivar el filtrado. También puede seleccionar opciones para varios campos simultáneamente manteniendo pulsada la tecla Mayús mientras realiza la selección.
- **Campo.** Muestra los campos tal y como salen del nodo Filtro. Los nombres duplicados se muestran en rojo. Puede editar los nombres de campos pulsando en esta columna e introduciendo un nombre nuevo. También se pueden eliminar campos pulsando en la columna *Filtro* para desactivar los campos duplicados.

Todas las columnas de la tabla se pueden organizar pulsando en el encabezado de la columna.

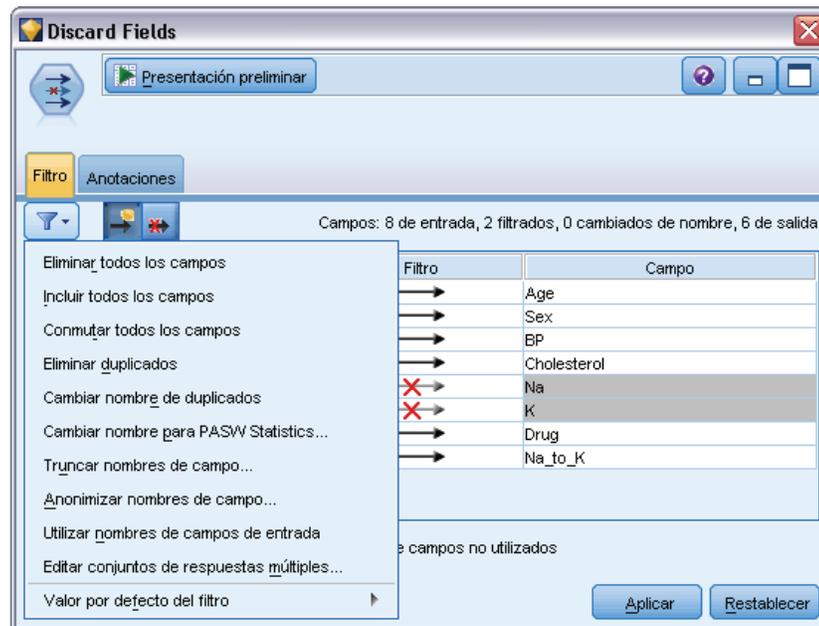
**Ver campos actuales.** Seleccione esta opción para ver los campos de los conjuntos de datos conectados de forma activa con el nodo Filtro. Esta opción está seleccionada por defecto y es el método más común para utilizar los nodos Filtro.

**Ver configuración de campos no utilizados.** Seleccione esta opción para ver los campos de los conjuntos de datos que han estado conectados con anterioridad al nodo Filtro pero ya no lo están. Esta opción resulta útil al copiar nodos Filtro de una ruta a otra o cuando se almacenan o se vuelven a cargar nodos Filtro.

### Menú del botón Filtro

Pulse en el botón Filtro en la esquina superior izquierda del cuadro de diálogo para acceder a un menú que proporciona diferentes accesos directos y otras opciones.

Figura 4-33  
Opciones del menú de filtrado



Puede seleccionar:

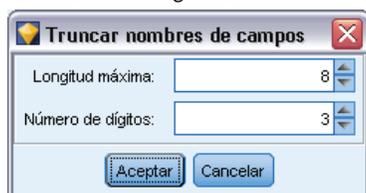
- Eliminar todos los campos.
- Incluir todos los campos.
- Conmutar todos los campos.
- Eliminar duplicados. *Nota:* al seleccionar esta opción, se eliminan todas las instancias del nombre duplicado, incluida la primera.
- Cambie el nombre de los campos y conjuntos de respuestas múltiples para que cumplan con otras aplicaciones. Si desea obtener más información, consulte el tema [Cambio de nombre o filtrado de campos para IBM SPSS Statistics](#) en el capítulo 8 el p. 503.
- Truncar nombres de campos.
- Anonimizar campo y conjunto de respuestas múltiples.
- Utilizar nombres de campos de entrada.

- Editar conjuntos de respuestas múltiples Si desea obtener más información, consulte el tema [Edición de conjuntos de respuestas múltiples](#) el p. 158.
- Establecer el estado del filtro por defecto.

También puede utilizar los botones de flechas situados en la parte superior del cuadro de diálogo para determinar si desea incluir o descartar campos por defecto. Esta opción es útil para los conjuntos de datos grandes en los que solamente se incluyen unos pocos campos a partir de ese punto de la ruta. Por ejemplo, puede seleccionar únicamente los campos que quiera conservar e indicar que se descarte el resto (en lugar de seleccionar uno a uno todos los campos que se vayan a descartar).

### Truncado de nombres de campos

Figura 4-34  
Cuadro de diálogo Truncar nombres de campos



En el menú del botón Filtro (esquina superior izquierda de la pestaña Filtro), puede seleccionar truncar nombres de campo.

**Longitud máxima.** Permite especificar un número de caracteres para limitar la longitud de los nombres de campos.

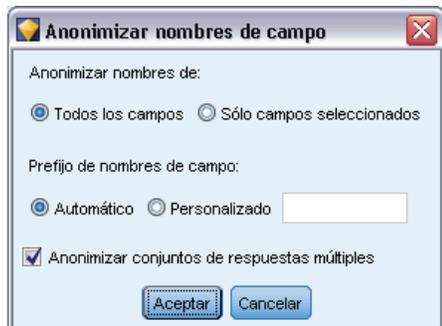
**Número de dígitos.** Si los nombres de campos, cuando están truncados, dejan de ser exclusivos, se truncarán más y se distinguirán mediante dígitos añadidos. Puede especificar el número de dígitos que utilizar. Utilice las flechas para ajustar el número.

Por ejemplo, la siguiente tabla ilustra cómo se truncan los nombres de campos de un conjunto de datos médicos mediante la configuración por defecto (longitud máxima = 8 y número de dígitos = 2).

Nombres de campos	Nombres de campos truncados
Entrada de paciente 1	Pacien01
Entrada de paciente 2	Pacien02
Ritmo cardíaco	RitmoCar
PS	PS

### **Anonimización de nombres de campos**

Figura 4-35  
Cuadro de diálogo Anonimizar nombres de campos



Puede anonimizar los nombres de los campos de cualquier nodo que incluya una pestaña Filtro pulsando en el menú del botón Filtro en la esquina superior izquierda y seleccionando Anonimizar nombres de campos. Los nombres de los campos anonimizados están compuestos por un prefijo de cadena y un valor numérico único.

**Es posible anonimizar los nombres de:** Seleccione Solamente campos seleccionados para anonimizar solamente los nombres de los campos seleccionados en la pestaña Filtro. La opción seleccionada por defecto es Todos los campos, que anonimiza todos los nombres de campos.

**Prefijo de nombres de campo:** El prefijo por defecto para los nombres de campo anonimizados es anon\_; seleccione Personalizado y escriba su propio prefijo si desea uno diferente.

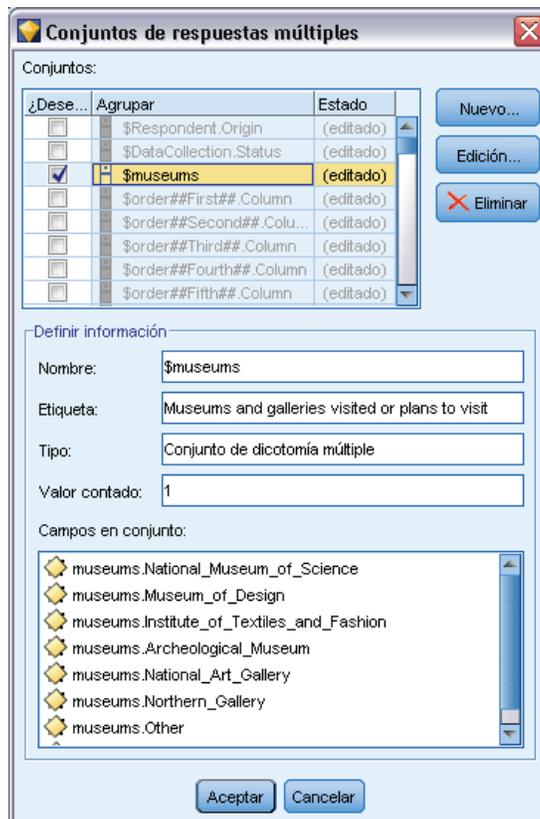
**Anonimizar conjuntos de respuestas múltiples.** Anonimiza los nombres de conjuntos de respuestas múltiples de la misma manera como campos. Si desea obtener más información, consulte el tema [Edición de conjuntos de respuestas múltiples](#) el p. 158.

Para restaurar los nombres de campo originales, seleccione Utilizar nombres de campos de entrada en el menú del botón de filtro.

### **Edición de conjuntos de respuestas múltiples**

Puede añadir o editar conjuntos de respuestas múltiples de cualquier nodo que incluya una pestaña Filtro pulsando en el menú del botón Filtro en la esquina superior izquierda y seleccionando Editar conjuntos de respuestas múltiples.

Figura 4-36  
Cuadro de diálogo Conjuntos de respuestas múltiples



Los conjuntos de respuestas múltiples se utilizan para registrar datos que pueden tener más de un valor para cada caso, por ejemplo, si pregunta a los encuestados los museos que han visitado o las revistas que han leído. Los conjuntos de respuestas múltiples se pueden importar a IBM® SPSS® Modeler mediante un nodo de fuente Data Collection o un nodo de fuente de archivo Statistics y se pueden definir en SPSS Modeler mediante un nodo Filtro.

- Pulse en Nuevo para crear un conjunto de respuestas múltiples nuevo o pulse en Editar para modificar un conjunto existente.

Figura 4-37  
Edición de un conjunto de respuestas múltiples

**Nombre y etiqueta.** Especifica el nombre y la descripción del conjunto.

**Tipo.** Las preguntas de respuestas múltiples se pueden gestionar de dos formas:

- **Conjunto de dicotomía múltiple.** Se crea un campo de marca independiente para cada respuesta posible, además si hay 10 revistas, habrá 10 campos de marca, cada una con valores como 0 ó 1 para *true* o *false*. El valor de recuento permite especificar el valor que se recontará como true. Este método es muy útil si desea dejar que los encuestados puedan seleccionar todas las opciones aplicables.
- **Conjunto de categoría múltiple.** Se crea un campo nominal para cada respuesta, hasta el número máximo de respuestas de un encuestado concreto. Cada campo nominal tiene valores que representan las posibles respuestas, como 1 para *Muy Interesante*, 2 para *National Geographic* y 3 para *PC Actual*. Este método es de gran utilidad si desea limitar el número de respuestas, por ejemplo, si se pide a los encuestados que seleccionen entre las tres revistas que leen con mayor frecuencia.

**Campos del conjunto.** Utilice los iconos a la derecha para añadir o eliminar campos.

Figura 4-38  
Pregunta de respuestas múltiples

**Comentarios**

- Todos los campos incluidos en un conjunto de respuestas múltiples deben estar en la misma ubicación.
- Los conjuntos son diferentes de los campos que incluyen. Por ejemplo, si elimina un conjunto, no se eliminarán los campos que incluyen, únicamente los enlaces entre los campos. El conjunto seguirá siendo visible desde el punto anterior de su eliminación, pero no desde un punto posterior.
- Si cambia el nombre de los campos mediante un nodo Filtro (directamente en la pestaña o seleccionando las opciones Cambiar nombre para IBM® SPSS® Statistics, Truncar o Anonimizar del menú Filtro), también se actualizarán todas las referencias a estos campos utilizados en conjuntos de respuestas múltiples. Sin embargo, los campos de cualquier conjunto de respuestas múltiples que el nodo Filtro active, no se eliminarán del conjunto de respuestas múltiples. Aunque estos campos ya no sean visibles en la ruta, el conjunto de respuestas múltiples seguirá haciendo referencia, lo que puede ser de gran ayuda cuando realice exportaciones, por ejemplo.

**Nodo Conjunto**

El nodo Conjunto combina dos o más nugget de modelo para obtener pronósticos más precisos que pueden conseguirse de los modelos individuales. Al combinar predicciones de varios modelos, pueden evitarse las limitaciones en modelos individuales que dan como resultado una precisión global superior. Los modelos combinados de esta forma suelen ejecutarse tan bien como el mejor de los modelos individuales y, en ocasiones, mejor.

Esta combinación de nodos ocurre de forma automática en los nodos de modelado automático Clasificador automático, Conglomerado automático y Autonumérico.

Después de utilizar un nodo Conjunto, puede utilizar un nodo Análisis o Evaluación para comparar la precisión de los resultados combinados con cada uno de los modelos de entrada. Para ello, asegúrese de que la opción Filtrar campos generados por modelos de conjunto no está seleccionada en la pestaña Configuración en el nodo Conjunto.

**Campos de resultado**

Cada nodo Conjunto genera un campo que contiene las puntuaciones combinadas. El nombre se basa en el campo objetivo especificado y con el prefijo  $\$XF_$ ,  $\$XS_$  o  $\$XR_$ , dependiendo del nivel de medición (marca, nominal (conjunto) o continuo (rango), respectivamente). Por ejemplo, si el objetivo es un campo de marca denominado *respuesta*, el campo de resultado sería  $\$XF\_respuesta$ .

**Confianza o campos de propensión.** Para los campos nominales y marca, se crean campos de propensión o confianza adicionales en función del método de conjunto, tal y como se detalla en la tabla siguiente:

Método de conjunto	Nombre de campo
Voting Votación ponderada de confianza Votación ponderada de propensión bruta Votación ponderada de propensión ajustada La mayor confianza gana	$\$XFC_{<campo>}$
Propensión bruta media	$\$XFRP_{<campo>}$
Propensión bruta ajustada media	$\$XFAP_{<campo>}$

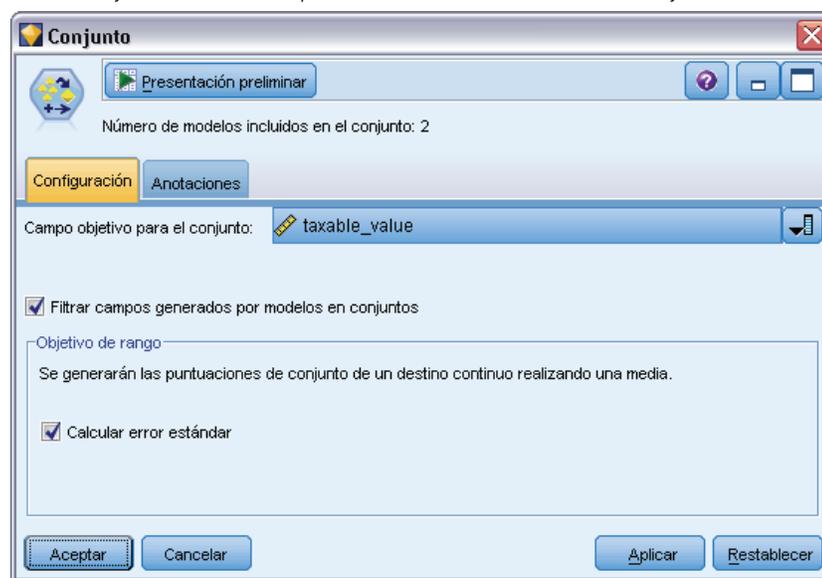
### Configuración de nodo Conjunto

**Campo objetivo de conjunto.** Seleccione un único campo que se utilice como objetivo por dos o más modelos anteriores. Los modelos anteriores pueden utilizar objetivos marca, nominales o continuos, pero al menos dos de estos modelos deben compartir el mismo objetivo para combinar las puntuaciones.

**Filtrar campos generados por modelos de conjunto.** Elimina los resultados de todos los campos adicionales que generan los modelos individuales que contiene el nodo Conjunto. Seleccione esta casilla de verificación si solamente está interesado en la puntuación combinada de todos los modelos de entrada. Asegúrese de que esta opción no está seleccionada si, por ejemplo, desea utilizar un nodo Análisis o Evaluación para comparar la precisión de la puntuación combinada con la de cada uno de los modelos de entrada individuales.

Figura 4-39

Nodo Conjunto con un campo continuo seleccionado como objetivo



Las configuraciones disponibles dependen del nivel de medición del campo seleccionado como objetivo.

### Objetivos continuos

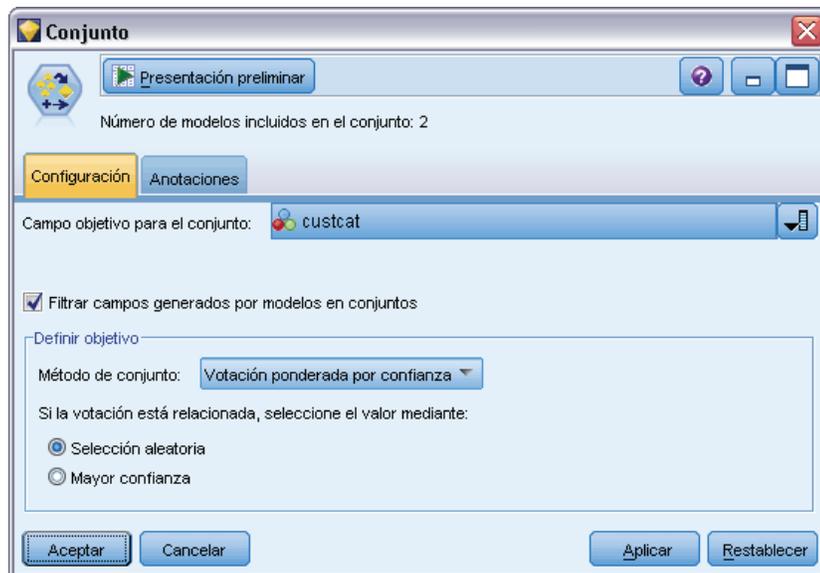
En un objetivo continuo, se calculará el promedio de las puntuaciones. Es el único método disponible para combinar puntuaciones.

Al promediar promedios o estimaciones, el nodo Conjunto utiliza un cálculo de error estándar para calcular la diferencia entre los valores medidos o estimados y los valores true, y para mostrar si las estimaciones coinciden. Los cálculos de error estándar se generan para los nuevos modelos por defecto; sin embargo, puede cancelar la selección de la casilla de verificación de los modelos existentes, por ejemplo, si se van a regenerar.

### Objetivos categóricos

Los objetivos categóricos admiten diferentes métodos, incluyendo **votación**, que funciona cuadrando el número de veces que cada posible valor pronosticado se elige y seleccionando el valor con mayor valor total. Por ejemplo, si tres de los cinco modelos pronostican *sí* y las otras dos pronostican *no*, *sí* gana por 3 a 2. Además, los votos se pueden **ponderar** en función del valor de confianza o propensión de cada predicción. Las ponderaciones se suman y el valor con el mayor total se vuelve a seleccionar. La confianza de la predicción final es la suma de las ponderaciones del valor seleccionado dividido por el número de modelos que se incluyen en el conjunto.

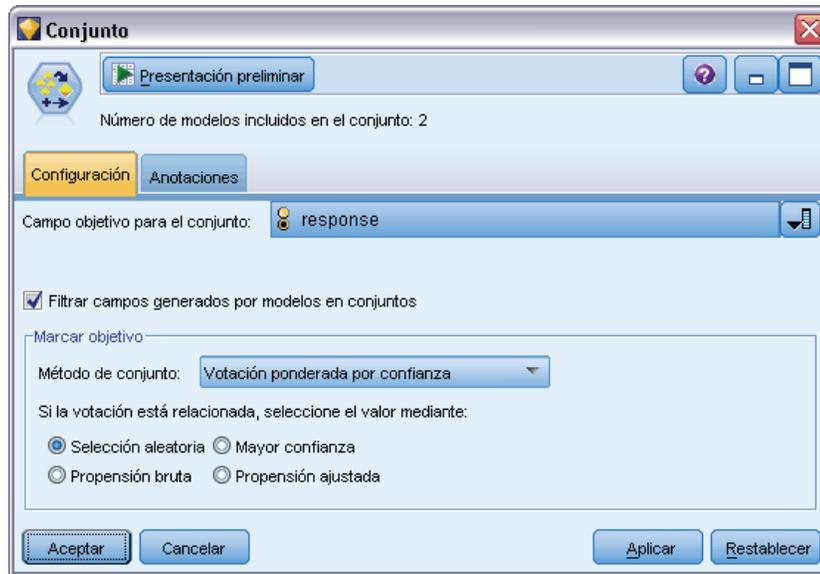
Figura 4-40  
Nodo Conjunto con un campo nominal seleccionado como objetivo



**Todos los campos categóricos.** Los campos marca y nominales admiten los métodos siguientes:

- Voting
- Votación ponderada de confianza
- La mayor confianza gana

Figura 4-41  
 Nodo Conjunto con un campo de marca seleccionado como objetivo



**Campos de marcas únicamente.** En los campos de marca únicamente, también existen disponibles diferentes métodos según su propensión:

- Votación ponderada de propensión bruta
- Votación ponderada de propensión ajustada
- Propensión bruta media
- Propensión ajustada media

**Empates de votaciones.** En los métodos de votación puede especificar cómo se resuelven los empates.

- **Selección aleatoria.** Se seleccionará aleatoriamente uno de los valores empatados.
- **La mayor confianza.** El valor con el mayor índice de confianza gana. Tenga en cuenta que no es necesariamente la misma opción que la del índice de mayor confianza de todos los valores pronosticados.
- **Propensión bruta o ajustada (campos con marcas únicamente).** El valor empatado pronosticado con la mayor propensión absoluta, donde la propensión absoluta se calcula como:

$$\frac{\text{abs}(0.5 - \text{propensión})}{2}$$

O, en el caso de la propensión ajustada:

$$\text{abs}(0,5 - \text{propensión ajustada}) * 2$$

## Nodo Derivar

Una de las funciones más eficaces de IBM® SPSS® Modeler es la capacidad de modificar valores de datos y derivar campos nuevos a partir de datos existentes. Durante proyectos minería de datos de larga duración, es común realizar varias derivaciones, como extraer un ID de cliente a partir de una cadena de datos del registro Web o crear un valor de por vida de clientes basado en los datos demográficos y de transacción. Todas estas transformaciones pueden realizarse, utilizando diversos nodos de operaciones con campos.

Existen diversos nodos que ofrecen la capacidad de derivar campos nuevos:



El nodo Derivar modifica los valores de datos o crea campos nuevos desde uno o más campos existentes. Crea campos del tipo fórmula, marca, nominal, estado, recuento y condicional. Si desea obtener más información, consulte el tema [Nodo Derivar](#) el p. 165.



El nodo Reclasificar transforma un conjunto de valores categóricos en otro. La reclasificación es útil para contraer categorías o reagrupar datos para su análisis. Si desea obtener más información, consulte el tema [Nodo Reclasificar](#) el p. 185.



El nodo Intervalos crea automáticamente nuevos campos nominales (conjunto) en función de los valores de uno o más campos continuos (rango numérico) existentes. Por ejemplo, puede transformar un campo de ingresos continuo en un campo categórico nuevo que contenga grupos de ingresos como desviaciones desde la media. Una vez creados los intervalos para el campo nuevo, puede generar un nodo Derivar en función de los puntos de corte. Si desea obtener más información, consulte el tema [Nodo Intervalos](#) el p. 189.



El nodo Marcas deriva varios campos de marcas en función de los valores categóricos definidos para uno o más campos nominales. Si desea obtener más información, consulte el tema [Nodo Marcas](#) el p. 207.



El nodo Reestructurar convierte un campo nominal o marca en un grupo de campos que se puede rellenar con los valores todavía de otro campo. Por ejemplo, para un campo determinado llamado *tipo de pago*, con valores de *crédito*, *efectivo*, y *débito*, se crearían tres campos nuevos (*crédito*, *efectivo*, *débito*), que contendría cada uno el valor del pago real realizado. Si desea obtener más información, consulte el tema [Nodo Reestructurar](#) el p. 209.



El nodo Histórico se utiliza para crear campos nuevos que contienen datos de los campos de registros anteriores. Los nodos Histórico se suelen utilizar para los datos secuenciales, como los datos de series temporales. Antes de utilizar un nodo Histórico, puede desear ordenar los datos utilizando un nodo Ordenar. Si desea obtener más información, consulte el tema [Nodo Histórico](#) el p. 237.

### Utilización del nodo Derivar

Al utilizar el nodo Derivar, puede crear seis tipos de campos nuevos desde uno o más campos existentes:

- **Fórmula.** El campo nuevo es el resultado de una expresión CLEM arbitraria.

- **Marca.** El campo nuevo es una marca que representa una condición especificada.
- **Nominal.** El campo nuevo es nominal, lo cual supone que sus miembros conforman un grupo de valores especificados.
- **Estado.** El campo nuevo es uno de dos estados. El cambio entre estos estados los desencadena una condición especificada.
- **Recuento.** El campo nuevo está basado en el número de veces que una condición es verdadera.
- **Condicional.** El campo nuevo es el valor de una de las dos expresiones, dependiendo del valor de una condición.

Cada uno de estos nodos contiene un conjunto de opciones especiales en el cuadro de diálogo del nodo Derivar. Estas opciones se describen en los siguientes temas.

### **Opciones básicas de configuración del nodo Derivar**

En la parte superior del cuadro de diálogo de los nodos Derivar se incluyen varias opciones para seleccionar el tipo de nodo Derivar que se necesita.

Figura 4-42  
Cuadro de diálogo del nodo Derivar



**Moda.** Seleccione Único o Múltiple, dependiendo de si desea derivar múltiples campos. Al seleccionar Múltiple, el cuadro de diálogo muestra opciones para derivar múltiples campos.

**Derivar campo.** En el caso de los nodos Derivar únicos, especifique el nombre del campo que desea derivar y añádale a cada registro. El nombre por defecto es Derivar $N$ , donde  $N$  es el número de nodos Derivar que se han creado durante la sesión actual.

**Derivar como.** Seleccione un tipo de nodo Derivar, como Fórmula o Nominal, de la lista desplegable. Para cada tipo, se crea un campo nuevo en función de las condiciones que se especifiquen en el cuadro de diálogo de cada tipo.

Al seleccionar una opción de la lista desplegable, se añade un conjunto nuevo de controles al cuadro de diálogo principal de acuerdo con las propiedades de cada tipo de nodo Derivar.

**Tipo de campo.** Seleccione un nivel de medición (como continuo, categórico o marca) para el nodo recién derivado. Esta opción es común para todas las formas de nodos Derivar.

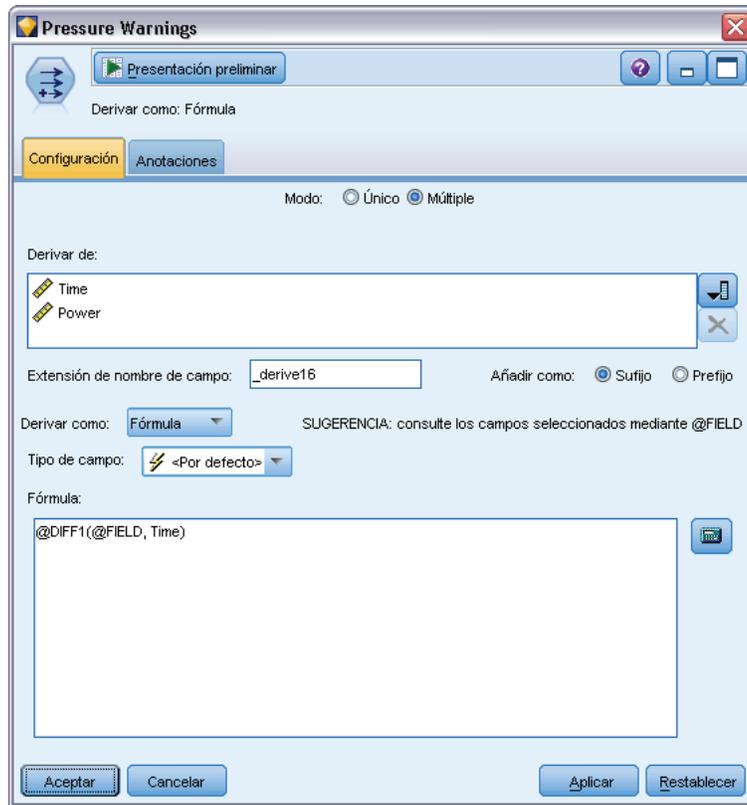
*Nota:* para derivar campos nuevos, se suele requerir el uso de funciones especiales o expresiones matemáticas. Para facilitar la creación de estas expresiones, existe un generador de expresiones en el cuadro de diálogo para todos los tipos de nodos Derivar, que ofrece comprobación de reglas y una lista completa de expresiones del lenguaje para la manipulación de expresiones CLEM.

### ***Derivación de múltiples campos***

La selección del modo Múltiple en un nodo Derivar permite derivar múltiples campos en función de la misma condición dentro de un único nodo. Esta función permite ahorrar tiempo cuando se desea realizar transformaciones idénticas en varios campos del conjunto de datos. Por ejemplo, si desea generar un modelo de regresión que pronostique el salario actual en función del salario inicial y la experiencia previa, puede resultar útil aplicar una transformación de registro a las tres variables desviadas. En lugar de añadir un nodo Derivar nuevo a cada transformación, puede aplicar la misma función a todos los campos a la vez. Solamente hay que seleccionar todos los campos desde los que se va a derivar un campo nuevo y, a continuación, escribir la expresión de derivación con la función @FIELD dentro de los paréntesis del campo.

*Nota:* la función @FIELD es una herramienta importante para derivar múltiples campos al mismo tiempo. Permite hacer referencia a los contenidos del campo actual o los campos sin especificar el nombre de campo exacto. Por ejemplo, una expresión CLEM utilizada para aplicar una transformación de registro a múltiples campos es  $\log(@FIELD)$ .

Figura 4-43  
Derivación de múltiples campos



Las siguientes opciones se añaden al cuadro de diálogo al seleccionar el modo Múltiple:

**Derivar de.** Permite utilizar el selector de campos para elegir los campos a partir de los cuales derivar campos nuevos. Se generará un campo de salida para cada campo seleccionado. *Nota:* los campos seleccionados no tienen que coincidir en el tipo de almacenamiento; sin embargo, la operación Derivar fallará si la condición no es válida para *todos* los campos.

**Extensión de nombre de campo.** Permite escribir la extensión que desea añadir a los nombres de los campos nuevos. Por ejemplo, en el caso de un campo nuevo que contenga el registro *Salario Actual*, podría añadir la extensión *reg\_* al nombre de campo, que resultaría en *reg\_Salario Actual*. Utilice los botones de radio para seleccionar si desea añadir la extensión como un prefijo (al principio) o como un sufijo (al final) del nombre de los campos. El nombre por defecto es Derivar*N*, donde *N* es el número de nodos Derivar que se han creado durante la sesión actual.

Como en el caso del nodo Derivar en modo único, debe crear una expresión para utilizarla al derivar un campo nuevo. Existen distintas opciones para crear una condición en función del tipo de operación de derivación seleccionada. Estas opciones se describen en los siguientes temas. Para crear una expresión, solamente hay que escribir en los campos de la fórmula o utilizar el generador de expresiones, al que se accede mediante el botón de calculadora. Recuerde utilizar la función @FIELD al hacer referencia a las manipulaciones en múltiples campos.

### **Selección de varios campos**

Todos los nodos que realizan operaciones en múltiples campos de entrada, como Derivar (modo múltiple), Agregar, Ordenar, G. múltiple y Gráfico de tiempo, permiten seleccionar los campos de forma sencilla a través del siguiente cuadro de diálogo Seleccionar campos.

Figura 4-44  
Selección de varios campos



**Ordenar por.** Puede ordenar campos disponibles para su visualización eligiendo una de las siguientes opciones:

- **Natural.** El orden de los campos es aquél en que pasaron desde la parte anterior de la ruta de datos al nodo actual.
- **Nombre.** Ordena los campos siguiendo un orden alfabético para su visualización.
- **Tipo.** Ordena los campos en función de su nivel de medición. Esta opción es útil cuando se seleccionan campos con un nivel de medición en particular.

Seleccione los campos de la lista de uno en uno o mantenga pulsada la tecla Mayús o Ctrl mientras selecciona otros campos para seleccionar varios campos. También puede utilizar los botones que se muestran bajo la lista para seleccionar grupos de campos en función de su nivel de medición o seleccionar y anular la selección de todos los campos de la tabla.

### **Opciones de configuración de derivación como fórmula**

Los nodos de derivación como fórmula crean un campo nuevo para cada registro de un conjunto de datos a partir de los resultados de una expresión CLEM. Tenga en cuenta que esta expresión no puede ser condicional. Para derivar valores en función de una expresión condicional, utilice la marca o el tipo condicional del nodo Derivar.

Figura 4-45  
Opciones de configuración como fórmula de un nodo Derivar



**Fórmula.** Permite especificar una fórmula mediante el lenguaje CLEM con el fin de derivar un valor para el campo nuevo.

### ***Opciones de configuración de derivación como marca***

Los nodos de derivación como marca se utilizan para indicar una condición específica, como una alta presión sanguínea o una inactividad de cuenta del cliente. Se crea un campo de marcas para cada registro y, cuando se cumple la condición verdadera, el valor de marca verdadero se añade al campo.

Figura 4-46  
Derivación de un campo de marcas

healthy

Presentación preliminar

Derivar como: Marca

Configuración Anotaciones

Modo:  Único  Múltiple

Derivar campo:  
healthy

Derivar como: Marca

Tipo de campo: Marca

Valor para Verdadero: T Valor para Falso: F

Verdadero si:  
'fruitveg' == 'T' and 'fish' == 'T'

Aceptar Cancelar Aplicar Restablecer

**Valor para Verdadero.** Permite especificar un valor para incluirlo en el campo de marcas para los registros que cumplen la condición que se especifica a continuación. El valor por defecto es T (del inglés 'True').

**Valor para Falso.** Permite especificar un valor para incluirlo en el campo de marcas de los registros que *no* cumplen la condición que se especifica a continuación. El valor por defecto es F.

**Verdadero si.** Permite especificar una condición CLEM para evaluar ciertos valores de cada registro y proporcionar un valor verdadero o falso al registro (definido anteriormente). Tenga en cuenta que el valor verdadero se asignará a los registros en caso de que existan valores numéricos no falsos.

*Nota:* para devolver una cadena vacía, es necesario escribir comillas de apertura y cierre sin nada entre ellas, como "". Las cadenas vacías se suelen utilizar a menudo, por ejemplo, como el valor falso para permitir que los valores verdaderos destaquen con más claridad en la tabla. De igual manera, las comillas deben utilizarse si desea un valor de cadena que, de otro modo, se trataría como número.

**Ejemplo**

En las versiones de IBM® SPSS® Modeler anteriores a la 12.0 importaban respuestas múltiples en un solo campo, con valores separados por comas.

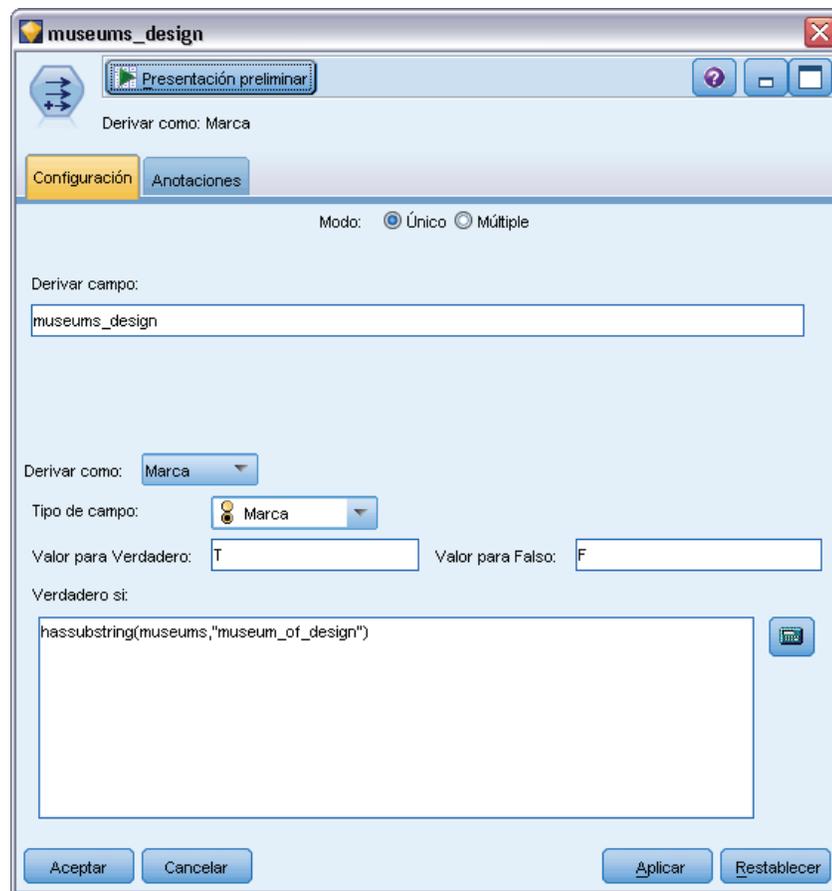
<b>museos</b>
museo_de_diseño,instituto_de_moda_y_textil
museo_de_diseño
museo_arqueológico
\$null\$
museo_nacional_de_arte,museo_nacional_de_ciencias,otro

Para preparar los datos para el análisis, puede utilizar la función `hasubstring` para generar un campo de marcas independiente para cada respuesta con una expresión como:

```
hasubstring(museums,"museo_de_diseño")
```

Figura 4-47

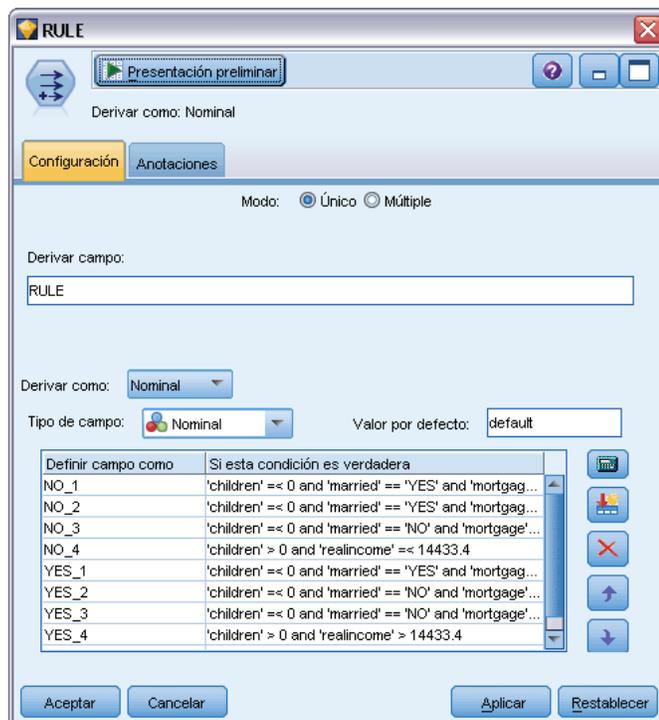
Derivación de un campo de marca con la función `hasubstring`



## Opciones de configuración de derivación como conjunto

Los nodos de derivación como conjunto se utilizan para ejecutar un conjunto de condiciones CLEM que determine la condición que satisface cada registro. Cuando se cumple una condición para cada registro, se añadirá un valor (que indica el conjunto de condiciones que se cumple) al nuevo campo derivado.

Figura 4-48  
Uso de un nodo de derivación como conjunto



**Valor por defecto.** Permite especificar un valor que utilizar en el campo nuevo si no se cumple ninguna de las condiciones.

**Definir campo como.** Permite especificar un valor que introducir en el campo nuevo cuando se cumple una determinada condición. Cada valor de la lista tiene una condición asociada que se puede especificar en la columna adyacente.

**Si esta condición es verdadera.** Permite especificar una condición para cada miembro en el campo de conjuntos que se va a enumerar. Utilice el generador de expresiones para seleccionar de entre los campos y las funciones disponibles. Puede utilizar los botones de flecha y eliminación para reorganizar o eliminar las condiciones.

Las condiciones funcionan comprobando los valores de un campo determinado del conjunto de datos. Al comprobar cada condición, los valores especificados con anterioridad se asignan al nuevo campo para indicar la condición que se cumple, (caso de haber alguna). Si no se cumple ninguna de las condiciones, se utiliza el valor por defecto.

## Opciones de configuración de derivación como estado

En cierto modo, los nodos de derivación como estado son similares a los nodos de derivación como marca. Los nodos de derivación como marca establecen valores en función del cumplimiento de una condición *única* del registro actual, pero los nodos de derivación de estado pueden cambiar los valores de un campo dependiendo del modo en que cumplen *dos condiciones independientes*. Esto quiere decir que el valor cambiará (activado o desactivado) según se cumplan las condiciones.

Figura 4-49  
Uso de un nodo de derivación como estado

**Estado inicial.** Permite seleccionar si se va a dar a cada registro del nuevo campo el valor Activado o Desactivado en un principio. Tenga en cuenta que este valor puede cambiar cada condición que se cumpla.

**Valor “Activado”.** Permite especificar el valor del campo nuevo cuando se cumple la condición Activado.

**Cambiar a “Activado” cuando.** Permite especificar una condición CLEM que cambia el estado a Activado cuando la condición es verdadera. Pulse en el botón de calculadora para abrir el generador de expresiones.

**Valor “Desactivado”.** Permite especificar el valor del campo nuevo cuando se cumple la condición Desactivado.

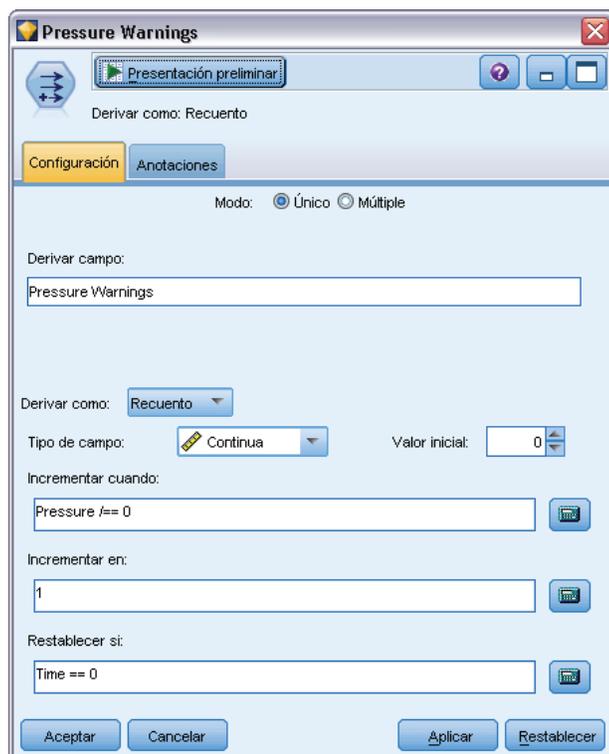
**Cambiar a "Desactivado" cuando.** Permite especificar una condición CLEM que cambia el estado a Desactivado cuando la condición sea falsa. Pulse en el botón de calculadora para abrir el generador de expresiones.

*Nota:* para especificar una cadena vacía, es necesario escribir comillas de apertura y cierre sin nada entre ellas, como "". De igual manera, las comillas deben utilizarse si desea un valor de cadena que, de otro modo, se trataría como número.

### Opciones de configuración de derivación como recuento

Los nodos de derivación como recuento se utilizan para aplicar una serie de condiciones a los valores de un campo numérico en el conjunto de datos. Al cumplirse cada condición, el valor del campo de recuento derivado aumenta en un incremento determinado. Este tipo de nodo Derivar resulta útil para los datos de series temporales.

Figura 4-50  
Opciones de recuento del cuadro de diálogo del nodo Derivar



**Valor inicial.** Establece un valor que utilizar en la ejecución del nuevo campo. El valor inicial debe ser una constante numérica. Utilice los botones de flecha para aumentar o reducir el valor.

**Incrementar cuando.** Permite especificar la condición CLEM que, de cumplirse, cambia el valor derivado en función del número especificado en Incrementar en. Pulse en el botón de calculadora para abrir el generador de expresiones.

**Incrementar en.** Permite especificar el valor utilizado para aumentar el recuento. Puede utilizar una constante numérica o el resultado de una expresión CLEM.

**Restablecer si.** Permite especificar una condición que, de cumplirse, restablece el valor derivado al valor inicial. Pulse en el botón de calculadora para abrir el generador de expresiones.

### ***Opciones de configuración de derivación como condicional***

Los nodos de derivación como condicional utilizan una serie de instrucciones Si-Entonces-En caso contrario para derivar el valor del campo nuevo.

Figura 4-51

*Uso de un nodo de derivación como condicional*

The screenshot shows the 'PowerChange' configuration window. At the top, it says 'Derivar como: Condicional'. Below this, there are two tabs: 'Configuración' (selected) and 'Anotaciones'. Under 'Configuración', there are radio buttons for 'Modo: Único' (selected) and 'Múltiple'. The 'Derivar campo:' field contains 'PowerChange'. The 'Derivar como:' dropdown is set to 'Condicional'. The 'Tipo de campo:' dropdown is set to '<Por defecto>'. The 'Si:' field contains 'Time <= 5'. The 'Entonces:' field contains '0.0'. The 'En caso contrario:' field contains '@AVE(PowerInc, 5)'. At the bottom, there are buttons for 'Aceptar', 'Cancelar', 'Aplicar', and 'Restablecer'.

**Si.** Permite especificar una condición CLEM que se evaluará en cada registro tras la ejecución. Si la condición es verdadera (o no falsa, en el caso de los números), el nuevo campo tomará el valor especificado a continuación por la expresión Entonces. Pulse en el botón de calculadora para abrir el generador de expresiones.

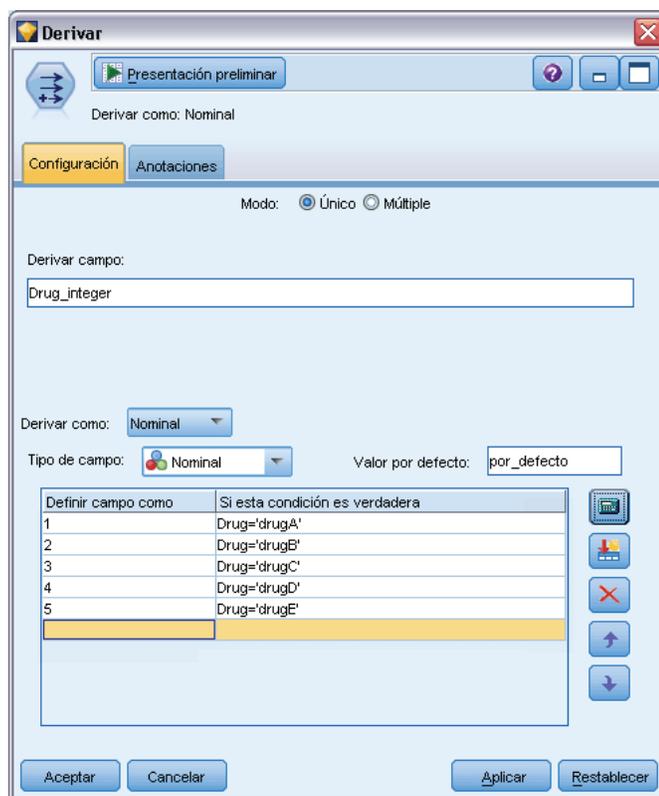
**Entonces.** Permite especificar un valor o expresión CLEM para el nuevo campo cuando la instrucción Si anterior es verdadera (o no falsa). Pulse en el botón de calculadora para abrir el generador de expresiones.

**En caso contrario.** Permite especificar un valor o expresión CLEM para el nuevo campo cuando la instrucción Si anterior es falsa. Pulse en el botón de calculadora para abrir el generador de expresiones.

## Nueva codificación de valores con el nodo Derivar

Los nodos Derivar sirven también para recodificar valores, por ejemplo, convirtiendo un campo de cadena con valores categóricos en un campo nominal (conjunto) numérico.

Figura 4-52  
Nueva codificación de valores de cadena

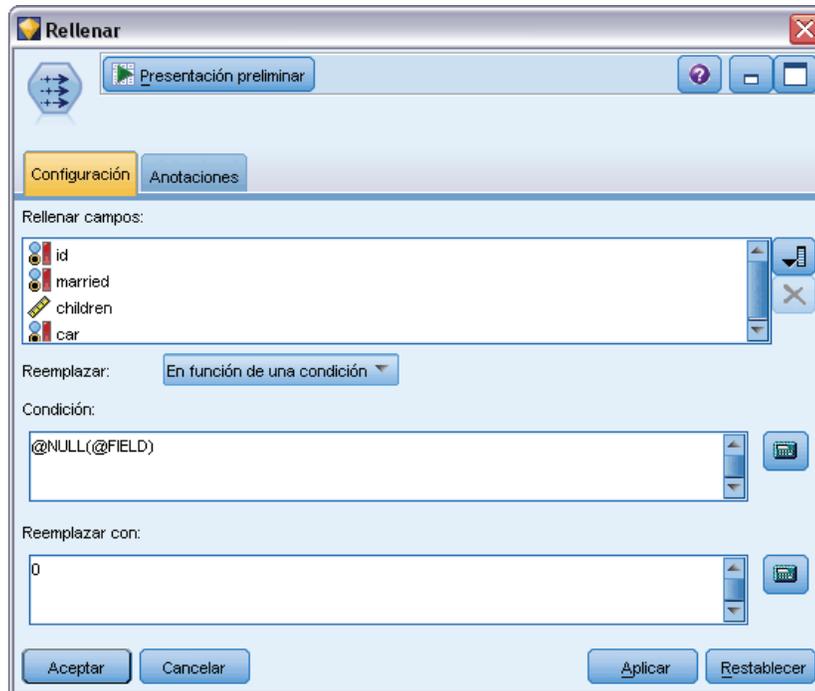


- ▶ En Derivar como, seleccione el tipo de campo (Nominal, Marca, etc.) que proceda.
- ▶ Especifique las condiciones para recodificar los valores. Por ejemplo, puede establecer el valor a 1 si Drug='drugA', 2 si Drug='drugB', etc.

## Nodo Rellenar

Los nodos Rellenar se utilizan para sustituir valores de campos y cambiar el almacenamiento. Puede sustituir los valores en función de una condición CLEM especificada, como @BLANK(FIELD). También puede sustituir todos los espacios vacíos o valores nulos por un valor específico. Los nodos Rellenar suelen utilizarse junto con el nodo Tipo para sustituir valores perdidos. Por ejemplo, puede rellenar espacios vacíos con el valor medio de un campo especificando una expresión del tipo @GLOBAL\_MEAN. Esta expresión rellenará todos los espacios vacíos con el valor medio tal y como lo calculó un nodo de valores globales.

Figura 4-53  
Cuadro de diálogo del nodo Rellenar



**Rellenar campos.** El selector de campos (botón situado a la derecha del campo de texto) permite seleccionar campos del conjunto de datos cuyos valores se examinarán y sustituirán. El comportamiento por defecto consiste en sustituir los valores en función de la condición y por las expresiones especificadas a continuación. También puede seleccionar un método alternativo de sustitución con las opciones de sustitución que se muestran a continuación.

*Nota:* al seleccionar varios campos para sustituirlos por un valor definido por el usuario, es importante que los tipos de campo sean similares (todos numéricos o simbólicos).

**Reemplazar.** Seleccione esta opción para reemplazar los valores de los campos seleccionados mediante uno de estos métodos:

- **En función de una condición.** Esta opción activa el campo Condición y el generador de expresiones para crear una expresión utilizada como una condición de sustitución por el valor especificado.
- **Siempre.** Sustituye todos los valores del campo seleccionado. Por ejemplo, podría utilizar esta opción para convertir el almacenamiento de los ingresos en una cadena mediante la siguiente expresión CLEM: (to\_string(income)).
- **Valor vacío.** Sustituye todos los valores vacíos especificados por el usuario en el campo seleccionado. La condición estándar @BLANK(@FIELD) se utiliza para seleccionar valores vacíos. *Nota:* puede definir valores vacíos mediante la pestaña Tipos del nodo de fuente o con un nodo Tipo.

- **Valores nulos.** Sustituye todos los valores nulos del sistema del campo seleccionado. La condición estándar @NULL(@FIELD) se utiliza para seleccionar valores nulos.
- **Valores vacíos y nulos.** Sustituye los valores vacíos y los valores nulos del sistema del campo seleccionado. Esta opción es útil cuando no se está seguro de si los valores nulos se han definido como valores perdidos.

**Condición.** Esta opción está disponible cuando se ha seleccionado la opción En función de una condición. Utilice este cuadro de texto para especificar una expresión CLEM para evaluar los campos seleccionados. Pulse en el botón de calculadora para abrir el generador de expresiones.

**Reemplazar con.** Permite especificar una expresión CLEM para dar un nuevo valor a los campos seleccionados. También puede reemplazar el valor por un valor nulo escribiendo undef en el cuadro de texto. Pulse en el botón de calculadora para abrir el generador de expresiones.

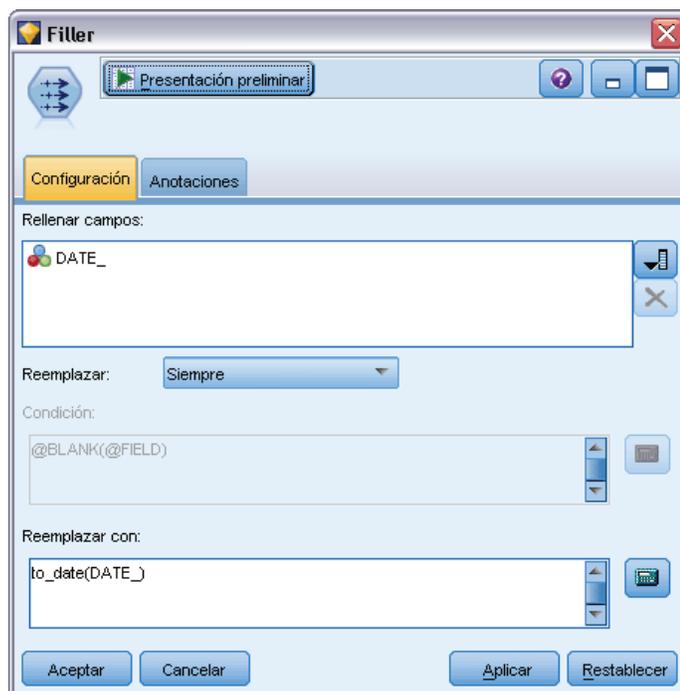
*Nota:* cuando los campos seleccionados son cadenas, debe reemplazarlos con un valor de cadena. Al utilizar el valor por defecto 0 u otro valor numérico como valor de sustitución para los campos de cadena, se producirá un error.

## Conversión del almacenamiento mediante el nodo Rellenar

Al utilizar la condición Reemplazar de un nodo Rellenar, puede convertir fácilmente el almacenamiento del campo de uno o varios campos. Por ejemplo, al utilizar la función de conversión to\_integer, se puede convertir la cadena *ingreso* en valor entero con la siguiente expresión CLEM: to\_integer(income).

Figura 4-54

Uso de un nodo Rellenar para convertir almacenamiento de campos



El generador de expresiones permite ver funciones de conversión disponibles y crear automáticamente una expresión CLEM. En la lista desplegable de funciones, seleccione Conversión para ver una lista de funciones de conversión de almacenamiento. Las funciones de conversión disponibles son:

- `to_integer(ITEM)`
- `to_real(ITEM)`
- `to_number(ITEM)`
- `to_string(ITEM)`
- `to_time(ITEM)`
- `to_timestamp(ITEM)`
- `to_date(ITEM)`
- `to_datetime(ITEM)`

**Conversión de valores de fecha y hora.** Tenga en cuenta que las funciones de conversión (y muchas otras funciones que requieren un tipo específico de entrada como un valor de fecha y hora) dependen de los formatos actuales especificados en el cuadro de diálogo Opciones de ruta. Por ejemplo, si desea convertir un campo de cadena con los valores *Ene 2003*, *Feb 2003*, etc. en un almacenamiento de fecha, seleccione MES AAAA como formato de fecha por defecto de la ruta.

Las funciones de conversión también están disponibles en el nodo Derivar para la conversión temporal durante un cálculo de derivación. También puede utilizar el nodo Derivar para realizar otras manipulaciones, como volver a codificar los campos con valores categóricos. Si desea obtener más información, consulte el tema [Nueva codificación de valores con el nodo Derivar](#) el p. 177.

## ***Nodo Anonimizar***

El nodo Anonimizar permite ocultar nombres de campo, valores de campo o ambos, cuando se trabaja con datos que se van a incluir en un modelo que se encuentra en un punto posterior de la ruta al nodo. De esta forma, el modelo generado se puede distribuir libremente (por ejemplo, al servicio de asistencia técnica) sin peligro de que usuarios no autorizados puedan acceder a datos confidenciales como registros de empleados o registros médicos de pacientes.

Según dónde coloque el nodo Anonimizar en la ruta, puede que necesite realizar cambios en otros nodos. Por ejemplo, si inserta un nodo Anonimizar antes de un nodo Seleccionar, será necesario cambiar los criterios de selección del nodo Seleccionar para que actúen sobre los valores que ahora han sido anonimizados.

El método que se debe utilizar para la anonimización depende de varios factores. En los nombres de campos y en todos los valores de campo, excepto los de niveles de medición continuo, los datos se sustituyen por una ruta con el siguiente formato:

*prefix\_Sn*

donde *prefijo\_* es, bien una cadena especificada por el usuario, o bien la cadena por defecto *anon\_*, y *n* es un valor entero que empieza en 0 y aumenta en cada valor único (por ejemplo, *anon\_S0*, *anon\_S1*, etc.)

Los valores de campos de tipo Continuo deben transformarse, ya que los rangos numéricos trabajan con valores enteros o reales en lugar de cadenas. Como tal, pueden anonimizarse transformando solamente el rango en otro diferente, ocultando de esta forma los datos originales. La transformación de un valor  $x$  en el rango se realiza de la siguiente forma:

$$A*(x + B)$$

donde:

$A$  es un factor de escala que debe ser mayor que 0.

$B$  es un desplazamiento de traducción que se suma a los valores.

### Ejemplo

En el caso del campo *EDAD*, donde el factor de escala  $A$  está definido como 7 y el desplazamiento de traducción  $B$  como 3, los valores para *EDAD* se transforman en:

$$7*(EDAD + 3)$$

## Opciones de configuración del nodo Anonimizar

Aquí, puede seleccionar los campos cuyos valores se van a ocultar a partir de ese punto de la ruta.

Recuerde que los campos se deben instanciar antes del nodo Anonimizar para que se puedan realizar las operaciones de anonimización. Para instanciar los datos, pulse en el botón Leer valores de un nodo Tipo o en la pestaña Tipos de un nodo de fuente.

Figura 4-55  
Configuración de las opciones de anonimización



**Campo.** Muestra los campos del conjunto de datos actual. Si ya se han anonimizado algunos nombres de campo, se muestran los nombres anonimizados.

**Medida.** El nivel de medición del campo.

**Anonimizar valores.** Seleccione uno o más campos, pulse en esta columna y elija Sí para anonimizar el valor del campo mediante el prefijo por defecto anon\_; elija Especificar para mostrar un cuadro de diálogo donde podrá escribir su propio prefijo o, en el caso de valores de campo de tipo *Continuo*, especificar si la transformación de los valores de campo va a utilizar valores aleatorios o especificados por el usuario. Tenga en cuenta que no se pueden especificar en la misma operación los tipos de campo *Continuo* y no *Continuo*, sino que deberá hacerlo por separado para cada tipo de campo.

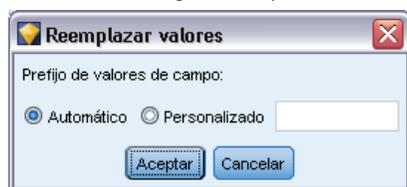
**Ver campos actuales.** Seleccione esta opción para ver los campos de los conjuntos de datos conectados de forma activa con el nodo Anonimizar. Esta opción está seleccionada por defecto.

**Ver configuración de campos no utilizados.** Seleccione esta opción para ver los campos de los conjuntos de datos que han estado conectados con anterioridad al nodo pero ya no lo están. Esta opción resulta útil al copiar nodos de una ruta a otra o cuando se guardan y se vuelven a cargar nodos.

### ***Especificación del método de anonimización de los valores de campo***

El cuadro de diálogo Reemplazar valores permite elegir si se va a utilizar el prefijo por defecto para los valores de campo anonimizados o un prefijo personalizado. Al pulsar en Aceptar en este cuadro de diálogo, la opción Anonimizar valores de la pestaña Configuración cambia a Sí para el campo o campos seleccionados.

Figura 4-56  
Cuadro de diálogo Reemplazar valores



**Prefijo de valores de campo.** El prefijo por defecto para los valores de campo anonimizados es anon\_; seleccione Personalizado y escriba su propio prefijo si desea uno diferente.

El cuadro de diálogo Transformar valores se muestra solamente para los campos de tipo Continuo y permite especificar si en la transformación de los valores de campo se van a utilizar valores aleatorios o especificados por el usuario.

Figura 4-57  
Cuadro de diálogo Transformar valores



**Aleatorios.** Seleccione esta opción para utilizar valores aleatorios para la transformación. Establecer semilla aleatoria está seleccionado por defecto; especifique un valor en el campo Semilla o utilice el valor por defecto.

**Fija.** Seleccione esta opción para especificar sus propios valores aleatorios para la transformación.

- **Escalar por.** El número por el que se multiplicarán los valores de campo en la transformación. El valor mínimo es 1 y el máximo suele ser 10, pero puede reducirse para evitar desbordamientos.
- **Trasladar por.** El número que se sumará a los valores de campo en la transformación. El valor mínimo es 0 y el máximo suele ser 1000, pero puede reducirse para evitar desbordamientos.

### ***Anonimización de valores de campo***

Los campos seleccionados para la anonimización en la pestaña Configuración tienen sus valores anonimizados:

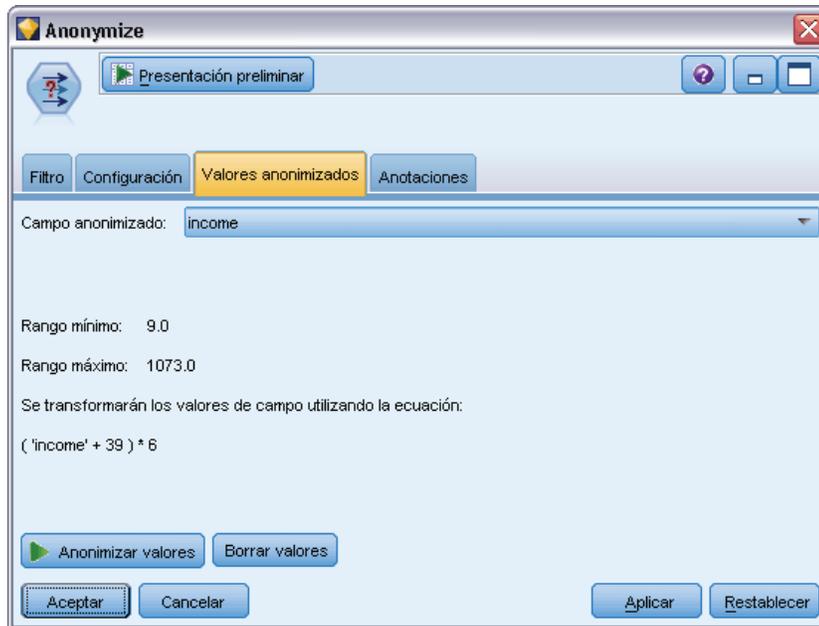
- cuando se ejecuta la ruta que contiene el nodo Anonimizar
- cuando se previsualizan los valores

Para ver la presentación preliminar de los valores, pulse en el botón Anonimizar valores en la pestaña Valores anonimizados. A continuación, seleccione un nombre de campo desde la lista desplegable.

Si el nivel de medición es Continuo, la pantalla muestra:

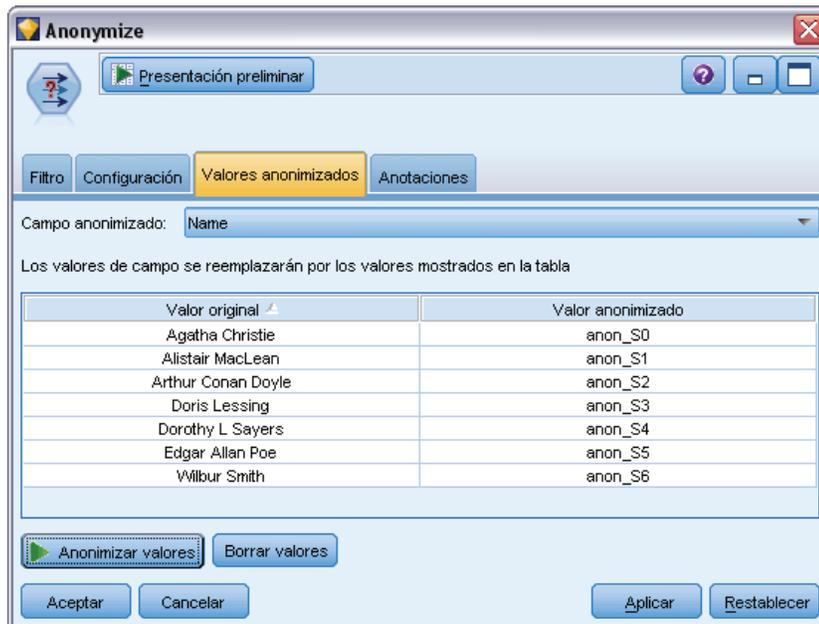
- valores mínimo y máximo del rango original
- ecuación utilizada para transformar los valores

Figura 4-58  
Anonimización de valores de campo



Si el nivel de medición es diferente de Continuo, la pantalla muestra los valores original y anonimizado de este campo.

Figura 4-59  
Anonimización de valores de campo



Cuando la visualización aparece con un fondo amarillo, indica que la configuración para el campo seleccionado ha cambiado desde la última anonimización de los valores o que se han realizado cambios a los datos en un punto de la ruta anterior al nodo Anonimizar y, por lo tanto, los valores de anonimización puede que ya no sean correctos. Se muestra el conjunto actual de valores; vuelva a pulsar en el botón Anonimizar valores para generar un nuevo conjunto de valores de acuerdo con la configuración actual.

**Anonimizar valores.** Crea valores anonimizados para el campo seleccionado y los muestra en la tabla. Si está utilizando semillas aleatorias para un campo de tipo Continuo, pulse varias veces en este botón para crear un conjunto de valores diferente cada vez.

**Borrar valores.** Borra los valores originales y anonimizados de la tabla.

## Nodo Reclasificar

El nodo Reclasificar permite la transformación desde un conjunto de valores categóricos a otro. La reclasificación es útil para contraer categorías o reagrupar datos para su análisis. Por ejemplo, se pueden reclasificar los valores *Producto* en tres grupos, como por ejemplo *Utensilios de cocina*, *Baño y ropa de cama* y *Electrodomésticos*. A menudo, esta operación se realiza directamente desde un nodo de distribución agrupando valores y generando un nodo Reclasificar. Si desea obtener más información, consulte el tema [Utilización de un nodo Distribución](#) en el capítulo 5 el p. 311.

La reclasificación puede realizarse mediante uno o varios campos simbólicos. También puede sustituir los nuevos valores por el campo existente o generar un campo nuevo.

Antes de utilizar un nodo Reclasificar, tenga en cuenta si existe otro nodo de operaciones con campos que sea más adecuado para la tarea en cuestión:

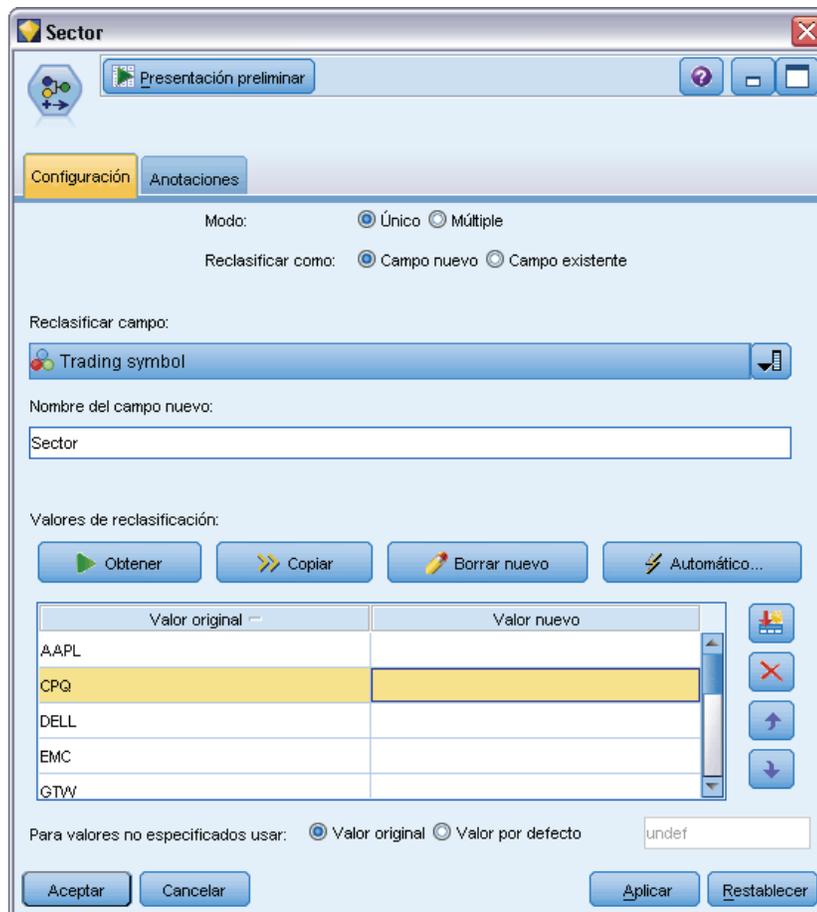
- Para transformar rangos numéricos en conjuntos con un método automático, como por ejemplo rangos o percentiles, se debe utilizar un nodo Intervalos. Si desea obtener más información, consulte el tema [Nodo Intervalos](#) el p. 189.
- Para clasificar rangos numéricos en conjuntos manualmente, se debe utilizar un nodo Derivar. Por ejemplo, si desea contraer valores de salario en categorías de rango de salario específicas, se debe utilizar un nodo Derivar para definir cada categoría manualmente.
- Para crear uno o más campos de marcas en función de los valores de un campo categórico, como por ejemplo *Mortgage\_type*, se debe utilizar un nodo Marcas.
- Para convertir un campo categórico en almacenamiento numérico, puede usar un nodo Derivar. Así, podría convertir los valores *No* y *Sí* a 0 y 1 respectivamente. Si desea obtener más información, consulte el tema [Nueva codificación de valores con el nodo Derivar](#) el p. 177.

## Opciones de configuración del nodo Reclasificar

Hay tres pasos para utilizar el nodo Reclasificar:

- ▶ Primero, seleccione si desea reclasificar varios campos o un campo único.
- ▶ A continuación, elija si desea recodificar en el campo existente o crear un campo nuevo.
- ▶ A continuación, utilice las opciones dinámicas del cuadro de diálogo del nodo Reclasificar para establecer las correspondencias de conjuntos que desee.

Figura 4-60  
Cuadro de diálogo del nodo Reclasificar



**Modo.** Seleccione Único para reclasificar las categorías de un campo. Seleccione Múltiple para activar las opciones que permiten la transformación de varios campos al mismo tiempo.

**Reclasificar como.** Seleccione Campo nuevo para conservar el campo nominal original y derivar un campo adicional que contenga los valores clasificados. Seleccione Campo existente para sobrescribir los valores en el campo original con las nuevas clasificaciones. Se trata fundamentalmente de una operación de “relleno”.

Una vez especificado el modo y las opciones de sustitución, debe seleccionar el campo de transformación y especificar los nuevos valores de clasificación mediante las opciones dinámicas de la mitad inferior del cuadro de diálogo. Estas opciones varían en función del modo seleccionado con anterioridad.

**Reclasificar campos.** Utilice el botón de selección de campos de la parte derecha para seleccionar un campo discreto (modo Único) o más campos categóricos (modo Múltiple).

**Nombre del campo nuevo.** Permite especificar un nombre para el nuevo campo nominal que contiene valores recodificados. Esta opción está disponible solamente en modo Único si se ha seleccionado Campo nuevo con anterioridad. Cuando Campo existente está seleccionado, el

nombre del campo original se conserva. Al trabajar en modo Múltiple, esta opción se sustituye por controles para especificar una extensión añadida a cada nuevo campo. Si desea obtener más información, consulte el tema [Reclasificación de múltiples campos](#) el p. 188.

**Valores de reclasificación.** Esta tabla permite establecer una correspondencia clara entre los antiguos valores del conjunto y los que se especifican aquí.

- **Valor original.** Esta columna enumera los valores existentes para los campos seleccionados.
  - **Valor nuevo.** Utilice esta columna para escribir nuevos valores de categoría o seleccionar uno de la lista desplegable. Al generar automáticamente un nodo Reclasificar con los valores del cuadro Distribución, estos valores se incluyen en la lista desplegable. Esto permite establecer con rapidez una correspondencia entre los valores existentes y un conjunto de valores conocido. Por ejemplo, en ocasiones las organizaciones de salud agrupan los diagnósticos de forma distinta en función de la configuración de red o local. Después de una fusión o una adquisición, todas las partes deben reclasificar los datos nuevos e incluso los existentes de forma coherente. En lugar de escribir manualmente cada valor objetivo de una larga lista, puede leer la lista maestra de valores en IBM® SPSS® Modeler, ejecutar un cuadro Distribución para el campo *Diagnóstico* y generar un nodo Reclasificar (valores) para este campo directamente desde el cuadro. Este proceso hará que todos los valores objetivo de Diagnóstico estén disponibles en la lista desplegable Valores nuevos.
- ▶ Pulse en Obtener para leer los valores originales de uno o varios campos seleccionados con anterioridad.
  - ▶ Pulse en Copiar para pegar valores originales en la columna *Valor nuevo* en el caso de los campos en los que aún no ha establecido ninguna correspondencia. Los valores originales sin correspondencia se añaden a la lista desplegable.
  - ▶ Pulse en Borrar nuevo para borrar todas las especificaciones de la columna *Valor nuevo*. *Nota:* esta opción no elimina los valores de la lista desplegable.
  - ▶ Pulse en Automático para generar automáticamente números enteros consecutivos para cada uno de los valores originales. Solamente se pueden generar valores enteros (no valores reales como 1,5 o 2,5).

Figura 4-61

Cuadro de diálogo de clasificación automática



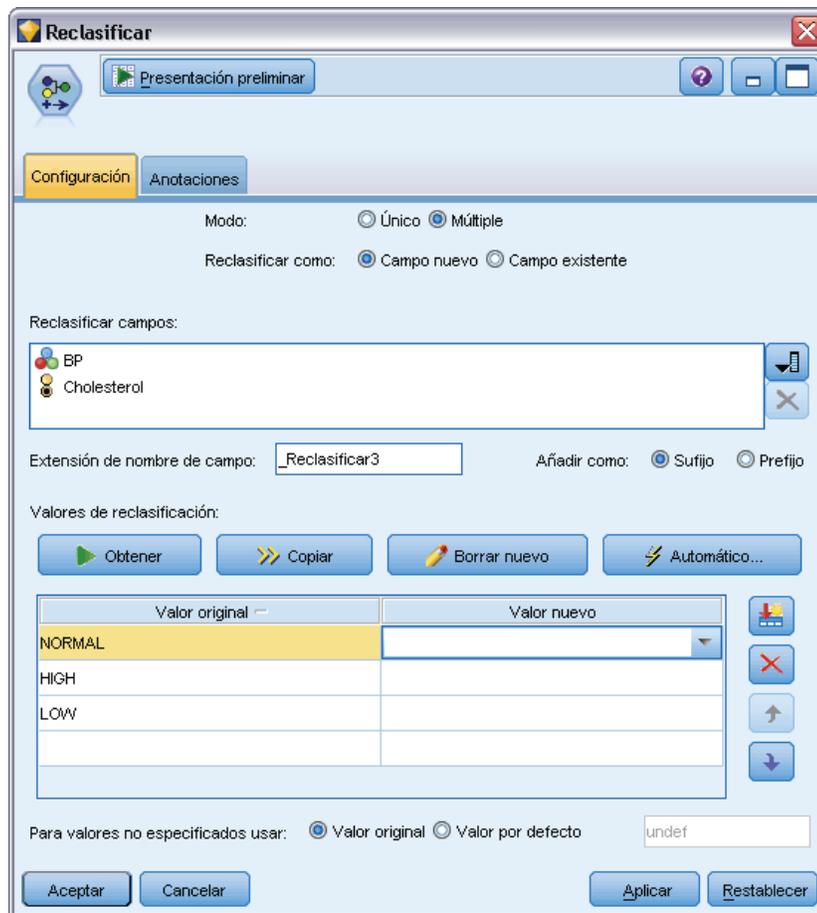
Por ejemplo, puede generar automáticamente números de ID de producto consecutivos para los nombres de productos o números de cursos para ofertas de clasificación universitaria. Esta función se corresponde con la transformación de la recodificación automática para los conjuntos de IBM® SPSS® Statistics.

**Para valores no especificados usar.** Esta opción se utiliza para rellenar valores no especificados del campo nuevo. También puede conservar el valor original seleccionando Valor original o especificar un valor por defecto.

## Reclasificación de múltiples campos

Para establecer correspondencias de valores de categorías para más de un campo a la vez, active el modo Múltiple. Esto permite establecer nuevas configuraciones en el cuadro de diálogo Reclassificar, que se describen a continuación.

Figura 4-62  
Opciones del cuadro de diálogo dinámico para reclasificar múltiples campos



**Reclasificar campos.** Utilice el botón de selección de campos de la parte derecha para seleccionar los campos que desea transformar. El botón de selección de campos permite seleccionar todos los campos a la vez o campos de tipos similares, como nominal o marca.

**Extensión de nombre de campo.** Al recodificar múltiples campos simultáneamente, es más eficaz especificar una extensión común añadida a los todos los campos nuevos en lugar de los nombres de campos individuales. Especifique una extensión del tipo `_recode` y seleccione si desea añadir esta extensión antes o después de los nombres de campos originales.

## Almacenamiento y nivel de medición de los campos reclasificados

El nodo Reclasificar siempre crea un campo nominal desde la operación de recodificación. En algunos casos, esto puede cambiar el nivel de medición de campo al utilizar el modo de reclasificación Campo existente.

El almacenamiento de los campos nuevos (la forma de *almacenar* los datos en lugar de cómo se *utilizan*) se calcula dependiendo de las siguientes opciones de la pestaña Configuración:

- Si se establece que los valores sin especificar utilicen un valor por defecto, el tipo de almacenamiento vendrá determinado al examinar los nuevos valores y el valor por defecto y determinar el almacenamiento adecuado. Por ejemplo, si todos los valores se pueden analizar como números enteros, el campo tendrá un tipo de almacenamiento entero.
- Si los valores no especificados se configuran para utilizar valores originales, el tipo de almacenamiento se basará en el almacenamiento del campo original. Si todos los valores pueden analizarse como el almacenamiento del campo original, dicho almacenamiento se conservará; de lo contrario, el almacenamiento vendrá determinado por el tipo de almacenamiento más adecuado que cubra tanto valores antiguos como nuevos. Por ejemplo, al reclasificar un conjunto de números enteros { 1, 2, 3, 4, 5 } con la reclasificación  $4 \Rightarrow 0$ ,  $5 \Rightarrow 0$  se genera un conjunto de números enteros nuevos { 1, 2, 3, 0 }, mientras que con la reclasificación  $4 \Rightarrow$  “Sobre 3”,  $5 \Rightarrow$  “Sobre 3” se genera el conjunto de cadena { “1”, “2”, “3”, “Sobre 3” }.

*Nota:* si el tipo original estaba sin instanciar, el nuevo tipo será también sin instanciar.

## Nodo Intervalos

El nodo Intervalos permite crear automáticamente nuevos campos nominales en función de los valores de uno o varios campos continuos (rango numérico) existentes. Por ejemplo, puede transformar un campo de ingresos continuo en un campo categórico nuevo que contenga grupos de ingresos con el mismo ancho, o como desviaciones desde la media. Si lo prefiere, puede seleccionar un campo categórico “supervisor” para conservar la fuerza de la asociación original entre los dos campos.

Los intervalos pueden resultar útiles por varias razones, entre ellas:

- **Requisitos de algoritmos.** Algunos algoritmos como Bayesiano Naive y Regresión logística requieren entradas categóricas.
- **Rendimiento.** El rendimiento de algoritmos como los de regresión logística multinomial es mayor si se reduce el número de valores distintos de los campos de entrada. Por ejemplo, utilice el valor de la media o la mediana para cada intervalo en lugar de los valores originales.
- **Privacidad de los datos.** La información personal confidencial, como los salarios, se puede registrar en rangos en lugar de cifras salariales reales para proteger la privacidad.

Existen varios métodos de agrupación en intervalos disponibles. Una vez creados los intervalos para el nuevo campo, puede generar un nodo Derivar a partir de los puntos de corte.

Antes de utilizar un nodo Intervalos, considere si existe otra técnica más adecuada para la tarea en cuestión:

- Para especificar manualmente los puntos de corte para las categorías, como rangos de salario predefinidos específicos, utilice un nodo Derivar. Si desea obtener más información, consulte el tema [Nodo Derivar](#) el p. 165.
- Para crear nuevas categorías para conjuntos existentes, utilice un nodo Reclasificar. Si desea obtener más información, consulte el tema [Nodo Reclasificar](#) el p. 185.

### ***Gestión de valores perdidos***

El nodo Intervalos gestiona valores perdidos de las siguientes formas:

- **Elementos vacíos especificados por el usuario.** Los valores perdidos especificados como valores vacíos se incluyen durante la transformación. Por ejemplo, si designó -99 para indicar un valor vacío usando el nodo Tipo, este valor se incluirá en el proceso de intervalos. Para omitir los valores vacíos durante el intervalo, se debe usar el nodo Rellenar, que permite sustituir valores vacíos por valores nulos del sistema.
- **Valores perdidos del sistema (\$null\$).** Los valores nulos se omiten durante la transformación de intervalos y permanecen así tras la transformación.

La pestaña Configuración proporciona opciones para las técnicas disponibles. La pestaña Ver muestra puntos de corte establecidos para los datos ejecutados previamente en el nodo.

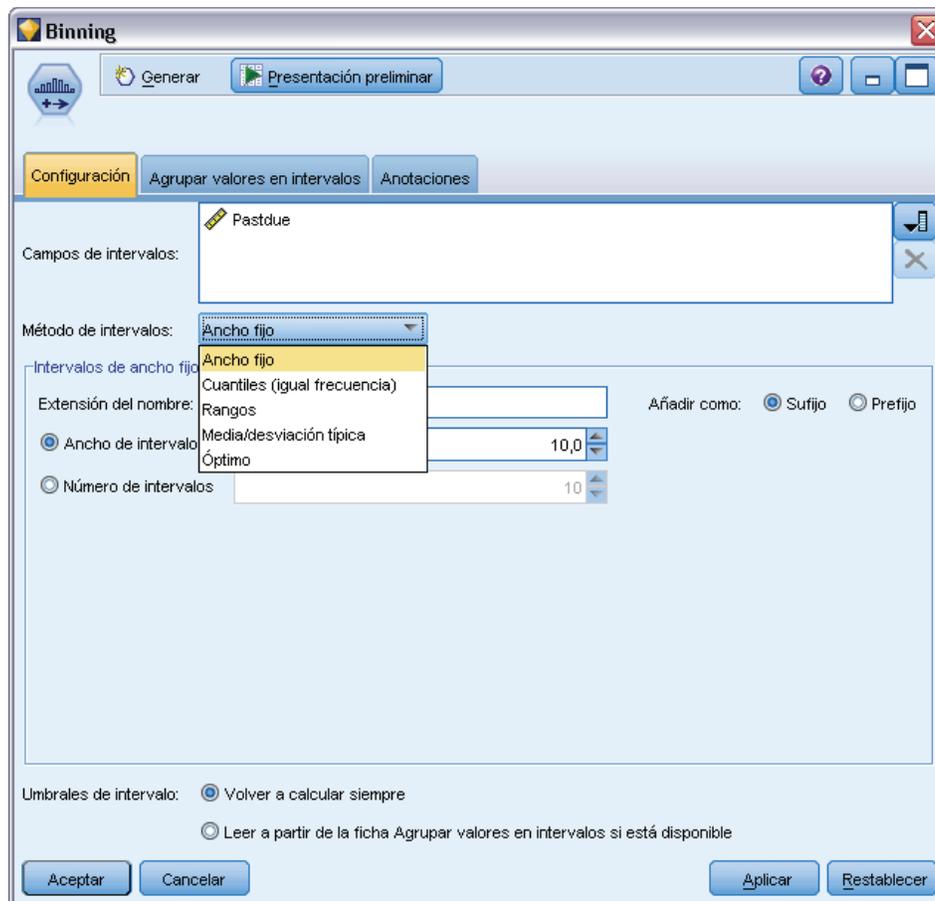
### ***Opciones de configuración del nodo Intervalos***

Con el nodo Intervalos puede generar intervalos (categorías) automáticamente mediante las siguientes técnicas:

- Intervalos de ancho fijo
- Cuantiles (igual frecuencia o igual suma)
- Media y desviación típica
- Rangos
- Optimizado respecto a un campo categórico “supervisor”

La mitad inferior del cuadro de diálogo cambia de forma dinámica en función del método de intervalos seleccionado.

Figura 4-63  
Cuadro de diálogo del nodo Intervalos, pestaña Configuración



**Campos de intervalos.** Los campos continuos (rango numérico) pendientes de transformación se muestran aquí. El nodo Intervalos permite segmentar varios campos simultáneamente. Añada campos utilizando los botones de la derecha.

**Método de intervalos.** Permite seleccionar el método utilizado para determinar los puntos de corte de los intervalos de campo nuevos (categorías). Los temas siguientes describen las opciones disponibles en cada caso.

**Umbral de intervalo.** Especifica cómo se calculan los umbrales de intervalo.

- **Recalcular siempre.** Los puntos de corte y las asignaciones de umbrales siempre se recalculan cuando se ejecuta el nodo.
- **Leer de la pestaña Agrupar valores en intervalos si está disponible.** Los puntos de corte y las asignaciones de umbrales se calculan únicamente cuando sea necesario (por ejemplo, cuando se han añadido nuevos datos).

Los siguientes temas detallan las opciones de los métodos de intervalos disponibles.

## Intervalos de ancho fijo

Al seleccionar Ancho fijo como método de intervalos, se muestra un conjunto de opciones en el cuadro de diálogo.

Figura 4-64

Cuadro de diálogo del nodo Intervalos (pestaña Configuración) con opciones para los intervalos de ancho fijo

**Extensión del nombre.** Permite especificar una extensión que utilizar al generar campos. *\_INT* es la extensión por defecto. También puede determinar si la extensión se debe añadir al principio (prefijo) o al final (sufijo) del nombre de campo. Por ejemplo, puede generar un campo nuevo denominado *ingresos\_INT*.

**Ancho de intervalo.** Permite especificar un valor (entero o real) para calcular el “ancho” del intervalo. Por ejemplo, puede utilizar el valor por defecto 10 para segmentar el campo *Edad*. Como *Edad* tiene un rango de 18 a 65, los intervalos generados serían los siguientes:

Tabla 4-1

Intervalos para *Edad* con el rango 18 a 65

Intervalo 1	Intervalo 2	Intervalo 3	Intervalo 4	Intervalo 5	Intervalo 6
$\geq 13$ a $< 23$	$\geq 23$ a $< 33$	$\geq 33$ a $< 43$	$\geq 43$ a $< 53$	$\geq 53$ a $< 63$	$\geq 63$ a $< 73$

El inicio de los intervalos se calcula mediante el valor menor explorado menos la mitad del ancho de intervalo (tal y como se especifica). Por ejemplo, en los intervalos del ejemplo, el 13 se utiliza para iniciar los intervalos según el siguiente cálculo:  $18 [\text{menor valor de datos}] - 5 [0,5 \times (\text{ancho de intervalo de } 10)] = 13$ .

**Número de intervalos.** Utilice esta opción para especificar un número entero para determinar el número de intervalos de ancho fijo (categorías) para los nuevos campos.

Una vez ejecutado el nodo Intervalos en una ruta, puede ver los umbrales de intervalo generados pulsando en la pestaña Presentación preliminar del cuadro de diálogo del nodo Intervalos. Si desea obtener más información, consulte el tema [Vista preliminar de intervalos generados](#) el p. 199.

## Cuantiles (igual frecuencia o igual suma)

Con el método de intervalos de cuantil se crean campos nominales que pueden servir para dividir registros explorados en grupos de percentiles (o cuantiles, deciles, etc.) para que, de este modo, cada grupo contenga el mismo número de registros, o bien la suma de los valores de cada uno de ellos sea la misma. Los registros se ordenan por rango en orden ascendente en función del valor del campo de intervalo especificado, y así, a los registros con los valores más bajos de la variable de intervalo seleccionada se les asigna un rango de 1, al siguiente conjunto de registros un rango de 2, y así sucesivamente. Los valores de umbral de cada intervalo se generan automáticamente según los datos y el método de mosaico empleados.

Figura 4-65

Cuadro de diálogo del nodo Intervalos (pestaña Configuración) con opciones para los intervalos de igual frecuencia

Intervalos de igual frecuencia

Extensión del nombre de cuantil:  Añadir como:  Sufijo  Prefijo

Nota: el recuento de cuantiles se añadirá al final de la extensión

Extensión de cuantil personalizada:  Añadir como:  Sufijo  Prefijo

Cuartil (4)  Quintil (5)  Decil (10)

Veintil (20)  Percentil (100)

Personalizar N

Método de cuantiles:  Recuento de registros  Suma de valores

Empates:  Añadir al siguiente  Mantener en el actual  Asignar aleatoriamente

**Extensión del nombre de cuantil.** Permite especificar una extensión utilizada para los campos generados con p-tilas estándar. La extensión por defecto es *\_TILE* más *N*, donde *N* es el número de cuantil. También puede determinar si la extensión se debe añadir al principio (prefijo) o al final (sufijo) del nombre de campo. Por ejemplo, puede generar un campo nuevo denominado *ingresos\_INT4*.

**Extensión de cuantil personalizada.** Permite especificar una extensión utilizada para un rango de cuantiles personalizados. El valor por defecto es *\_TILN*. Observe que, en este caso, la *Nno* se sustituye por el número personalizado.

Los p-tilas disponibles son:

- **Cuartil.** Genera 4 intervalos, cada uno con el 25% de los casos.
- **Quintil.** Genera 5 intervalos, cada uno con el 20% de los casos.
- **Decil.** Genera 10 intervalos, cada uno con el 10% de los casos.
- **Veintil.** Genera 20 intervalos, cada uno con el 5% de los casos.
- **Percentil.** Genera 100 intervalos, cada uno con el 1% de los casos.
- **Personalizar N.** Seleccione esta opción para indicar el número de intervalos. Por ejemplo, un valor de 3 produciría 3 categorías agrupadas (2 puntos de corte), cada una de las cuales contendría el 33,3% los casos.

Tenga en cuenta que si hay menos valores discretos en los datos que cantidad de cuantiles especificados, no se utilizarán todos los cuantiles. En estos casos la nueva distribución suele reflejar la distribución original de los datos.

**Método de mosaico.** Especifica el método utilizado para asignar registros a los intervalos.

- **Recuento de registros.** Trata de asignar el mismo número de registros a cada intervalo.
- **Suma de los valores.** Trata de asignar registros a intervalos de forma que la suma de los valores de cada intervalo sea la misma. Al intentar dirigir los esfuerzos de ventas, por ejemplo, este método puede servir para asignar posibles clientes a grupos de deciles en función del valor de cada registro, de forma que los posibles clientes con los valores más altos se incluyan en el

intervalo superior. Así, una empresa farmacéutica puede clasificar a los médicos en grupos de deciles en función del número de prescripciones que realizan. Si bien cada decil contendría aproximadamente el mismo número de procesos, el número de individuos que contribuyen en dichos procesos no sería el mismo, concentrándose los individuos que realizan más procesos en el decil 10. Observe que en este método se da por hecho que todos los valores son superiores a cero, de modo que puede arrojar resultados inesperados en caso de que no sea así.

**Empates.** Una condición de empate se produce cuando los valores en ambas partes de un punto de corte son idénticos. Por ejemplo, si está asignando deciles y más del 10% de los registros tienen el mismo valor para el campo de intervalo, no todos podrán encajar en el mismo intervalo a menos que se fuerce el umbral de una forma u otra. Así, los empates pueden pasar al siguiente intervalo o permanecer en el actual, pero deberán resolverse de manera que todos los registros con valores idénticos se encuentren en el mismo intervalo, aun cuando esto suponga que algunos intervalos contengan más registros de lo previsto. En consecuencia, los umbrales de los intervalos siguientes también pueden ajustarse, lo que hará que los valores se asignen de manera diferente para el mismo conjunto de números según el método empleado para solucionar el empate.

- **Añadir al siguiente.** Seleccione esta opción para mover los valores de empate hacia el siguiente intervalo superior.
- **Mantener en el actual.** Mantiene los valores de empate en el intervalo actual (más bajo). Este método puede dar como resultado un menor número total de intervalos creados.
- **Asignar aleatoriamente.** Seleccione esta opción para relacionar los valores aleatoriamente a un intervalo. Se intenta mantener así el mismo número de registros en cada intervalo.

**Ejemplo: Mosaico por recuento de registros**

La tabla que aparece a continuación ilustra cómo los valores de campo simplificados se clasifican como cuartiles cuando se usa el mosaico por recuento de registros. Tenga en cuenta que los resultados variarán en función de la opción de empate seleccionada.

Valores	Añadir al siguiente	Mantener en el actual
10	1	1
13	2	1
15	3	2
15	3	2
20	4	3

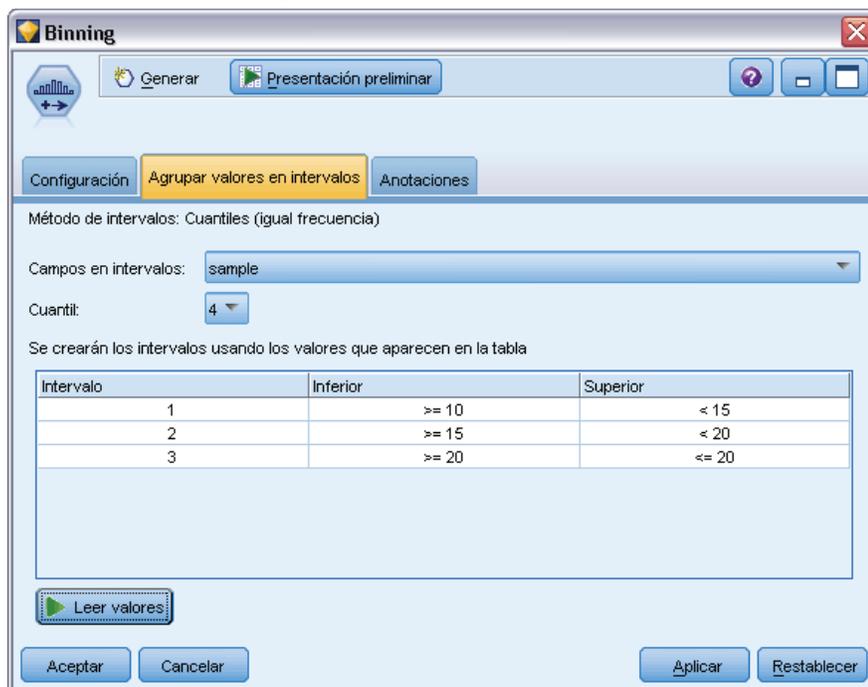
El número de elementos por intervalo se calcula así:

número total de valor / número de cuantiles

En el ejemplo simplificado anterior, el número de elementos deseado por intervalo es 1,25 (5 valores / 4 cuartiles). El valor 13 (siendo el número de valor 2) divide el umbral del recuento deseado de 1,25 y, por lo tanto, se trata de forma distinta en función de la opción de empate seleccionada. En el modo Añadir al siguiente, se añade al intervalo 2. En el modo Mantener en el actual, se deja en el intervalo 1, forzando el rango de valores del intervalo 4 fuera de los valores

de datos existentes. Como resultado, solamente se crearán tres intervalos y los umbrales para cada uno de ellos se ajustarán en consecuencia.

Figura 4-66  
Umbrals para intervalos generados

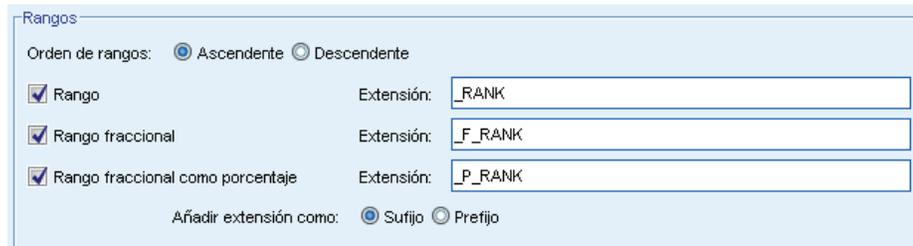


*Nota:* la velocidad de creación de intervalos por cuantiles puede ser mayor si se activa el procesamiento paralelo.

## Asignar rangos a los casos

Al seleccionar Rangos como método de intervalos, se muestra un conjunto de opciones en el cuadro de diálogo.

Figura 4-67  
Cuadro de diálogo del nodo Intervalos (pestaña Configuración) con opciones para los rangos



Los rangos crean nuevos campos que contienen rangos, rangos fraccionales y valores de percentiles para campos numéricos en función de las opciones especificadas más adelante.

**Orden de rangos.** Permite seleccionar Ascendente (el valor más bajo está marcado con 1) o Descendente (el valor más alto está marcado con 1).

**Rango.** Seleccione esta opción para establecer rangos de casos en orden ascendente o descendente tal y como se especifica anteriormente. El rango de valores del campo nuevo será  $1-N$ , donde  $N$  es el número de valores discretos del campo original. Los valores empatados toman la media de sus rangos.

**Rango fraccional.** Seleccione esta opción para establecer rangos de casos en los que el valor del campo nuevo es igual al rango dividido por la suma de las ponderaciones de los casos que no están perdidos. Los rangos fraccionales están dentro del rango de 0 a 1.

**Rango fraccional como porcentaje.** Cada rango se divide entre el número de registros con valores válidos y se multiplica por 100. Los rangos fraccionales como porcentaje están dentro del rango de 1 a 100.

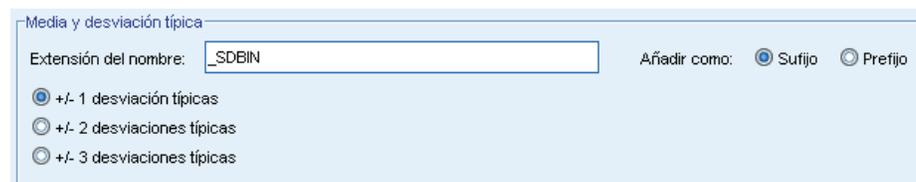
**Extensión.** En todas las opciones de rangos puede crear extensiones personalizadas y especificar si la extensión se debe añadir al principio (prefijo) o al final (sufijo) del nombre del campo. Por ejemplo, puede generar un campo nuevo denominado *ingresos\_P\_RANGO*.

## Media/desviación típica

Al seleccionar Media/desviación típica como método de intervalos, se muestra un conjunto de opciones en el cuadro de diálogo.

Figura 4-68

Cuadro de diálogo del nodo Intervalos (pestaña Configuración) con opciones para media/desviación típica



Este método genera uno o varios campos nuevos con categorías de unión en función de los valores de la media y la desviación típica de la distribución de los campos especificados. Seleccione el número de desviaciones que desee utilizar a continuación.

**Extensión del nombre.** Permite especificar una extensión que utilizar al generar campos. *\_DTINT* es la extensión por defecto. También puede determinar si la extensión se debe añadir al principio (prefijo) o al final (sufijo) del nombre de campo. Por ejemplo, puede generar un campo nuevo denominado *ingresos\_DTINT*.

- **+/- 1 desviación típica.** Seleccione esta opción para generar tres intervalos.
- **+/- 2 desviaciones típicas.** Seleccione esta opción para generar cinco intervalos.
- **+/- 3 desviaciones típicas.** Seleccione esta opción para generar siete intervalos.

Por ejemplo, al seleccionar  $\pm 1$  desviación típica, se producen tres intervalos tal y como se calcula a continuación:

Intervalo 1	Intervalo 2	Intervalo 3
$x < (\text{Mean} - \text{Std. Dev})$	$(\text{Mean} - \text{Std. Dev}) \leq x \leq (\text{Mean} + \text{Std. Dev})$	$x > (\text{Mean} + \text{Std. Dev})$

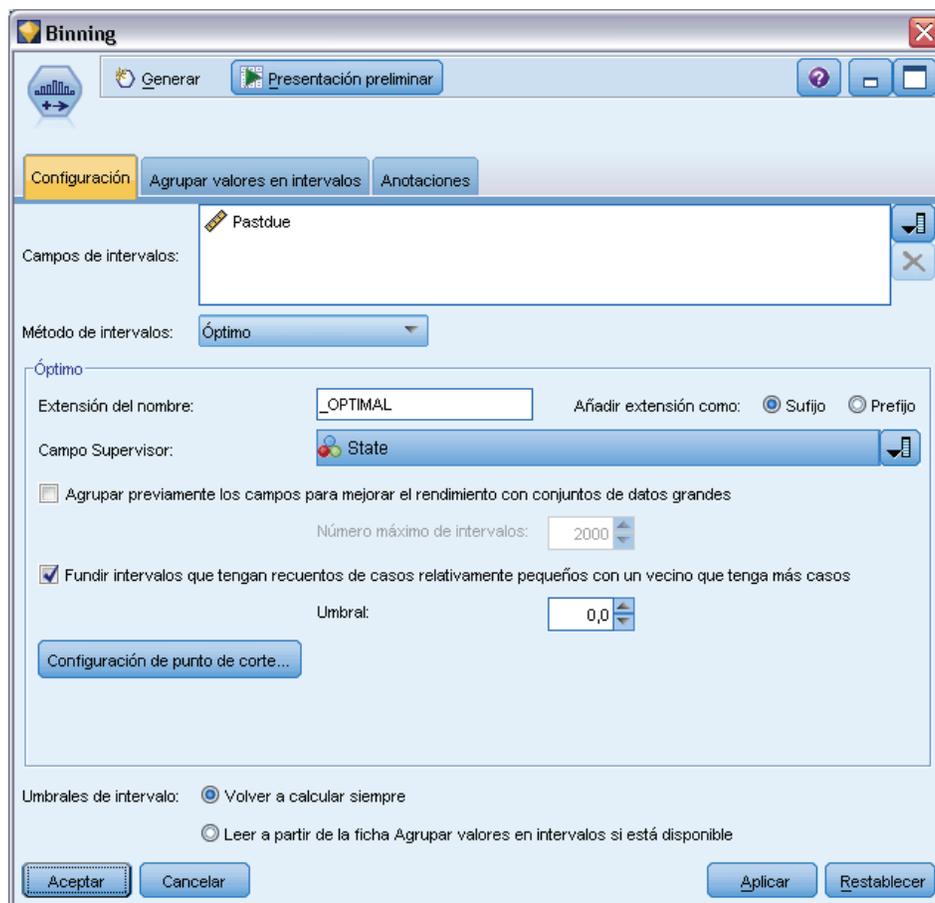
En una distribución normal, el 68% de los casos entran dentro de una desviación típica de la media, el 95% en dos desviaciones estándar y el 99% en tres desviaciones estándar. No obstante, tenga en cuenta que al crear categorías de uniones en función de las desviaciones típicas, algunos intervalos pueden definirse fuera del rango de datos real e incluso fuera del rango de los valores de datos posibles (por ejemplo, un rango de salario negativo).

### ***Intervalos óptimos***

Si el campo que desea dividir en intervalos está fuertemente asociado a otro campo categórico, puede seleccionar el campo categórico como “supervisor” para crear los intervalos de forma que se conserve la fuerza de la asociación original entre los dos campos.

Por ejemplo, suponga que ha utilizado el análisis de conglomerados para agrupar estados basados en los porcentajes de impago de préstamos hipotecarios, con los porcentajes más altos en el primer conglomerado. En este caso, podría seleccionar *Porcentaje en mora* y *Porcentaje de ejecuciones hipotecarias* como campos de intervalos, y el campo de pertenencia al conglomerado generado por el modelo, como campo supervisor.

Figura 4-69  
Opciones de los intervalos óptimos y supervisados



**Extensión del nombre.** Especifique una extensión para utilizarla en los campos generados y especifique si debe añadirse al inicio (Prefijo) o al final (Sufijo) del nombre de campo. Por ejemplo, puede generar un campo nuevo denominado *emora\_ÓPTIMO* y otro denominado *enejecución\_ÓPTIMO*.

**Campo Supervisor.** Campo categórico empleado para construir los intervalos.

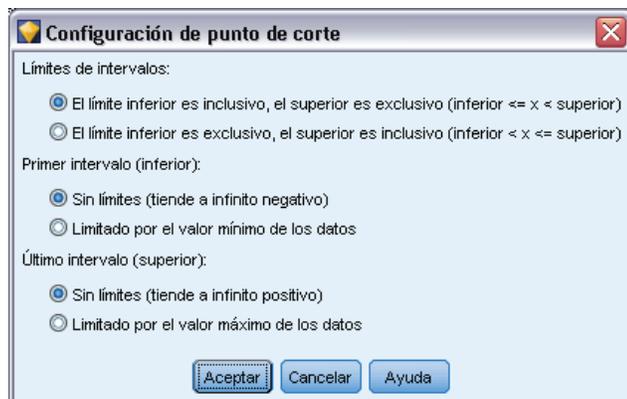
**Agrupar previamente los campos para mejorar el rendimiento con conjuntos de datos grandes.** Indica si se debe realizar un procesamiento previo para simplificar los intervalos óptimos. Este método agrupa los valores de escala en un número mayor de intervalos mediante un sencillo método de intervalos sin supervisión, representa los valores dentro de cada intervalo mediante la media y ajusta la ponderación de casos en consecuencia antes de continuar con los intervalos supervisados. En términos prácticos, este método logra un equilibrio entre el grado de precisión y la velocidad, por lo que resulta recomendable para conjuntos de datos de gran tamaño. También puede especificar el número máximo de intervalos en que cualquier variable debe terminar después del procesamiento cuando se utiliza esta opción.

**Fusiona los intervalos que tienen un recuento de casos relativamente pequeño con un vecino de mayor tamaño.** Si está activada, indica que se ha fusionado un intervalo si el cociente de su tamaño (número de casos) entre el del intervalo del vecino es inferior al del umbral especificado; tenga en cuenta que los umbrales mayores pueden resultar en una fusión de más elementos.

### **Configuración de puntos de corte**

El cuadro de diálogo Configuración de punto de corte permite especificar opciones avanzadas para el algoritmo de intervalos óptimos. Estas opciones indican al algoritmo cómo calcular los intervalos utilizando el campo objetivo.

Figura 4-70  
Configuración de puntos de corte para intervalos óptimos



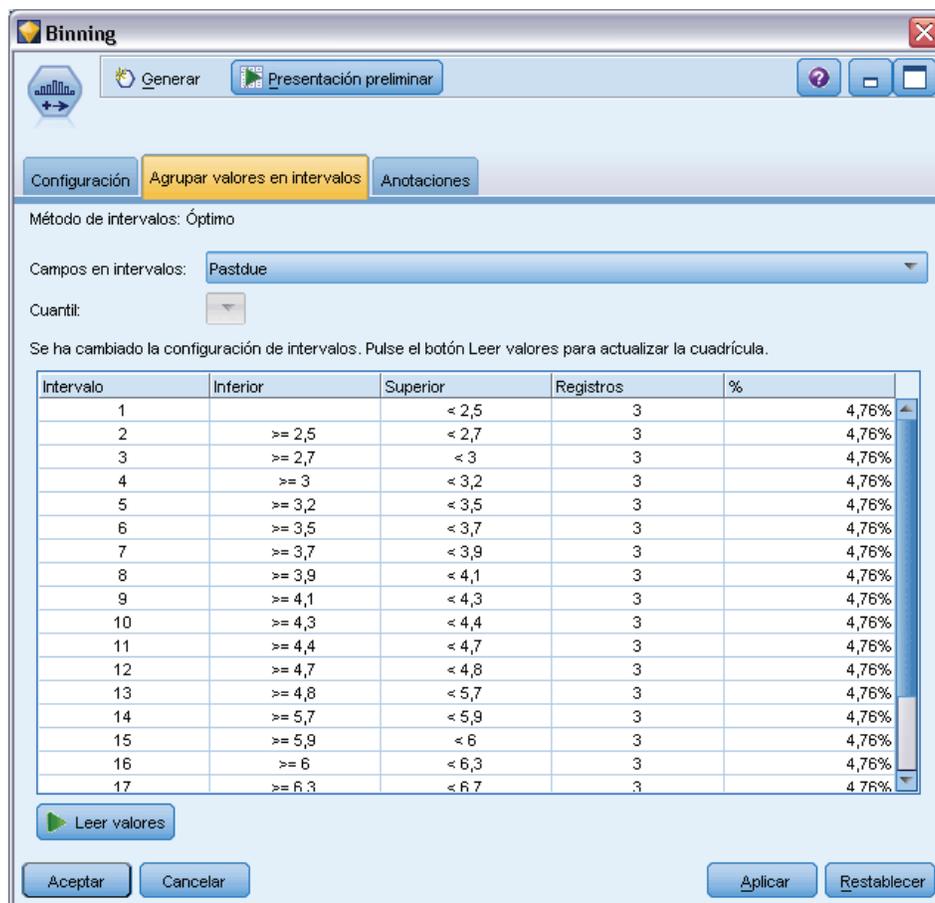
**Límites de intervalos.** Puede especificar si deben incluirse ( $\text{inferior} \leq x$ ) o excluirse ( $\text{inferior} < x$ ) los límites inferior y superior.

**Primer y último intervalo.** Tanto para el primero como para el último intervalo, puede especificar si los intervalos deben tener límites (puntos de datos superior e inferior) o no tenerlos (tendiendo hacia el infinito positivo o negativo).

### **Vista preliminar de intervalos generados**

La pestaña Agrupar valores en intervalos del nodo Intervalos permite ver los umbrales de los intervalos generados. Mediante el menú Generar, también puede generar un nodo Derivar que puede utilizar para aplicar estos umbrales de un conjunto de datos a otro.

Figura 4-71  
Cuadro de diálogo del nodo Intervalos, pestaña Agrupar valores en intervalos



**Campos en intervalos.** Utilice la lista desplegable para seleccionar un campo que visualizar. Los nombres de campos mostrados utilizan el nombre de campo original para mayor claridad.

**Cuantil.** Utilice la lista desplegable para seleccionar un cuantil que visualizar, como 10 o 100. Esta opción está disponible solamente cuando los intervalos se generan al utilizar el método de mosaico (igual frecuencia o igual suma).

**Umbral de intervalo.** Aquí se muestran los valores de umbral para cada intervalo generado, junto con el número de registros correspondiente a cada intervalo. Solamente en el método de intervalos óptimos, se muestra el número de registros de cada intervalo como porcentaje del total. Recuerde que los umbrales no tienen aplicación cuando se usa el método de intervalos de rangos.

**Leer valores.** Lee los valores agrupados en intervalos del conjunto de datos. Recuerde que los umbrales también se sobrescribirán cuando los datos nuevos se ejecuten en la ruta.

### **Generación de un nodo Derivar**

Puede utilizar el menú Generar para crear un nodo Derivar basado en los umbrales actuales. Esto resulta útil para aplicar los umbrales de intervalo establecidos en un conjunto de datos a otro. Además, una vez se conocen estos puntos de división, la operación Derivar es más eficiente (más rápida) que una operación Intervalos al trabajar con conjuntos de datos grandes.

## **Nodo Análisis de RFM**

El nodo Análisis de RFM (actualidad, frecuencia, monetario) permite determinar cuantitativamente qué clientes son los mejores examinando cuándo ha sido la compra más reciente de un cliente (actualidad), cuántas veces suelen comprar (frecuencia) y cuánto gasta el cliente en todas las transacciones (valor monetario).

La razón del análisis de RFM es que los clientes que han adquirido el producto o servicio una vez tienen más posibilidades de volver a comprarlo. Los datos del cliente categorizado se separan en un número de intervalos, con los criterios ajustados a sus necesidades. En cada uno de los intervalos se asigna una puntuación a los clientes y las puntuaciones se combinan para proporcionar una puntuación RFM global. Esta puntuación es una representación de la pertenencia del cliente a los intervalos creados por cada uno de los parámetros RFM. Los datos de los intervalos pueden ser suficientes para satisfacer sus necesidades, por ejemplo, identificando a los clientes más frecuentes o de mayor valor; además, se pueden pasar a una ruta para su posterior modelado y análisis.

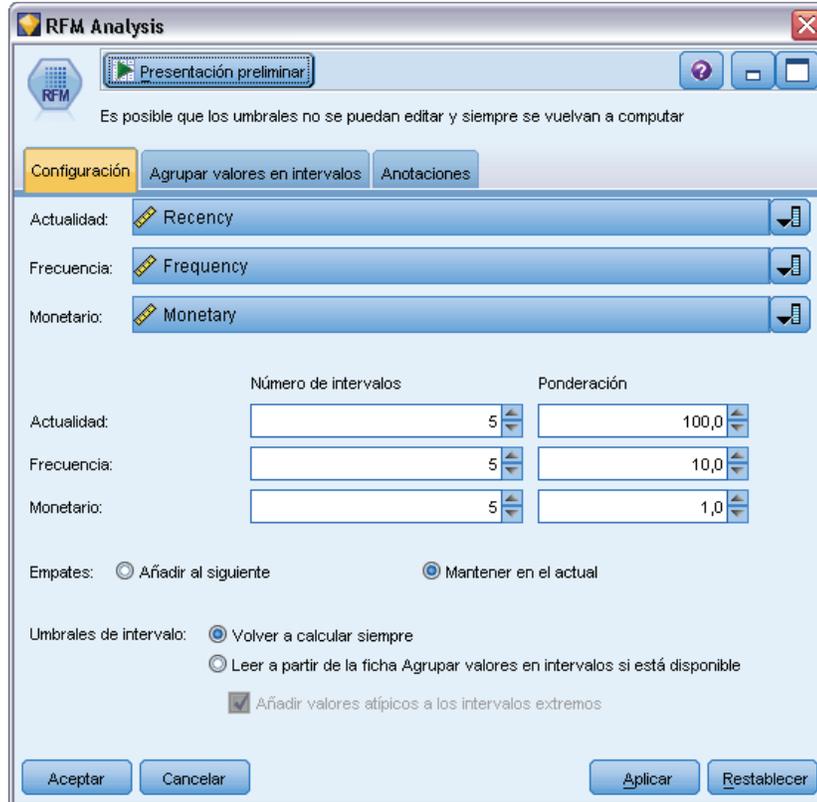
Tenga en cuenta, sin embargo, que aunque la posibilidad de analizar y clasificar puntuaciones RFM es una herramienta de gran utilidad, debe tener en cuenta diferentes factores a la hora de utilizarla. Puede tener la tentación de dirigirse a los clientes con las puntuaciones más altas; sin embargo, si se excede en esta actitud, puede provocar que el efecto contrario que opten por no volver a adquirir el producto. También cabe recordar que los clientes con menos puntuación no se deben despreciar, sino intentar que suban de categoría. Por el contrario, las puntuaciones más altas no son necesariamente un indicativo de mayores posibilidades de venta, dependiendo del mercado. Por ejemplo, un cliente en el intervalo 5 de actualidad (que ha adquirido un producto muy recientemente), no es el mejor objetivo para una empresa que venda productos de un alto precio o muy duraderos como coches o televisores.

*Nota:* Dependiendo de cómo almacene sus datos, es posible que necesite ejecutar el nodo Análisis de RFM, junto con el nodo Adición de RFM para transformar los datos en un formato utilizable. Por ejemplo, los datos de entrada deben estar en el formato del cliente, con una fila por cliente; si los datos de los clientes están en formato de transacciones, utilice un nodo Adición de RFM anterior para derivar los campos de actualidad, frecuencia y monetario. Si desea obtener más información, consulte el tema [Nodo Adición de RFM](#) en el capítulo 3 el p. 85.

Los nodos Adición de RFM y Análisis de RFM de IBM® SPSS® Modeler están configurados para utilizar intervalos independientes; es decir, ordenan por rango y establecen en intervalos los datos de cada medida de actividades recientes, frecuencia y valor monetario, sin considerar sus valores o las otras dos medidas.

## Configuración del nodo Análisis de RFM

Figura 4-72  
Configuración de las opciones de Análisis de RFM



**Actualidad.** Con el selector de campos (botón situado a la derecha del campo de texto), seleccione el campo de actualidad. Puede ser una fecha, marca de tiempo o un número simple. Tenga en cuenta que si una fecha o marca de tiempo representa la fecha de la transacción más reciente, el mayor valor se considera el más reciente; si se especifica un número, representa el tiempo transcurrido desde la transacción más reciente y el menor valor se considera el más reciente.

*Nota:* Si el nodo Análisis de RFM está precedido en la ruta por un nodo Adición de RFM, los campos de actualidad, frecuencia y monetario generados por el nodo Adición de RFM de deben seleccionar como entradas en el nodo Análisis de RFM.

**Frecuencia.** Si utiliza el selector de campos, seleccione el campo de frecuencia que se utilizará.

**Monetario.** Si utiliza el selector de campos, seleccione el campo de monetario que se utilizará.

**Número de intervalos.** En cada uno de los tres tipos de entrada, seleccione el número de intervalos que se crearán. El valor predeterminado es 5.

*Nota:* El número mínimo de intervalos es 2 y el máximo 9.

**Ponderación.** Por defecto, se da mayor importancia a los datos recientes, seguidos de frecuencia y monetario, cuando calcula puntuaciones. Si fuera necesario, puede modificar la ponderación que afecta a uno o varios de los tipos de datos para modificar al que se le concede mayor importancia.

La puntuación RFM se calcula de la siguiente forma: (puntuación de actualidad x ponderación de actualidad) + (puntuación de frecuencia x ponderación de frecuencia) + (puntuación de monetario x ponderación de monetario).

**Empates.** Especifique cómo se distribuirán por intervalos las puntuaciones idénticas (empatadas). Las opciones son:

- **Añadir al siguiente.** Seleccione esta opción para mover los valores de empate hacia el siguiente intervalo superior.
- **Mantener en el actual.** Mantiene los valores de empate en el intervalo actual (más bajo). Este método puede dar como resultado un menor número total de intervalos creados. (Éste es el valor por defecto.)

**Umbrales de intervalo.** Especifique si las puntuaciones RFM y las asignaciones de umbrales se vuelven a computar siempre cuando se ejecuta el nodo, o que solamente se computan cuando sea necesario (por ejemplo, si se han añadido nuevos datos). Si selecciona Leer de la pestaña Agrupar valores en intervalos si está disponible, puede editar los puntos de corte superior e inferior de los diferentes intervalos en la pestaña Agrupar valores en intervalos.

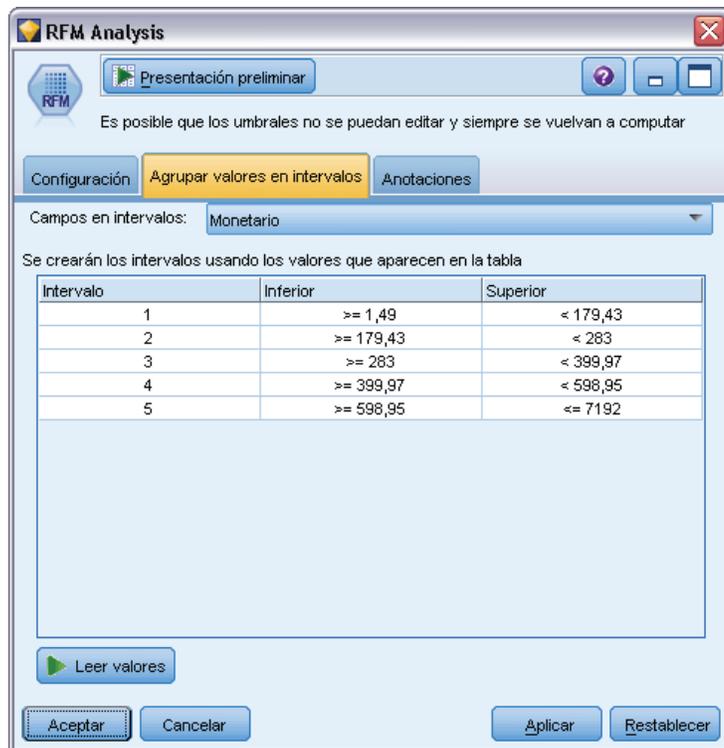
Cuando se ejecuta, el nodo Análisis de RFM distribuye por intervalos los campos actualidad, frecuencia y monetarios brutos y añade los siguientes nuevos campos al conjunto de datos:

- Puntuación de actualidad. Un rango (valor de intervalo) para actualidad
- Puntuación de frecuencia. Un rango (valor de intervalo) para frecuencia
- Valor monetario. Un rango (valor de intervalo) para monetario
- Puntuación de RFM. La suma ponderada de las puntuaciones de actualidad, frecuencia y monetario.

**Añadir valores atípicos a intervalos finales.** Si selecciona esta casilla de verificación, los registros por debajo del intervalo más inferior se añaden al intervalo inferior y los registros por encima, se añaden al intervalo superior, de otro modo, tendrán un valor nulo. Esta casilla solamente está disponible si selecciona Leer de la pestaña Agrupar valores en intervalos si está disponible.

## Intervalos del nodo Análisis de RFM

Figura 4-73  
Configuración de los valores de intervalo de Análisis de RFM



La pestaña Agrupar valores permite visualizar y, en algunos casos, modificar los umbrales de los intervalos generados.

*Nota:* Solamente puede modificar los valores de esta pestaña si selecciona Leer de la pestaña Agrupar valores en intervalos si está disponible en la pestaña Configuración.

**Campos en intervalos.** Utilice la lista desplegable para seleccionar un campo para dividir en valores. Los valores disponibles son los que seleccione de la pestaña Configuración.

**Tabla de valores en intervalos.** Aquí se muestran los valores de umbrales para cada intervalo generado. Si selecciona Leer de la pestaña Agrupar valores en intervalos si está disponible en la pestaña Configuración, puede modificar los puntos de corte superior e inferior de cada intervalo pulsando dos veces en la casilla correspondiente.

**Leer valores.** Lee valores de intervalos del conjunto de datos y completa los valores de intervalo de la tabla. Tenga en cuenta que si selecciona Recalcular siempre en la pestaña Configuración, los umbrales de los intervalos se sobrescribirán cuando los nuevos datos se ejecuten en la ruta.

## **Nodo Partición**

Los nodos de partición se utilizan para generar un campo de partición que divide los datos en subconjuntos o muestras independientes para las fases de entrenamiento, comprobación y validación en la generación del modelo. Si usa una muestra para generar el modelo y otra muestra distinta para probarlo, podrá obtener una buena indicación de la bondad del modelo a la hora de generalizarlo a conjuntos de datos de mayor tamaño similares a los datos actuales.

El nodo Partición genera un campo nominal con el papel definido a Partición. Si lo prefiere, en el caso de que un campo adecuado ya exista en los datos, puede designarse como una partición utilizando un nodo Tipo. En este caso no se requiere ningún nodo Partición independiente. Se puede utilizar como partición cualquier campo nominal instanciado con dos o tres valores, pero no los campos con marcas. Si desea obtener más información, consulte el tema [Definición del papel de campos](#) el p. 149.

En una ruta se pueden definir múltiples campos de partición pero, de hacerlo, será necesario seleccionar un campo de partición simple en la pestaña Campos de cada nodo de modelado que utilice la partición. (Si solamente hay una partición, se usará automáticamente siempre que se active la partición.)

**Activación de la partición.** Para utilizar la partición en un análisis, ésta debe estar activada en la pestaña Opciones de modelo en el nodo Análisis o la generación de modelos adecuada. Si se anula esta opción, se posibilita la desactivación de la partición sin eliminar el campo.

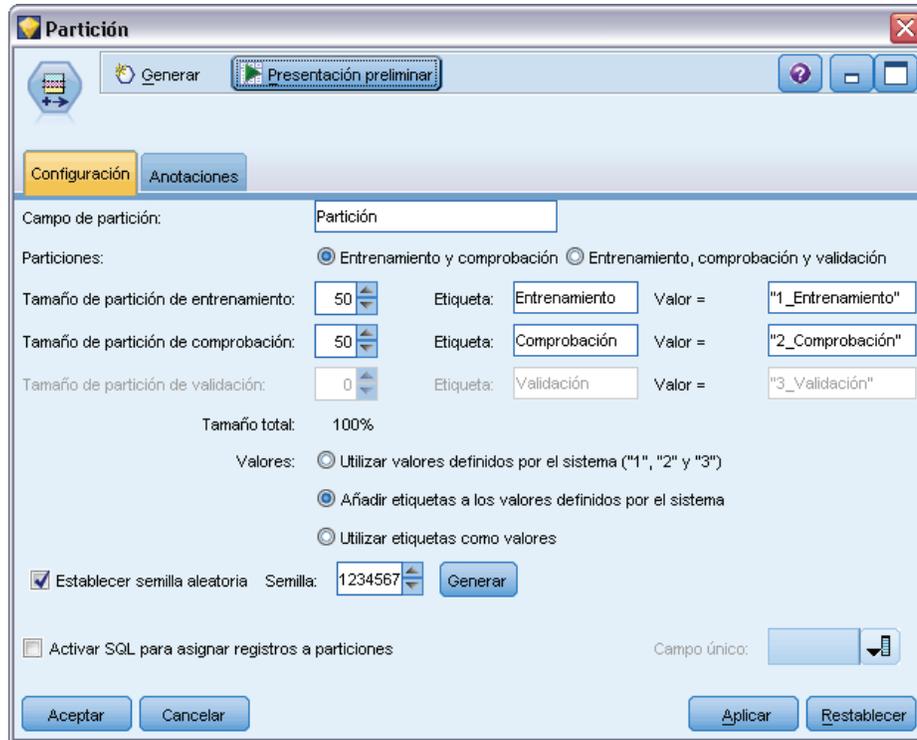
Para crear un campo de partición en función de algún otro criterio como un rango de fechas o una ubicación, también puede utilizar un nodo Derivar. Si desea obtener más información, consulte el tema [Nodo Derivar](#) el p. 165.

**Ejemplo.** Cuando genere una ruta de RFM para identificar clientes recientes que han respondido de forma positiva a campañas de marketing anteriores, el departamento de marketing de una empresa de ventas utiliza un nodo Partición para dividir los datos en particiones de prueba y entrenamiento.

## Opciones del nodo Partición

Figura 4-74

Cuadro de diálogo del nodo Partición, pestaña Configuración



**Campo de partición.** Especifica el nombre del campo creado por el nodo.

**Particiones.** Puede dividir los datos en dos muestras (Entrenamiento y comprobación) o en tres (Entrenamiento, comprobación y validación).

- **Entrenamiento y comprobación.** Divide los datos en dos particiones de muestra, lo que permite entrenar el modelo con una muestra y comprobarlo con otra.
- **Entrenamiento, comprobación y validación.** Divide los datos en tres particiones de muestra, lo que permite entrenar el modelo con una muestra, comprobarlo y ajustarlo con otra y validar los resultados con otra. Así se reduce el tamaño de cada partición en consecuencia y puede ser más adecuado al trabajar con un conjunto de datos muy grande.

**Tamaño de partición.** Especifica el tamaño relativo de cada partición. Si la suma de los tamaños de partición es inferior al 100%, los registros no incluidos en una partición se descartarán. Por ejemplo, si un usuario tiene 10 millones de registros y tamaños de partición del 5% para el entrenamiento y el 10% para la prueba, después de ejecutar el nodo, debería haber unos 500.000 registros de entrenamiento y un millón de registros de prueba, con el resto descartado.

**Valores.** Especifica los valores utilizados para representar cada muestra de partición de los datos.

- **Utiliza valores definidos por el sistema (1, 2 y 3).** Utiliza un número entero para representar cada partición; por ejemplo, todos los registros que se encuentran en la muestra de entrenamiento tienen un valor de 1 para el campo de partición. Así se garantiza que los datos sean intercambiables entre configuraciones regionales y que si el campo de partición se

reinstancia en otro lugar (por ejemplo, al leer los datos de una base de datos), se mantendrá el orden de clasificación (así, 1 seguirá representando la partición de entrenamiento). Sin embargo, los valores requieren alguna interpretación.

- **Añadir etiquetas a los valores definidos por el sistema.** Combina el número entero con una etiqueta; por ejemplo, los registros de partición de entrenamiento tienen un valor de *1\_Entrenamiento*. Así es posible identificar la identidad de los valores y mantener el orden de clasificación. Sin embargo, los valores son específicos de una configuración regional específica.
- **Utilizar etiquetas como valores.** Utiliza la etiqueta sin números enteros; por ejemplo, *Entrenamiento*. Esto permite especificar los valores al editar las etiquetas. Sin embargo, los datos de configuración se vuelven específicos de la configuración regional y, al volver a crear una instancia de una columna de partición, los valores se colocan en su orden natural, que puede no corresponder con su orden “semántico”.

**Establecer semilla aleatoria.** Cuando se muestran o particionan registros a partir de un porcentaje aleatorio, esta opción permite duplicar los mismos resultados en otra sesión. Al especificar el valor inicial utilizado por el generador de números aleatorios, puede garantizar que se asignan los mismos registros cada vez que se ejecuta el nodo. Introduzca el valor de semilla deseado, o pulse en el botón Generar para generar automáticamente un valor aleatorio. Si no se selecciona esta opción, se generará una muestra diferente cada vez que se ejecute el nodo.

*Nota:* cuando se utiliza la opción Establecer semilla aleatoria con registros leídos de una base de datos, puede ser necesario un nodo Ordenar, antes del muestreo con el fin de garantizar el mismo resultado cada vez que se ejecute el nodo. Esto se debe a que la semilla aleatoria depende del orden de registros, sin estar garantizado que sea el mismo en una base de datos relacional. Si desea obtener más información, consulte el tema [Nodo Ordenar](#) en el capítulo 3 el p. 87.

**Activar SQL para asignar registros a particiones.** (Solamente para bases de datos de nivel 1.) Marque este cuadro para utilizar retroacción SQL para asignar registros a particiones. En la lista desplegable Campo único, seleccione un campo con valores exclusivos (como un campo ID) para garantizar que los registros se asignan de forma aleatoria pero repetible.

Los niveles de la base de datos se explican en la descripción del nodo de fuente Base de datos. Si desea obtener más información, consulte el tema [Nodo de fuente Base de datos](#) en el capítulo 2 el p. 15.

### **Generación de nodos Seleccionar**

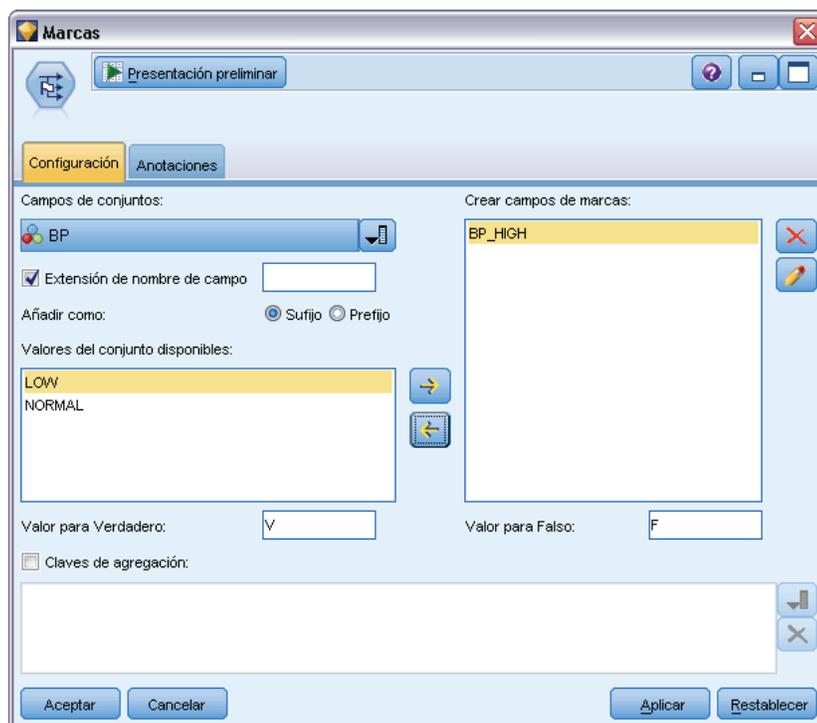
Mediante el menú Generar del nodo Partición se puede generar automáticamente un nodo Seleccionar para cada partición. Por ejemplo, se pueden seleccionar todos los registros de la partición de prueba para obtener una evaluación o análisis más profundos al utilizar solamente esta partición.

## **Nodo Marcas**

El nodo Marcas se utiliza para derivar campos de marcas en función de los valores categóricos definidos para uno o más campos nominales. Por ejemplo, su conjunto de datos puede contener un campo nominal, *PS* (presión sanguínea), con los valores *Alta*, *Normal* y *Baja*. Para manejar los

datos de una forma más sencilla, puede crear un campo de marcas para la presión sanguínea alta, que indicará si el paciente tiene la presión sanguínea alta o no.

Figura 4-75  
Creación de un campo marca para presión sanguínea alta



## Opciones de configuración del nodo Marcas

**Campos de conjuntos.** Enumera todos los campos de datos con un nivel de medición de *Nominal* (conjunto). Seleccione uno de la lista para mostrar los valores en el conjunto. Puede elegir entre estos valores para crear un campo de marcas. Tenga en cuenta que, para poder ver los campos nominales disponibles (y sus valores), los datos se deben instanciar completamente utilizando un nodo Tipo o de fuente en un punto anterior de la ruta. Si desea obtener más información, consulte el tema [Nodo Tipo](#) el p. 135.

**Extensión de nombre de campo.** Seleccione esta opción para activar los controles y especificar una extensión que añadir como sufijo o prefijo al nuevo campo de marcas. Por defecto, los nombres de los campos nuevos se crean automáticamente combinando el nombre del campo original con el valor del campo en una etiqueta, como *Nombredcampo\_valordecampo*.

**Valores del conjunto disponibles.** Aquí se muestran los valores del conjunto seleccionado anteriormente. Seleccione uno o varios valores para los que desea generar marcas. Por ejemplo, si los valores de un campo denominado *presión\_sanguínea* son *Alta*, *Media* y *Baja*, puede seleccionar *Alta* y añadirla a la lista de la derecha. Así se creará un campo con una marca para los registros que tendrá un valor indicativo de alta presión sanguínea.

**Crear campos de marcas.** Aquí se enumeran los campos de marcas creados recientemente. Puede especificar opciones para nombrar el nuevo campo mediante los controles de extensión de nombre de campo.

**Valor para Verdadero.** Permite especificar el valor para verdadero utilizado por el nodo al configurar una marca. Este valor es V por defecto.

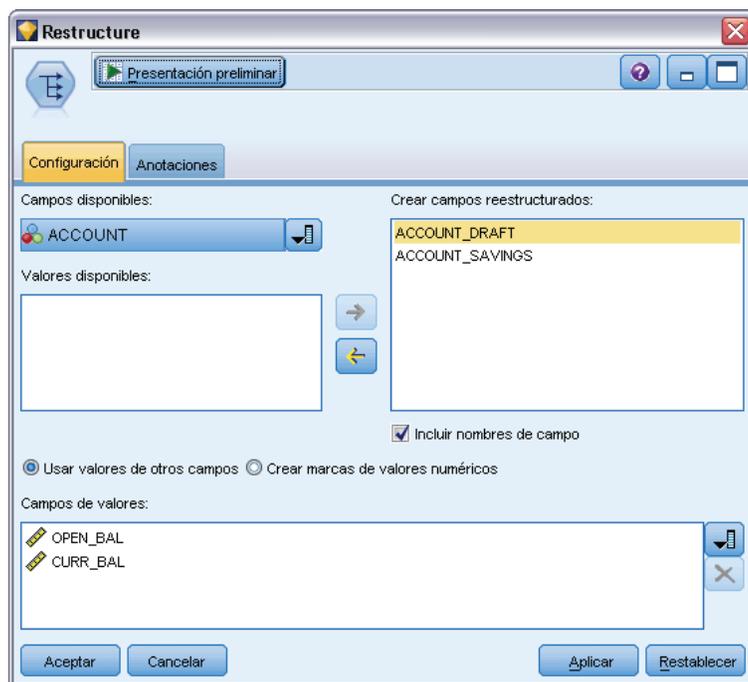
**Valor para Falso.** Permite especificar el valor para falso utilizado por el nodo al configurar una marca. Este valor es F por defecto.

**Claves de agregación.** Seleccione esta opción para agrupar registros en función de los campos clave especificados a continuación. Cuando se selecciona Claves de agregación, todos los campos de marcas de un grupo se “activan” si *cualquier* registro se ha configurado como verdadero. Utilice el selector de campos para especificar los campos clave que se utilizarán para agregar registros.

## Nodo Reestructurar

El nodo Reestructurar puede usarse para generar varios campos a partir de los valores de un campo nominal o de marcas. Estos campos recién generados pueden contener valores de otro campo o marcas numéricas (0 y 1). La funcionalidad de este nodo es similar a la que posee el nodo Marcas. Sin embargo, ofrece mayor flexibilidad. Con él, podrá crear campos de cualquier tipo (incluidas marcas numéricas) a partir de los valores de otro campo. Tras ello, puede llevar a cabo una agregación o cualquier otra manipulación con otros nodos a partir de ese punto de la ruta. (El nodo Marcas permite agregar campos en un paso, lo cual es conveniente si está creando campos de marcas.)

Figura 4-76  
Generación de campos reestructurados para Cuenta



Por ejemplo, el siguiente conjunto de datos contiene un campo nominal, *Cuenta*, con los valores *Ahorros* y *Cheque*. Se registra el saldo de apertura y el actual de cada cuenta, y son varios los clientes que tienen varias cuentas de cada tipo. Imaginemos que desea saber si cada cliente tiene un tipo de cuenta concreto y, de ser así, la cantidad de dinero que contiene cada tipo de cuenta. Usará el nodo Reestructurar para generar un campo para cada uno de los valores de *Cuenta* y seleccionará *Saldo\_actual* como el valor. Cada campo nuevo se rellenará con el saldo actual del registro correspondiente.

Tabla 4-2  
Datos de muestra antes de la reestructuración

IDclien	Cuenta	Saldo_ab	cheque_cuent
12701	Borrador	1000	1005.32
12702	Ahorros	100	144.51
12703	Ahorros	300	321.20
12703	Ahorros	150	204.51
12703	Borrador	1200	586.32

Tabla 4-3  
Datos de muestra tras la reestructuración

IDclien	Cuenta	Saldo_ab	cheque_cuent	Saldo_actual_ cheque_cuent	Saldo_actual_ ahorros_cuent
12701	Borrador	1000	1005.32	1005.32	\$null\$
12702	Ahorros	100	144.51	\$null\$	144.51
12703	Ahorros	300	321.20	\$null\$	321.20
12703	Ahorros	150	204.51	\$null\$	204.51
12703	Borrador	1200	586.32	586.32	\$null\$

### Uso del nodo Reestructurar con el nodo Agregar

En muchas ocasiones, puede que quiera emparejar el nodo Reestructurar con un nodo Agregar. En el ejemplo anterior, un cliente (con el ID 12703) tiene tres cuentas. Puede usar un nodo Agregar para calcular el saldo total de cada tipo de cuenta. El campo clave es *IDclien* y los campos agregados son los nuevos campos reestructurados, *Saldo\_actual\_cheque\_cuent* y *Saldo\_actual\_ahorros\_cuent*. La siguiente tabla recoge los resultados.

Tabla 4-4  
Datos de muestra tras la reestructuración y agregación

IDclien	Recuento_registro	Suma_saldo_actual_ ahorros_cuent	Suma_saldo_actual_ ahorros_cuent
12701	1	1005.32	\$null\$
12702	1	\$null\$	144.51
12703	3	586.32	525.71

### Opciones de configuración del nodo Reestructurar

**Campos disponibles.** Enumera todos los campos de datos con un nivel de medición de *Nominal* (conjunto) o *Marca*. Seleccione uno de la lista para mostrar los valores en el conjunto o marca y, a continuación, elija de entre estos valores para crear los campos reestructurados. Tenga en

cuenta que, para poder ver los campos disponibles (y sus valores), los datos se deben instanciar completamente utilizando un nodo Tipo o un nodo de fuente anterior de la ruta. Si desea obtener más información, consulte el tema [Nodo Tipo](#) el p. 135.

**Valores disponibles.** Aquí se muestran los valores del conjunto seleccionado anteriormente. Seleccione uno o varios valores para los que desea generar campos reestructurados. Por ejemplo, si los valores en un campo denominado *Presión sanguínea* son *Alta*, *Media* y *Baja*, puede seleccionar *Alta* y añadir esta opción a la lista de la derecha. De esta forma, se creará un campo con un valor especificado (consulte a continuación) para los registros con el valor *Alta*.

**Crear campos reestructurados.** Aquí se enumeran los campos reestructurados recién creados. Por defecto, los nombres de los campos nuevos se crean automáticamente combinando el nombre del campo original con el valor del campo en una etiqueta, como *Nombredcampo\_valordecampo*.

**Incluir nombres de campo.** Anule la selección para eliminar el nombre original del campo como prefijo de los nombres de los nuevos campos.

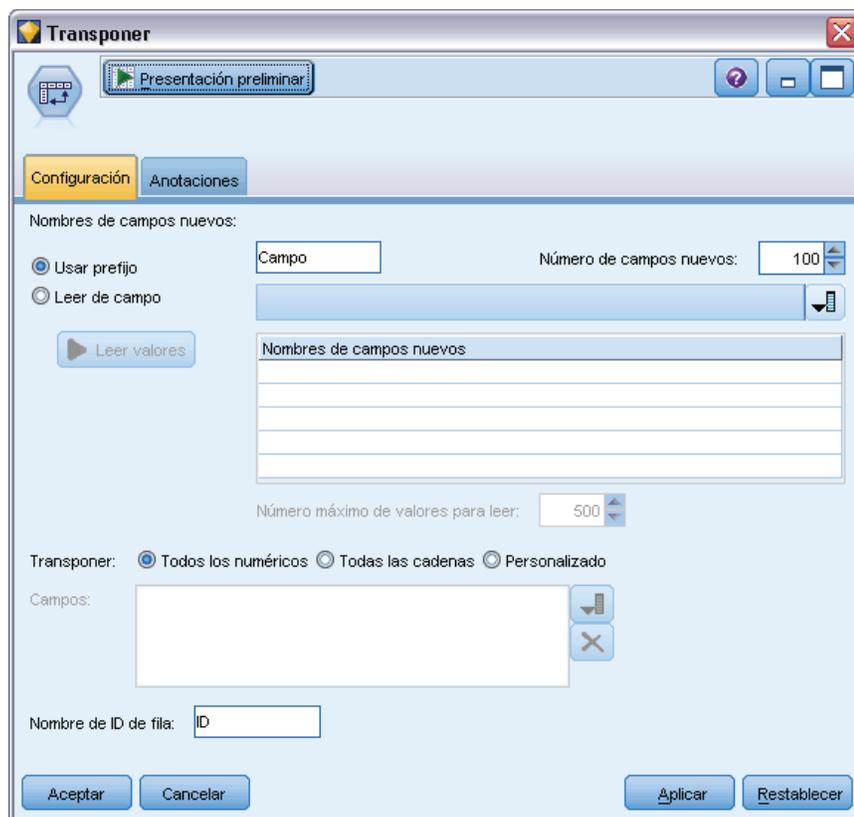
**Usar valores de otros campos.** Especifique uno o varios campos cuyos valores se vayan a usar para rellenar los campos reestructurados. Utilice el selector de campos para seleccionar uno o varios campos. Por cada campo seleccionado se creará un nuevo campo. El nombre del campo de valor se añadirá al nombre del campo reestructurado; por ejemplo, *PS\_alta\_edad* o *PS\_baja\_edad*. Todos los campos nuevos heredan el tipo del campo de valor original.

**Crear marcas de valores numéricos.** Seleccione para rellenar los nuevos campos con marcas de valores numéricos (0 para falso y 1 para verdadero), en lugar de usar un valor de otro campo.

## ***Nodo Transponer***

Por defecto, las columnas son campos y las filas, registros u observaciones. En caso de que sea necesario, puede usarse un nodo Transponer para intercambiar los datos de las filas y columnas de forma que los campos se convierten en registros y los registros, en campos. Por ejemplo, si tiene datos de series temporales donde cada serie es una fila y no una columna, puede transponer los datos antes de proceder al análisis.

Figura 4-77  
nodo Transponer, pestaña Configuración



## Opciones de configuración del nodo Transponer

### Nuevos nombres de campos

Se pueden generar nuevos nombres de campo automáticamente a partir de un prefijo concreto, o bien se pueden leer desde un campo existente en los datos.

**Usar prefijo.** Con esta opción se generan nuevos nombres de campos a partir del prefijo especificado (*Campo1*, *Campo2*, etc.). Puede personalizar el prefijo como considere conveniente. Con esta opción debe indicar el número de campos que se va a crear, independientemente del número de filas en los datos originales. Por ejemplo, si Número de campos nuevos está definido como 100, se descartarán todos los datos que no estén contenidos en las 100 primeras filas. Si hay menos de 100 filas en los datos originales, algunos de los campos serán nulos. (Puede aumentar el número de campos según sea necesario, pero el propósito de esta configuración reside en evitar que millones de registros se transpongan a millones de campos, dado que con esto se obtendría un resultado imposible de administrar.)

Por ejemplo, suponga que tiene datos con series en filas y un campo diferente (columna) para cada mes. Puede transponerlos para que cada serie esté en un campo diferente, con una fila para cada mes.

Figura 4-78  
 Datos originales con series en filas

	Jan	Feb	Mar	Apr
1	1	3	5	7
2	2	4	6	8

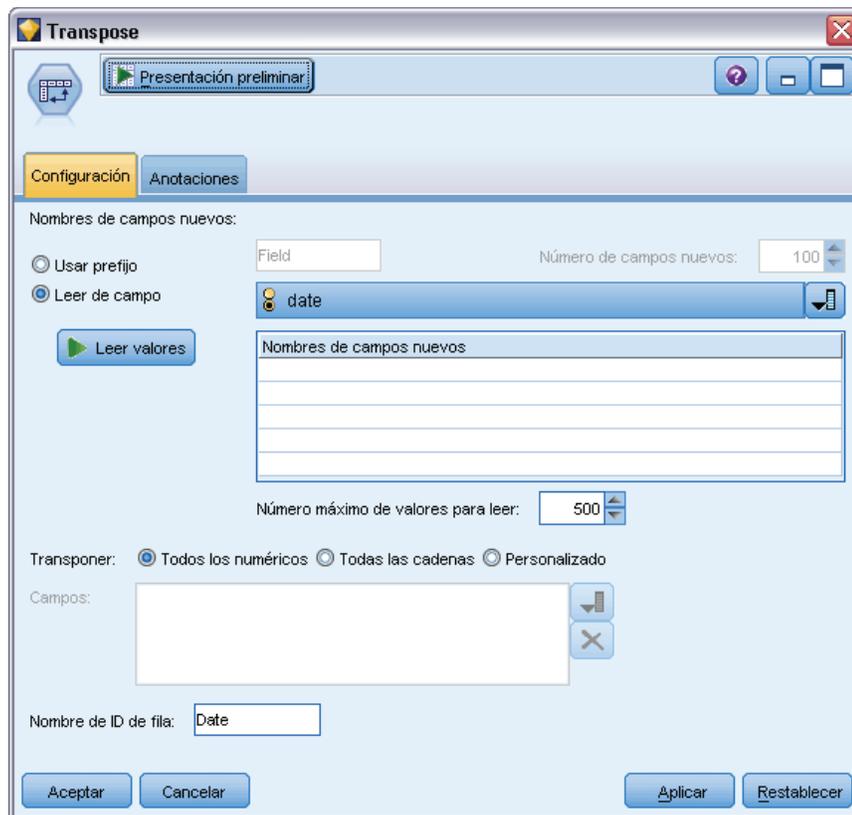
Figura 4-79  
 Datos transpuestos con series en columnas

	Month	Field1	Field2
1	Jan	1	2
2	Feb	3	4
3	Mar	5	6
4	Apr	7	8

*Nota:* para producir los resultados que se muestran, la opción Número de campos nuevos se ha cambiado de 100 a 2 y el Nombre de ID de fila de ID a Mes (vea a continuación).

**Leer de campo.** Lee nombres de campo de un campo existente. Con esta opción, el número de campos nuevos se establece a partir de los datos y hasta el máximo establecido. Cada valor del campo seleccionado se convierte en un nuevo campo en los datos de salida. El campo seleccionado admite cualquier tipo de almacenamiento (número entero, cadena, fecha, etc.), pero, para evitar nombres de campos duplicados, cada valor del campo seleccionado debe ser único (en otras palabras, el número de valores ha de coincidir con el número de filas). En caso de que existan nombres de campos duplicados, se mostrará una advertencia.

Figura 4-80  
Lectura de nombres de campo de un campo existente



- **Leer valores.** Si el campo seleccionado no se ha instanciado, seleccione esta opción para rellenar la lista de nuevos nombres de campos. En caso de que el campo sí esté instanciado, este paso no será necesario.
- **Número máximo de valores para leer.** Al leer nombres de campos desde los datos, se especifica un límite superior con el fin de evitar que se genere un número desmesurado de campos. (Tal y como se ha indicado anteriormente, la transposición de un millón de registros en un millón de campos produciría un resultado imposible de manejar.)

Por ejemplo, si la primera columna de los datos especifica el nombre de cada serie, puede utilizar estos valores como nombres de campo en los datos transpuestos.

Figura 4-81  
 Datos originales con series en una fila

	date	1949-01-01	1949-02-01	1949-04-01	1949-05-01	1949-06-01	1949-07-01	1949-08-01
1	passengers	112.000	118.000	129.000	121.000	135.000	148.000	148.000

Figura 4-82  
 Datos transpuestos con series en columnas

	Date	passengers
1	1949-01-01	112.000
2	1949-02-01	118.000
3	1949-04-01	129.000
4	1949-05-01	121.000
5	1949-06-01	135.000
6	1949-07-01	148.000
7	1949-08-01	148.000
8	1949-09-01	136.000
9	1949-10-01	119.000
10	1949-11-01	104.000
11	1949-12-01	118.000
12	1950-01-01	115.000
13	1950-02-01	126.000
14	1950-03-01	141.000
15	1950-04-01	135.000
16	1950-05-01	125.000
17	1950-06-01	149.000
18	1950-07-01	170.000
19	1950-08-01	170.000

**Transponer.** Por defecto, únicamente los campos continuos (rango numérico) se transponen (ya sean de número entero o de almacenamiento real). De forma opcional, puede elegir un subconjunto de campos numéricos o transponer campos de cadena en su lugar. No obstante, todos los campos transpuestos deben ser del mismo tipo de almacenamiento (numérico o de cadena, pero nunca ambos), dado que la mezcla de campos de entrada generaría valores mezclados en cada columna de resultados y, en consecuencia, se incumpliría la norma que establece que todos los valores de un campo deben tener el mismo almacenamiento. El resto de tipos de almacenamiento (fecha, hora, marca de tiempo) no puede transponerse.

- **Todos los numéricos.** Transpone todos los campos numéricos (número entero o almacenamiento real). El número de filas del resultado coincide con el número de campos numéricos en los datos originales.
- **Todas las cadenas.** Transpone todos los campos de cadena.
- **Personalizado.** Permite seleccionar un subconjunto de campos numéricos. El número de filas del resultado coincide con el número de campos seleccionado. *Nota:* esta opción solamente se encuentra disponible para los campos numéricos.

**Nombre ID de fila.** Especifica el nombre del campo de ID de fila creado por el nodo. Los valores de este campo se establecen a partir de los nombres de los campos en los datos originales.

*Sugerencia:* Cuando se transponen datos de series temporales de filas a columnas, si los datos originales incluyen una fila (como fecha, mes o año) que etiqueta el período para cada medida, asegúrese de que estas etiquetas se leen en IBM® SPSS® Modeler como nombres de campo (tal y como se describe en los ejemplos anteriores, que muestran el mes o la fecha como nombres de campo en los datos originales, respectivamente) en lugar de incluir la etiqueta en la primera fila de datos. De esta forma evitará mezclar etiquetas y valores en cada columna (lo que forzaría que los números se leyesen como cadenas, ya que no se pueden mezclar tipos de almacenamiento dentro de una columna).

## ***Nodo de intervalos de tiempo***

El nodo Intervalos de tiempo permite especificar intervalos y generar etiquetas para los datos que se utilizarán en un modelado de series temporales o un nodo Gráfico de tiempo con fines de estimación o pronóstico. Se admite una gama completa de intervalos de tiempo que abarca desde segundos a años. Por ejemplo, si tiene una serie con medidas diarias que empieza el 3 de enero de 2005, podrá etiquetar registros a partir de esa fecha, siendo la segunda fila el 4 de enero, y así sucesivamente. También puede indicar la periodicidad; por ejemplo, cinco días a la semana u ocho horas al día.

Además, puede especificar el rango de registros que se va a utilizar para la estimación. Puede elegir entre excluir los registros más recientes de la serie o especificar casos reservados. De esta forma, puede probar el modelo reservando los registros más recientes de los datos de series temporales para comparar sus valores conocidos con los valores estimados para dichos períodos.

También puede especificar cuántos períodos de tiempo desea pronosticar en el futuro y los valores futuros que se van a utilizar en la predicción en nodos de modelado de serie temporal a partir de ese punto de la ruta.

El nodo Intervalos de tiempo genera un campo *EtiquetaTiempo* en el formato adecuado para el intervalo y el período, así como un campo *ÍndiceTiempo*, que asigna un número entero único a cada registro. Puede que además se generen varios campos adicionales según el intervalo o período seleccionado (como el minuto o el segundo al que corresponde una medida).

Puede rellenar o agregar valores según sea necesario con el fin de asegurarse de que las medidas están espaciadas uniformemente. Los métodos para modelar datos de series temporales precisan de un intervalo uniforme entre cada medida. Los valores perdidos deben indicarse mediante filas vacías. Si sus datos no cumplen con este requisito, el nodo puede transformarlos para que lo hagan.

### Comentarios

- Es posible que los intervalos periódicos no coincidan con el tiempo real. Así, por ejemplo, una serie basada en una semana de trabajo estándar compuesta por cinco días consideraría el intervalo entre viernes y lunes como un solo día.
- El nodo de intervalos de tiempo da por hecho que cada serie corresponde a un campo o una columna, con una fila por medida. Si procede, puede transponer los datos para cumplir este requisito. Si desea obtener más información, consulte el tema [Nodo Transponer](#) el p. 211.
- En cuanto a aquellas series que no están espaciadas uniformemente, puede especificar un campo que identifique la fecha o la hora de cada medida. Recuerde que para ello se necesita un campo de fecha, hora o marca de tiempo con el formato apropiado para utilizarlo como entrada. Si es preciso, puede convertir a este formato un campo existente (por ejemplo, un campo de cadena o de etiqueta) mediante un nodo Rellenar. Si desea obtener más información, consulte el tema [Conversión del almacenamiento mediante el nodo Rellenar](#) el p. 179.
- A la hora de ver los detalles de la etiqueta y los campos de índice generados, a veces resulta práctico activar la visualización de etiquetas de valores. Por ejemplo, al visualizar una tabla con valores generados para los datos mensuales, puede pulsar en el icono de etiquetas de valores de la barra de herramientas para ver *enero, febrero, marzo*, etc. en lugar de *1, 2, 3*, etc.

Figura 4-83

icono de etiquetas de valores



### Especificación de intervalos de tiempo

Con la pestaña Intervalos podrá indicar el intervalo y la periodicidad para crear o etiquetar las series. La configuración particular depende del intervalo seleccionado: Así, si selecciona Horas por día, podrá establecer el número de días a la semana, el día en el que cada semana empieza, el número de horas al día y la hora a la que cada día empieza. Si desea obtener más información, consulte el tema [Intervalos admitidos](#) el p. 225.

Figura 4-84  
Configuración del intervalo de tiempo de una serie por horas

### **Etiquetado o creación de la serie**

Puede etiquetar los registros de manera consecutiva o crear la serie en función de una fecha, marca de tiempo o campo de tiempo concretos.

- **Iniciar etiquetado desde el primer registro.** Especifique la fecha y/o hora de inicio para etiquetar registros consecutivos. Si, por ejemplo, va a etiquetar las horas por día, especificaría la fecha y la hora a la que empezaría la serie, con un solo registro por cada hora sucesiva. Aparte de añadir etiquetas, este método no altera los datos originales. En su lugar, supone que los registros están espaciados uniformemente, con un intervalo uniforme entre cada medida. Cualquier medida perdida deberá indicarse mediante filas vacías en los datos.
- **Generar a partir de los datos.** En cuanto a aquellas series que no están espaciadas por igual, puede especificar un campo que identifique la fecha o la hora de cada medida. Recuerde que para ello se necesita un campo de fecha, hora o marca de tiempo con el formato apropiado para utilizarlo como entrada. Por ejemplo, si tiene un campo de cadena con valores como *Ene 2000*, *Feb 2000*, etc., puede convertirlo en un campo de fecha mediante el nodo Rellenar. Si desea obtener más información, consulte el tema [Conversión del almacenamiento mediante el nodo Rellenar](#) el p. 179. La opción Generar a partir de los datos también transforma los datos para que coincidan con el intervalo especificado rellenando o agregando registros cuando proceda; así, acumulando semanas en meses o sustituyendo los registros perdidos por espacios en blanco o valores extrapolados. Puede especificar las funciones utilizadas para

rellenar o agregar registros en la pestaña Generar. Si desea obtener más información, consulte el tema [Opciones de generación en intervalos de tiempo](#) el p. 219.

**Nueva extensión de nombre de campo.** Permite especificar un prefijo o un sufijo que se aplica a todos los campos que el nodo genera. Por ejemplo, si se usa el prefijo por defecto *\$TI*, los campos que el nodo cree se denominarán *\$TI\_ÍndiceTiempo*, *\$TI\_EtiquetaTiempo*, etc.

**Formato de fecha.** Especifica el formato para el campo *EtiquetaTiempo* que el nodo ha creado según proceda con respecto al intervalo actual. La disponibilidad de estas opciones depende de la selección actual.

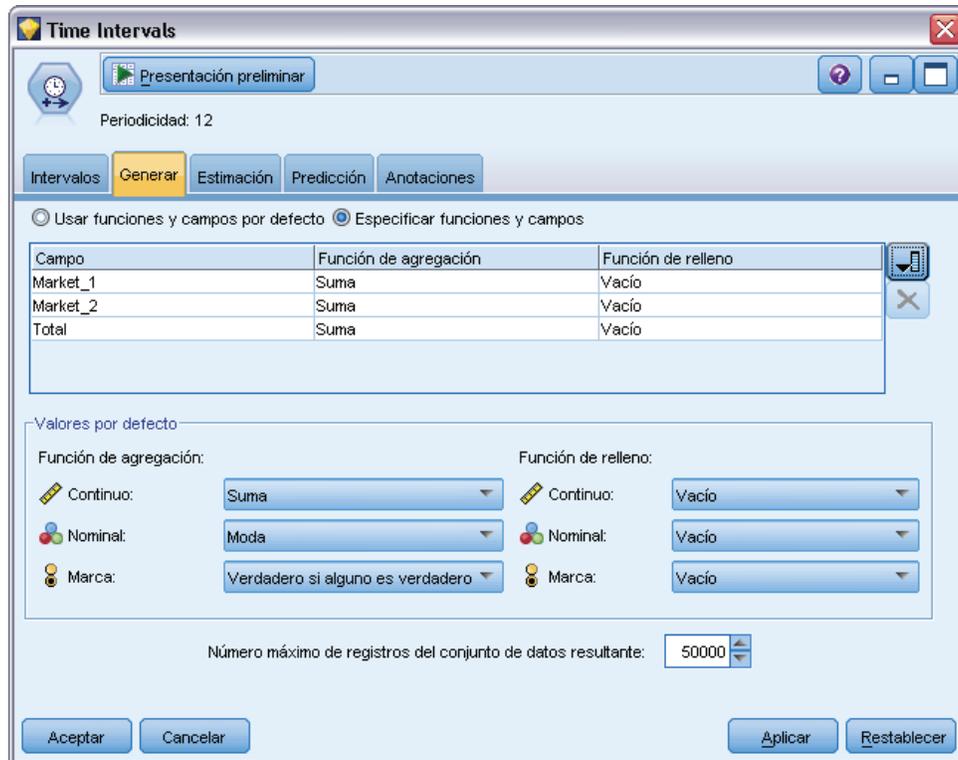
**Formato de hora.** Especifica el formato para el campo *EtiquetaTiempo* que el nodo ha creado según proceda con respecto al intervalo actual. La disponibilidad de estas opciones depende de la selección actual.

### ***Opciones de generación en intervalos de tiempo***

Con la pestaña Generar del nodo Intervalos de tiempo podrá especificar opciones para agregar o rellenar campos de manera que coincidan con el intervalo concreto. Esta configuración se aplica únicamente si se selecciona la opción Generar a partir de los datos de la pestaña Intervalos. Por ejemplo, si tiene una mezcla de datos semanales y mensuales, podría agregar o “acumular” los valores semanales con el fin de obtener un intervalo mensual uniforme. Otra opción consiste en establecer el intervalo en semanal y rellenar la serie insertando valores en blanco para las semanas que faltan, o bien extrapolando los valores que faltan mediante una función de relleno específica.

Cuando se rellenan o agregan datos, los campos de marca de tiempo o de fecha que existan se sustituyen eficazmente con los campos *EtiquetaTiempo* e *ÍndiceTiempo* generados y se eliminan del resultado. Los campos sin tipos se eliminan igualmente. Se conservan aquellos campos que miden el tiempo como una duración (como un campo que mide la duración de una llamada de servicio y no la hora a la que la llamada se inició), siempre y cuando se almacenen internamente como campos de tiempo y no de marca de tiempo. Si desea obtener más información, consulte el tema [Configuración de almacenamiento de campos y formato](#) en el capítulo 2 el p. 32. Otros campos se agregan en función de las opciones especificadas en la pestaña Generar.

Figura 4-85  
nodo Intervalos de tiempo, pestaña Generar



- **Usar funciones y campos por defecto.** Especifica que todos los campos han de agregarse o rellenarse según sea necesario, a excepción de los campos sin tipo, los de marca de tiempo y los de fecha, tal y como se ha indicado anteriormente. La función por defecto que se aplica depende del nivel de medición. Por ejemplo, los campos continuos se agregarán utilizando la media, mientras que los campos nominales se agregarán utilizando la moda. El valor por defecto para uno o varios niveles de medición puede cambiarse en la parte inferior del cuadro de diálogo.
- **Especificar funciones y campos.** Permite especificar los campos que van a rellenarse o agregarse, así como la función empleada para ambas acciones. Los campos que no se hayan seleccionado se eliminarán del resultado. Use los iconos de la derecha para añadir o eliminar campos de la tabla, o bien pulse en la casilla de la columna pertinente para cambiar la función de agregación o relleno utilizada para sustituir la función por defecto. Los campos sin tipo se excluyen de la lista y no se pueden añadir a la tabla.

**Valores por defecto.** Especifica las funciones de agregación y relleno que se usan por defecto para los distintos tipos de campos. Estos valores por defecto se aplican cuando se selecciona Usar funciones y campos por defecto y, asimismo, constituye la opción por defecto inicial para los nuevos campos que se añaden a la tabla. (Si se cambian los valores por defecto, la configuración existente en la tabla no se verá alterada, pero sí afectará a los campos que se añaden con posterioridad.)

**Función de agregación.** Las funciones de agregación disponibles son:

- **Continuo.** Las funciones disponibles para los campos continuos incluyen Media, Suma, Moda, Mín y Máx.

- **Nominal.** Las opciones incluyen Moda, Primeros y Último. Primeros indica los primeros valores no nulos (ordenados por fecha) en el grupo de agregación, mientras que Último señala el último valor no nulo.
- **Marca.** Las opciones incluyen Verdadero si alguno es verdadero, Moda, Primeros y Último.

**Función de relleno.** Las funciones de relleno disponibles son:

- **Continuo.** Las opciones incluyen Vacío y Media de los puntos recientes, que indica los tres valores no nulos más recientes antes del período de tiempo que se va a crear. Si no existen tres valores, el nuevo estará vacío. Los valores recientes solamente incluyen valores reales. Un valor relleno creado previamente no se tendrá en cuenta en la búsqueda de un valor no nulo.
- **Nominal.** Vacío y Valor más reciente. Con más reciente se hace referencia al valor no nulo más reciente antes del período de tiempo que se va a crear. De nuevo, solamente se tendrán en cuenta los valores reales en la búsqueda de un valor reciente.
- **Marca.** Las opciones incluyen Vacío, Verdadera y Falsa.

**Número máximo de registros del conjunto de datos resultante.** Especifica un límite máximo de número de registros creados, ya que, de lo contrario, podría ser muy elevado, en concreto cuando el intervalo de tiempo se establece en segundos (ya sea deliberadamente o no). Por ejemplo, una serie con solamente dos valores (1 ene, 2000 y 1 ene, 2001) generaría 31.536.000 registros si se rellenan en segundos (60 segundos x 60 minutos x 24 horas x 365 días). El sistema detendrá el proceso y mostrará una advertencia en caso de que se supere el máximo especificado.

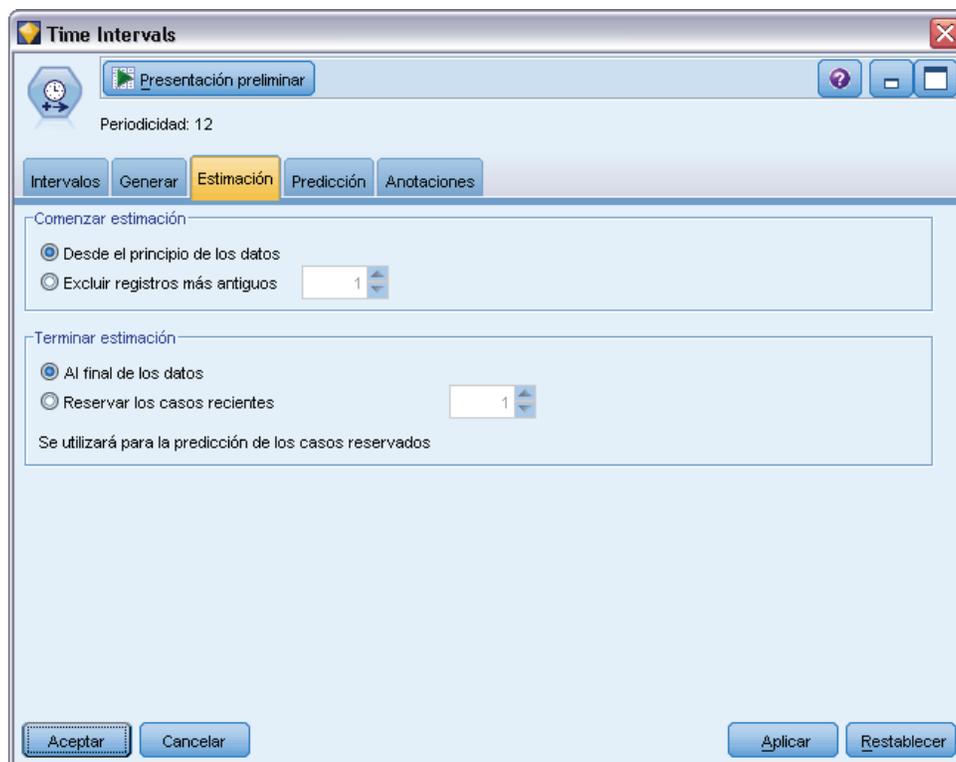
### ***Campo Recuento***

Cuando se agregan o rellenan valores, se crea un campo *Recuento* nuevo que indica el número de registros implicados en la definición del nuevo registro. Si, por ejemplo, se han agregado valores semanales en un solo mes, el recuento sería 4. En el caso de un registro relleno, el recuento es 0. El nombre del campo es *Recuento* más el prefijo o sufijo especificado en la pestaña Intervalos.

## ***Período de estimación***

La pestaña Estimación del nodo Intervalos de tiempo permite especificar el intervalo de registros utilizado en la estimación del modelo, además de los casos reservados. Esta configuración se puede sobrescribir en los nodos de modelado posteriores según sea necesario, pero es conveniente especificarla aquí en lugar de hacerlo para cada nodo individualmente.

Figura 4-86  
Nodo Intervalos de tiempo, pestaña Estimación



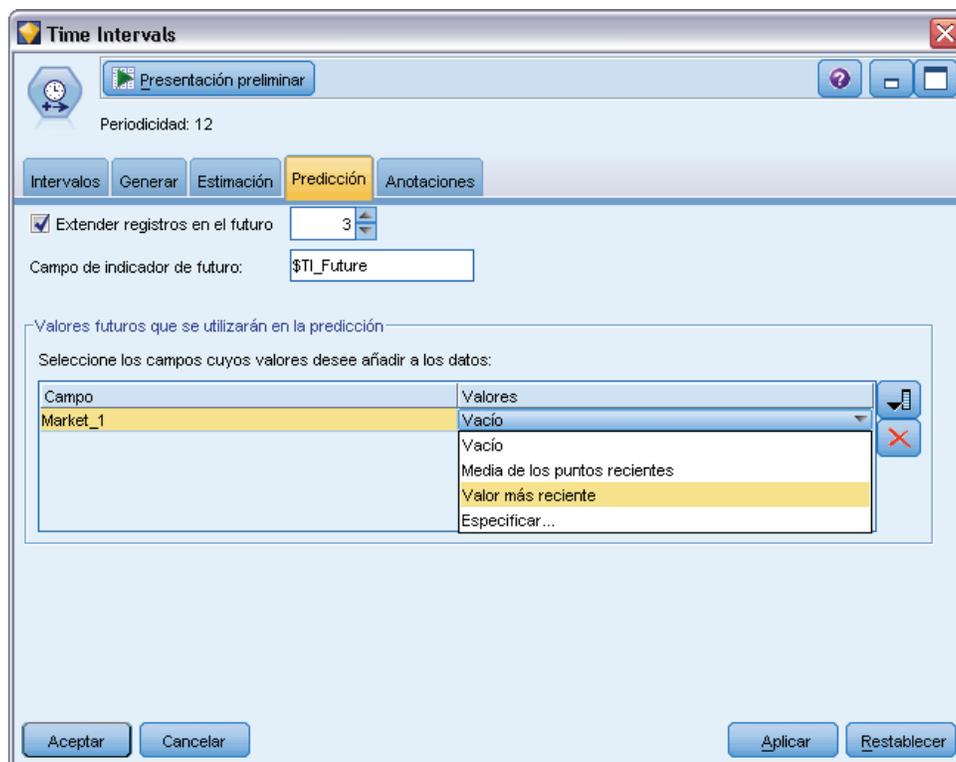
**Comenzar estimación.** Puede comenzar el período de estimación al principio de los datos o excluir los valores antiguos que tal vez no sean muy útiles para la predicción. Dependiendo de los datos, puede que al reducir el período de estimación se acelere el rendimiento (y disminuya el tiempo dedicado a la preparación de los datos) sin que suponga una pérdida significativa en la precisión de la predicción.

**Terminar estimación.** Puede estimar el modelo utilizando todos los registros hasta el final de los datos, o “reservar” los registros más recientes para evaluar el modelo. En el caso siguiente, en realidad está “prediciendo” valores que ya conoce, lo que permite comparar los valores observados y predictivos para evaluar la eficacia del modelo.

## Predicciones

La pestaña Predicción del nodo Intervalos de tiempo permite especificar el número de registros que se desea pronosticar y los valores futuros que se van a utilizar en la predicción mediante nodos de modelado de serie temporal a partir de ese punto de la ruta. Esta configuración se puede sobrescribir en los nodos de modelado posteriores según sea necesario, pero es conveniente especificarla aquí en lugar de hacerlo para cada nodo individualmente.

Figura 4-87  
Nodo Intervalos de tiempo, pestaña Predicción



**Extender registros en el futuro.** Especifica el número de registros que se pronosticarán más allá del período de estimación. Tenga en cuenta que estos registros pueden ser o no “predicciones” según el número de reservas que se haya especificado en la pestaña Estimación.

**Campo de indicador de futuro.** Etiqueta del campo generado que indica si un registro contiene datos de predicciones. El valor por defecto de la etiqueta es *\$TI\_Futuro*.

**Valores futuros que se utilizarán en la predicción.** Si utiliza campos predictores (Dirección = Entrada), para cada registro que desee predecir (con el papel definido a *Entrada*) deberá especificar los valores estimados para el período de predicción para cada predictor. Puede introducir los valores manualmente o seleccionarlos de una lista.

- **Campo.** Pulse en el botón selector de campos y seleccione los campos que se utilizarán como predictores. Tenga en cuenta que los campos aquí seleccionados podrán utilizarse o no en el modelado; para utilizar un campo como predictor debe seleccionarse en un nodo de modelado posterior. Este cuadro de diálogo permite especificar los valores futuros de forma cómoda para que puedan compartirlos varios nodos de modelado posteriores sin tener que especificarlos por separado en cada nodo. Recuerde también que la lista de campos disponibles puede restringirse mediante las selecciones realizadas en la pestaña Generar. Por ejemplo, si se selecciona Especificar funciones y campos en la pestaña Generar, se eliminarán de la ruta todos los campos que no se hayan agregado o rellenado y no podrán utilizarse en el modelado.

*Nota:* si se especifican valores futuros para un campo que ya no está disponible en la ruta (debido a que se ha eliminado o a las selecciones actualizadas realizadas en la pestaña Generar) el campo se mostrará en rojo en la pestaña Predicción.

- **Valores.** Para cada campo, puede seleccionar en una lista de funciones o pulsar en Especificar para introducir los valores manualmente o desde una lista de valores predefinidos. Si los campos predictores se refieren a los elementos que están bajo su control, o que se conocen de otra forma, debe introducir los valores manualmente. Por ejemplo, si está pronosticando los ingresos del próximo mes para un hotel a partir del número de reservas, puede especificar el número de reservas que tiene para ese período. Por el contrario, si un campo predictor está relacionado con algo que escapa a su control, como el precio de las acciones, puede utilizar una función como el valor más reciente o la media de los puntos recientes.

Las funciones disponibles dependen del nivel de medición del campo.

Nivel de medida	Funciones
Campo continuo o nominal	Vacío Media de puntos recientes Valor más reciente Especificar
Campo de marcas	Vacío Valor más reciente True False Especificar

Media de los puntos recientes: calcula el valor futuro a partir de la media de los tres últimos puntos de datos.

Valor más reciente: establece el valor futuro en el punto de datos más reciente.

Verdadero/falso: establece el valor futuro de un campo de marcas en Verdadero o Falso según lo especificado.

Especificar: abre un cuadro de diálogo que permite especificar los valores futuros manualmente o desde una lista predefinida.

Figura 4-88  
Especificación de valores futuros para predictores



### Valores futuros

Aquí, puede especificar valores futuros para utilizarlos en las predicciones realizadas en nodos de modelado de series temporales en puntos posteriores de la ruta. Esta configuración se puede sobrescribir en los nodos de modelado posteriores según sea necesario, pero es conveniente especificarla aquí en lugar de hacerlo para cada nodo individualmente.

Puede introducir los valores manualmente o pulsar en el botón selector situado en el lado derecho del cuadro de diálogo para seleccionarlos de una lista de valores definida para el campo actual.

El número de valores futuros que se puede especificar corresponde al número de registros para los que está extendiendo la serie temporal en el futuro.

### Intervalos admitidos

El nodo Intervalos de tiempo admite una amplia variedad de intervalos, desde segundos a años, así como períodos cíclicos (por ejemplo, estacionales) y no cíclicos. Puede especificar el intervalo en el campo Intervalos de tiempo, en la pestaña Intervalos.

### Períodos

Seleccione Períodos para etiquetar una serie no cíclica existente que no coincida con ninguno de los otros intervalos especificados. La serie debe hallarse ya en el orden correcto, con un intervalo uniforme entre cada medida. La opción Generar a partir de los datos no estará disponible si se selecciona este intervalo.

Figura 4-89  
Configuración del intervalo de tiempo de períodos no cíclicos

Intervalo de tiempo: Períodos

Iniciar etiquetado desde el primer registro  Generar a partir de los datos

Período: 1

Nueva extensión de nombre de campo: \$TI\_ Añadir como:  Prefijo  Sufijo

### Resultados de muestra

Los registros se etiquetan de forma incremental en función del valor de inicio especificado (*Período 1, Período 2, etc.*). Se crean nuevos campos del siguiente modo:

\$TI_ÍndiceTiempo (Entero)	\$TI_EtiquetaTiempo (Cadena)	\$TI_Período (Entero)
1	Período 1	1
2	Período 2	2
3	Período 3	3

\$TI_ÍndiceTiempo (Entero)	\$TI_EtiquetaTiempo (Cadena)	\$TI_Período (Entero)
4	Período 4	4
5	Período 5	5

### Períodos cíclicos

Seleccione **Períodos cíclicos** para etiquetar una serie existente con un ciclo repetitivo que no se ajuste a uno de los intervalos estándar. Por ejemplo, podría usar esta opción si tiene solamente 10 meses en el año fiscal. La serie debe hallarse ya en el orden correcto, con un intervalo uniforme entre cada medida. (La opción **Generar a partir de los datos** no estará disponible si se selecciona este intervalo.)

Figura 4-90  
Configuración del intervalo de tiempo de períodos cíclicos

### Resultados de muestra

Los registros se etiquetan de forma incremental en función del período y ciclo de inicio especificados (*Ciclo 1, Período 1, Ciclo 1, Período 2*, etc.). Por ejemplo, si el número de períodos por ciclo está definido como 3, se crearán nuevos campos del siguiente modo:

\$TI_ÍndiceTiempo (Entero)	\$TI_EtiquetaTiempo (Cadena)	\$TI_Ciclo (Entero)	\$TI_Período (Entero)
1	Ciclo 1, Período 1	1	1
2	Ciclo 1, Período 2	1	2
3	Ciclo 1, Período 3	1	3
4	Ciclo 2, Período 1	2	1
5	Ciclo 2, Período 2	2	2

### Años

En relación con los años, se puede especificar el año de inicio para etiquetar registros consecutivos o seleccionar **Generar a partir de los datos** para especificar un campo de marca de tiempo o de fecha que identifique el año de cada registro.

**Figura 4-91**  
Configuración del intervalo de tiempo de una serie por años

Intervalo de tiempo:

Iniciar etiquetado desde el primer registro  Generar a partir de los datos

Año:

Nueva extensión de nombre de campo:  Añadir como:  Prefijo  Sufijo

### Resultados de muestra

Se crean nuevos campos del siguiente modo:

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Cadena)	\$TI-Año (Entero)
1	2000	2000
2	2001	2001
3	2002	2002
4	2003	2003
5	2004	2004

### Trimestres

En el caso de una serie trimestral, puede especificar el mes en el que comienza el año fiscal. Asimismo, puede indicar el año o el trimestre de inicio (así, T1 2000) para etiquetar registros consecutivos o seleccionar Generar a partir de los datos para elegir un campo de marca de tiempo o de fecha que identifique el trimestre y el año de cada registro.

**Figura 4-92**  
Configuración del intervalo de tiempo de una serie por trimestres

Intervalo de tiempo:

El año fiscal comienza:

Iniciar etiquetado desde el primer registro  Generar a partir de los datos

Año:  Trimestre:

Nueva extensión de nombre de campo:  Añadir como:  Prefijo  Sufijo

**Resultados de muestra**

En relación con un año fiscal que comienza en enero, se crearían y rellenarían campos nuevos de la siguiente forma:

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Cadena)	\$TI-Año (Entero)	\$TI-Trimestre (Entero con etiquetas)
1	T1 2000	2000	1 (T1)
2	T2 2000	2000	2 (T2)
3	T3 2000	2000	3 (T3)
4	T4 2000	2000	4 (T4)
5	T1 2001	2001	1 (T1)

En caso de que el año comience en un mes que no sea enero, los nuevos campos serán los siguientes (suponiendo que un año fiscal empieza en julio). Para ver las etiquetas que identifican los meses de cada trimestre, active la visualización de las etiquetas de los valores pulsando en el icono de la barra de herramientas.

Figura 4-93

icono de etiquetas de valores



\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Cadena)	\$TI-Año (Entero)	\$TI-Trimestre (Entero con etiquetas)
1	T3 2000/2001	1	1 (T1 Jul-Sep)
2	T2 2000/2001	1	2 (T2 Oct-Dic)
3	T3 2000/2001	1	3 (T3 Ene-Mar)
4	T4 2000/2001	1	4 (T4 Abr-Jun)
5	T1 2001/2002	2	1 (T1 Jul-Sep)

**Meses**

Puede seleccionar el año y mes de inicio para etiquetar registros consecutivos o seleccionar Generar a partir de los datos para elegir un campo de marca de tiempo o de fecha que señale el mes de cada registro.

Figura 4-94

Configuración del intervalo de tiempo de una serie por meses

Intervalo de tiempo:

Iniciar etiquetado desde el primer registro  Generar a partir de los datos

Año:  Mes:

Nueva extensión de nombre de campo:  Añadir como:  Prefijo  Sufijo

**Resultados de muestra**

Se crean nuevos campos del siguiente modo:

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Fecha)	\$TI-Año (Entero)	\$TI-Meses (Entero con etiquetas)
1	Ene 2000	2000	1 (Enero)
2	Feb 2000	2000	2 (Febrero)
3	Mar 2000	2000	3 (Marzo)
4	Abr 2000	2000	4 (Abril)
5	May 2000	2000	5 (Mayo)

**Semanas (no periódicas)**

Para una serie semanal, puede seleccionar el día de la semana en el que desea que empiece el ciclo.

Tenga en cuenta que esta opción solamente puede ser no periódica, ya que diferentes meses, trimestres o años, no tienen necesariamente el mismo número de semanas. Si embargo, es muy sencillo agregar o rellenar datos con marcas de tiempo en un nivel semanal para modelos no periódicos.

Figura 4-95

Configuración del intervalo de tiempo de una serie semanal

**Resultados de muestra**

Se crean nuevos campos del siguiente modo:

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Fecha)	\$TI-Semana (Entero)
1	1999-12-27	1
2	2000-01-03	2
3	2000-01-10	3
4	2000-01-17	4
5	2000-01-24	5

El campo *\$TI-EtiquetaTiempo* para una semana muestra el primer día de la semana. En la tabla anterior, el usuario inicia el etiquetado el 1 de enero de 2000. Sin embargo, la semana empieza en Lunes, y el 1 de enero de 2000 es Sábado. Así, la semana que incluye el 1 de enero empieza el 27 de diciembre de 1999, y es la etiqueta del primer punto.

El formato de fecha determina las cadenas generadas para el campo *\$TI-EtiquetaTiempo*.

### Días por semana

En el caso de las medidas por días que se encuentran dentro de un ciclo semanal, puede especificar el número de días a la semana, así como el día en que cada semana empieza. Puede especificar la fecha de inicio para etiquetar registros consecutivos o seleccionar Generar a partir de los datos para elegir un campo de marca de tiempo o de fecha que señale la fecha de cada registro.

Figura 4-96  
Configuración del intervalo de tiempo de una serie por días

### Resultados de muestra

Se crean nuevos campos del siguiente modo:

<b>\$TI-ÍndiceTiempo (Entero)</b>	<b>\$TI-EtiquetaTiempo (Fecha)</b>	<b>\$TI-Semana (Entero)</b>	<b>\$TI-Día (Entero con etiquetas)</b>
1	Ene 5 2005	1	3 (Miércoles)
2	Ene 6 2005	1	4 (Jueves)
3	Ene 7 2005	1	5 (Viernes)
4	Ene 10 2005	2	1 (Lunes)
5	Ene 11 2005	2	2 (Martes)

*Nota:* la semana siempre comienza en 1 para el primer período de tiempo y no sigue un ciclo en función del calendario. Por lo tanto, a la semana 52 le sigue la semana 53, 54, etc. La semana no refleja la semana del año, sino simplemente el número de incrementos semanales que la serie contiene.

### **Días (no periódicos)**

Escoja días no periódicos en caso de que tenga medidas por días que no se ajusten a un ciclo semanal regular. Puede especificar la fecha de inicio para etiquetar registros consecutivos o seleccionar Generar a partir de los datos para elegir un campo de marca de tiempo o de fecha que señale la fecha de cada registro.

Figura 4-97

Configuración del intervalo de tiempo de una serie por días (no periódicos)

Intervalo de tiempo: **Días (no periódicos)**

Iniciar etiquetado desde el primer registro  Generar a partir de los datos

Año: **2000** Mes: **Enero** Día: **1**

Nueva extensión de nombre de campo: **\$TI\_** Añadir como:  Prefijo  Sufijo

Formato de fecha: **AAA-MM-DD**

### **Resultados de muestra**

Se crean nuevos campos del siguiente modo:

<b>\$TI-ÍndiceTiempo (Entero)</b>	<b>\$TI-EtiquetaTiempo (Fecha)</b>
1	Ene 5 2005
2	Ene 6 2005
3	Ene 7 2005
4	Ene 8 2005
5	Ene 9 2005

### **Horas por día**

En el caso de las medidas por horas que se ajusten a un ciclo diario, puede especificar el número de días a la semana, el de horas al día (como un día laboral de ocho horas, por ejemplo), el día en que la semana comienza y la hora a la que cada día empieza. Las horas pueden especificarse incluso por minutos de acuerdo a un reloj de 24 horas (por ejemplo, 14:05 = 2:05 p.m.).

**Figura 4-98**  
Configuración del intervalo de tiempo de una serie por horas

Puede especificar la fecha y la hora de inicio para etiquetar registros consecutivos o seleccionar Generar a partir de los datos para elegir un campo de marca de tiempo que señale la fecha y hora de cada registro.

### **Resultados de muestra**

Se crean nuevos campos del siguiente modo:

<b>\$TI-ÍndiceTiempo (Entero)</b>	<b>\$TI-EtiquetaTiempo (Marca de tiempo)</b>	<b>\$TI-Día (Entero con etiquetas)</b>	<b>\$TI-Hora (Entero con etiquetas)</b>
1	5 Ene 2005 8:00	3 (Miércoles)	8 (8:00)
2	5 Ene 2005 9:00	3 (Miércoles)	9 (9:00)
3	5 Ene 2005 10:00	3 (Miércoles)	10 (10:00)
4	5 Ene 2005 11:00	3 (Miércoles)	11 (11:00)
5	5 Ene 2005 12:00	3 (Miércoles)	12 (12:00)

### **Horas (no periódicas)**

Escoja esta opción en caso de que tenga medidas por horas que no se ajusten a un ciclo diario regular. Puede especificar la hora de inicio para etiquetar registros consecutivos o seleccionar Generar a partir de los datos para elegir un campo de marca de tiempo o de hora que señale la hora de cada registro.

**Figura 4-99**  
Configuración del intervalo de tiempo de datos anuales

Las horas se basan en un reloj de 24 horas (13:00 = 1:00 p.m.) y no se basan en ciclos de 24 horas (es decir, a la hora 24 le sigue la hora 25).

### **Resultados de muestra**

Se crean nuevos campos del siguiente modo:

<b>\$TI-ÍndiceTiempo (Entero)</b>	<b>\$TI-EtiquetaTiempo (Cadena)</b>	<b>\$TI-Hora (Entero con etiquetas)</b>
1	8:00	8 (8:00)
2	9:00	9 (9:00)
3	10:00	10 (10:00)
4	11:00	11 (11:00)
5	12:00	12 (12:00)

### **Minutos por día**

En el caso de las medidas por minuto que se encuentren dentro de un ciclo diario, puede especificar el número de días por semana, el día en el que empieza la semana, el número de horas al día y la hora a la que el día comienza. Las horas se especifican de acuerdo a un reloj de 24 horas y pueden especificarse incluso por minutos y segundos utilizando dos puntos (por ejemplo, 2:05:17 p.m. = 14:05:17). También puede especificar el número de minutos de incremento (cada minuto, cada dos minutos, etc., donde el incremento debe ser un valor múltiplo de 60).

Figura 4-100  
Configuración del intervalo de tiempo de minutos por día

Puede especificar la fecha y la hora de inicio para etiquetar registros consecutivos o seleccionar Generar a partir de los datos para elegir un campo de marca de tiempo que señale la fecha y hora de cada registro.

### Resultados de muestra

Se crean nuevos campos del siguiente modo:

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Marca de tiempo)	\$TI-Minuto
1	2005-01-05 08:00:00	0
2	2005-01-05 08:01:00	1
3	2005-01-05 08:02:00	2
4	2005-01-05 08:03:00	3
5	2005-01-05 08:04:00	4

### Minutos (no periódicos)

Escoja esta opción en caso de que tenga medidas por minuto que no se ajusten a un ciclo diario regular. Puede especificar el número de minutos de incremento (cada minuto, cada dos minutos, etc., donde el valor especificado debe ser múltiplo de 60).

**Figura 4-101**  
Configuración del intervalo de tiempo de minutos (no periódicos)

Intervalo de tiempo: **Minutos (no periódicos)** Incrementar en: **1**

Iniciar etiquetado desde el primer registro  Generar a partir de los datos

Hora:

Nueva extensión de nombre de campo:  Añadir como:  Prefijo  Sufijo

Puede especificar la hora de inicio para etiquetar registros consecutivos o seleccionar Generar a partir de los datos para elegir un campo de marca de tiempo o de hora que señale la hora de cada registro.

### Resultados de muestra

Se crean nuevos campos del siguiente modo:

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Cadena)	\$TI-Minuto
1	8:00	0
2	8:01	1
3	8:02	2
4	8:03	3
5	8:04	4

- La cadena *EtiquetaTiempo* se crea intercalando dos puntos entre la hora y los minutos. Los ciclos no son de 24 horas; es decir, a la hora 24 le sigue la hora 25.
- Los minutos se incrementan según el valor especificado en el cuadro de diálogo. Por ejemplo, si el incremento es 2, la *EtiquetaTiempo* será 8:00, 8:02, etc., mientras que los minutos serán 0, 2, y así sucesivamente.

### Segundos por día

En el caso de los intervalos en segundos que se encuentren dentro de un ciclo diario, puede especificar el número de días por semana, el día en el que empieza la semana, el número de horas al día y la hora a la que el día comienza. Las horas se determinan de acuerdo a un reloj de 24 horas y pueden especificarse incluso por minutos y segundos utilizando dos puntos (por ejemplo, 2:05:17 p.m. = 14:05:17). Asimismo, puede especificar el número de segundos de incremento (cada segundo, cada dos segundos, etc., donde el valor especificado debe ser múltiplo de 60).

Figura 4-102  
Configuración del intervalo de tiempo de segundos por día

Puede especificar la fecha y la hora de inicio para etiquetar registros consecutivos o seleccionar Generar a partir de los datos para elegir un campo de marca de tiempo que señale la fecha y hora de cada registro.

### Resultados de muestra

Se crean nuevos campos del siguiente modo:

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Marca de tiempo)	\$TI-Minuto	\$TI-Segundo
1	2005-01-05 08:00:00	0	0
2	2005-01-05 08:00:01	0	1
3	2005-01-05 08:00:02	0	2
4	2005-01-05 08:00:03	0	3
5	2005-01-05 08:00:04	0	4

### Segundos (no periódicos)

Escoja esta opción en caso de que tenga medidas por segundo que no se ajusten a un ciclo diario regular. Puede especificar el número de segundos de incremento (cada segundo, cada dos segundos, etc., donde el valor especificado debe ser múltiplo de 60).

Figura 4-103  
Configuración del intervalo de tiempo de segundos (no periódicos)

Especifique la hora de inicio para etiquetar registros consecutivos o seleccionar Generar a partir de los datos para elegir un campo de marca de tiempo o de hora que señale la hora de cada registro.

### Resultados de muestra

Se crean nuevos campos del siguiente modo:

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Cadena)	\$TI-Minuto	\$TI-Segundo
1	8:00:00	0	0
2	8:00:01	0	1
3	8:00:02	0	2
4	8:00:03	0	3
5	8:00:04	0	4

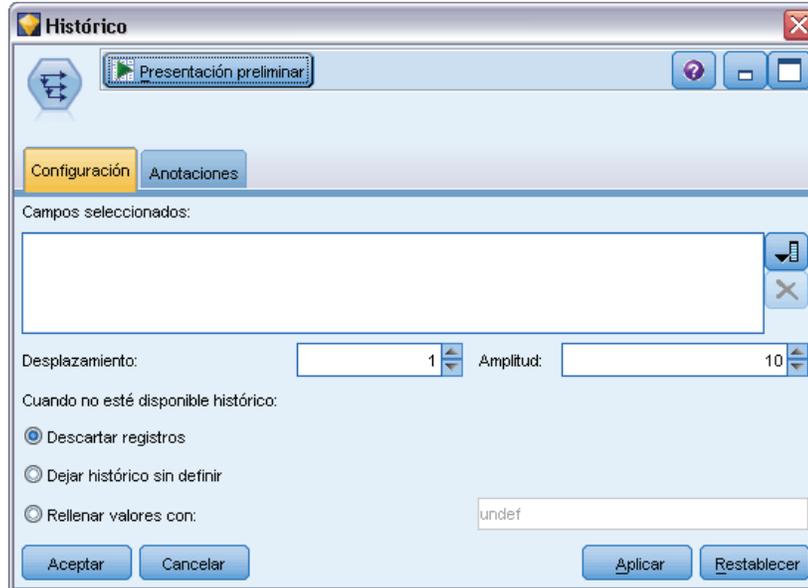
- La cadena *EtiquetaTiempo* se crea intercalando dos puntos entre la hora y los minutos y entre los minutos y los segundos. Los ciclos no son de 24 horas; es decir, a la hora 24 le sigue la hora 25.
- Los segundos aumentan según el número que se haya especificado como incremento. Si el incremento es 2, la *EtiquetaTiempo* será 08:00:00, 08:00:02, etc., mientras que los segundos serán 0, 2, y así sucesivamente.

## Nodo Histórico

Los nodos Histórico se suelen utilizar para los datos secuenciales, como los datos de series temporales. Se utilizan para crear campos nuevos que contienen datos de los campos de registros anteriores. Al utilizar un nodo Histórico, es posible que desee tener los datos ordenados previamente por un campo determinado. Puede utilizar un nodo Ordenar para hacerlo.

## Opciones de configuración del nodo Histórico

Figura 4-104  
Cuadro de diálogo del nodo Histórico



**Campos seleccionados.** El selector de campos (botón situado a la derecha del campo de texto) permite seleccionar campos de los que obtener un histórico. Cada campo seleccionado se utiliza para crear campos nuevos para todos los registros del conjunto de datos.

**Desplazamiento.** Permite especificar el último registro anterior al registro actual desde el que desea extraer valores del campo histórico. Por ejemplo, si el valor de Desplazamiento es 3, a medida que cada registro pasa por este nodo, los valores del campo del tercer registro anterior se incluirán en el registro actual. Utilice la opción Amplitud para especificar el límite hasta el que se pueden extraer registros. Utilice las flechas para ajustar el valor de desplazamiento.

**Amplitud.** Permite especificar la cantidad de registros anteriores de los que desea extraer valores. Por ejemplo, si se establece Desplazamiento en 3 y Amplitud en 5, cada registro que pase por el nodo tendrá cinco campos añadidos para cada campo especificado en la lista Campos seleccionados. Esto significa que cuando el nodo procesa el registro 10, los campos se añaden desde el registro 7 al 3. Utilice las flechas para ajustar el valor de amplitud.

**Cuando no esté disponible histórico.** Permite seleccionar una de las siguientes opciones para tratar los registros que no tengan valores históricos. Suele hacer referencia a los primeros registros de la parte superior del conjunto de datos, de los que no hay registros previos que utilizar como histórico.

- **Descartar registros.** Seleccione esta opción para descartar registros donde no existe ningún valor histórico disponible para el campo seleccionado.

- **Dejar histórico sin definir.** Seleccione esta opción para conservar registros donde no existe ningún valor histórico disponible. El campo histórico se rellenará con un valor no definido que se muestra como \$null\$.
- **Rellenar valores con.** Permite especificar un valor o cadena que utilizar en el caso de los registros en los que no existen valores históricos disponibles. El valor de sustitución por defecto es *undef*, el valor nulo del sistema. Los valores nulos se muestran con la cadena \$null\$.

Al seleccionar un valor de sustitución, tenga en cuenta las siguientes reglas para realizar una ejecución correcta.

- Los campos seleccionados deben ser del mismo tipo de almacenamiento.
- Si todos los campos seleccionados tienen un almacenamiento numérico, el valor de sustitución debe analizarse como un número entero.
- Si todos los campos seleccionados tienen un almacenamiento real, el valor de sustitución debe analizarse como un número real.
- Si todos los campos seleccionados tienen un almacenamiento simbólico, el valor de sustitución debe analizarse como una cadena.
- Si todos los campos seleccionados tienen un almacenamiento de fecha/hora, el valor de sustitución debe analizarse como un campo de fecha/hora.

Si ninguna de las condiciones anteriores se cumple, recibirá un error al ejecutar el nodo Histórico.

## ***Nodo Reorg. campos***

El nodo Reorg. campos permite definir el orden natural utilizado para mostrar los campos en la parte posterior de la ruta. Este orden afecta a la visualización de los campos en diversas ubicaciones, como las tablas, las listas y el selector de campos. Esta operación resulta útil, por ejemplo, al trabajar con conjuntos de datos amplios para aumentar la visibilidad de los campos de interés.

### ***Opciones de configuración de Reorg. campos***

Existen dos opciones de reorganización de los campos: orden personalizado y orden automático.

#### ***Orden personalizado***

Seleccione Orden personalizado para activar una tabla de nombres de campos y tipos donde pueda ver todos los campos y utilizar los botones de flechas para crear un orden personalizado.

Figura 4-105  
Reordenamiento para mostrar primero los campos de interés



Para reordenar campos:

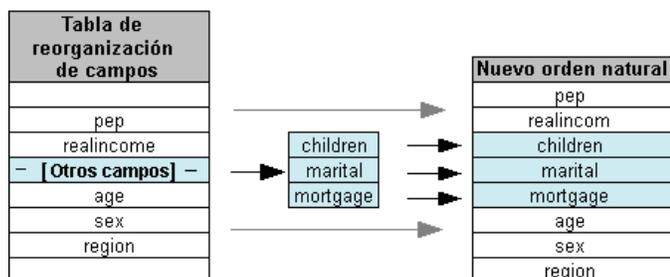
- ▶ Seleccione un campo de la tabla. Mantenga pulsada la tecla Ctrl mientras selecciona en otro campo para seleccionar series de campos.
- ▶ Utilice los botones de flechas sencillas para desplazar los campos una fila arriba o abajo.
- ▶ Utilice los botones de flecha con línea para mover los campos a la parte superior o inferior de la lista.
- ▶ Especifique el orden de los campos no incluidos desplazando la fila separadora hacia arriba o hacia abajo (con el texto [otros campos]).

**Otros campos.** El objetivo de la fila separadora [otros campos] es dividir la tabla en dos mitades.

- Los campos que aparecen por encima de la fila separadora se ordenarán (como aparecen en la tabla) en la parte superior de todos los órdenes naturales utilizados para mostrar los campos hacia abajo de este nodo.
- Los campos que aparecen por debajo de la fila separadora se ordenarán (como aparecen en la tabla) en la parte inferior de todos los órdenes naturales utilizados para mostrar los campos hacia abajo de este nodo.

Figura 4-106

Diagrama que ilustra cómo se incorporan "otros campos" al nuevo orden de campos



- Los demás campos que no aparecen en la tabla de reorganización de campos se incluyen entre los campos "superior" e "inferior" tal y como queda indicado por la ubicación de la fila separadora.

Las opciones de ordenación personalizada adicionales son:

- Ordenar los campos en orden ascendente o descendente pulsando en las flechas ubicadas sobre cada encabezado de columna (Tipo, Nombre y Almacenamiento). Al ordenar por columna, los campos que no están especificados aquí (indicados por la fila [otros campos]) se ordenan en último lugar en orden natural.
- Pulsar en Borrar no utilizados para eliminar todos los campos no utilizados del nodo Reorg. campos. Los campos no utilizados se muestran en la tabla con letras de color rojo. Esto indica que el campo se ha eliminado en las operaciones anteriores.
- Especificar orden para cualquier campo nuevo (se muestra con un icono luminoso para indicar un campo nuevo o sin especificar). Al pulsar en Aceptar o Aplicar, el icono desaparece.

*Nota:* si los campos se añaden hacia arriba (parte anterior de la ruta) tras aplicar un orden personalizado, los campos nuevos se añaden al final de la lista personalizada.

### Ordenación automática

Seleccione Ordenación automática para especificar un parámetro de ordenación. Las opciones del cuadro de diálogo cambian de forma dinámica para proporcionar opciones de ordenación automática.

Figura 4-107

Reordenación de todos los campos con las opciones de ordenación automáticas

**Ordenar por.** Permite seleccionar una de las tres formas de ordenar los campos leídos en el nodo Reorg. campos. Los botones de flecha indican si el orden será ascendente o descendente. Seleccione uno para realizar un cambio.

- Nombre

- Tipo
- Almacenamiento

Los campos del nodo Reorg. campos añadidos anteriormente tras haber aplicado la ordenación automática se colocarán automáticamente en su posición correcta en función del tipo de ordenación seleccionado.

# Nodos de gráficos

## Características comunes de nodos de gráficos

Muchas fases del proceso de minería de datos utilizan gráficos y diagramas para explorar los datos introducidos en IBM® SPSS® Modeler. Por ejemplo, puede conectar un nodo Gráfico o Distribución a una fuente de datos para comprender mejor los tipos de datos y las distribuciones. A continuación puede llevar a cabo manipulaciones de registros y campos para preparar los datos para las operaciones de modelado posteriores de la ruta. Otro uso común de los gráficos consiste en comprobar la distribución y las relaciones entre campos recién derivados.

La paleta Gráficos contiene los siguientes nodos:



El nodo Tablero ofrece muchos tipos diferentes de gráficos en un único nodo. Con este nodo puede seleccionar los campos de datos que desee explorar y seleccionar un gráfico de los disponibles para los datos seleccionados. El nodo filtra automáticamente cualquier tipo de gráfico que no sea compatible con las selecciones de campo. Si desea obtener más información, consulte el tema [Nodo Tablero](#) el p. 251.



El nodo Gráfico muestra la relación entre los campos numéricos. Puede crear un gráfico mediante puntos (un diagrama de dispersión) o líneas. Si desea obtener más información, consulte el tema [Nodo Gráfico](#) el p. 299.



El nodo Distribución muestra las instancias de valores simbólicos (categóricos), como el tipo de hipoteca o el género. Normalmente, podría usar el nodo Distribución para mostrar los desequilibrios de los datos, que pueden rectificarse mediante el nodo Equilibrar antes de crear un modelo. Si desea obtener más información, consulte el tema [Nodo Distribución](#) el p. 308.



El nodo Histograma muestra las instancias de valores de los campos numéricos. Se suele utilizar para explorar los datos antes de las manipulaciones y la generación de modelos. Al igual que con el nodo Distribución, con frecuencia el nodo Histograma detecta desequilibrios en los datos. Si desea obtener más información, consulte el tema [Pestaña Gráfico de Histograma](#) el p. 314.



El nodo Colección muestra la distribución de valores de un campo numérico relativo a los valores de otro. (Crea gráficos parecidos a los histogramas.) Es útil para ilustrar una variable o un campo cuyos valores cambian con el tiempo. Con los gráficos 3D también puede incluir un eje simbólico que muestra las distribuciones por categoría. Si desea obtener más información, consulte el tema [Pestaña Gráfico de Colección](#) el p. 318.



El nodo G. múltiple crea un gráfico que muestra varios campos *Y* sobre un campo *X* único. Los campos *Y* están representados como líneas coloreadas; cada uno equivale a un nodo Gráfico con el estilo establecido en Línea y el Modo para *X* establecido en Ordenar. Los gráficos múltiples son útiles cuando quiere explorar la fluctuación de varias variables a través del tiempo. Si desea obtener más información, consulte el tema [Nodo G. múltiple](#) el p. 322.



El nodo Malla muestra la fuerza de las relaciones entre los valores de dos o más campos simbólicos (categóricos). El gráfico utiliza líneas de varios anchos para indicar la fuerza de la conexión. Podría utilizar un nodo Malla, por ejemplo, para explorar las relaciones existentes entre la compra de un conjunto de elementos en un sitio de comercio electrónico. Si desea obtener más información, consulte el tema [Nodo Malla](#) el p. 326.



El nodo Gráfico de tiempo muestra uno o más conjuntos de datos de series temporales. Normalmente, primero se utilizaría un nodo Intervalos de tiempo para crear un campo *EtiquetaTiempo*, que se utilizaría para etiquetar el eje *x*. Si desea obtener más información, consulte el tema [Nodo Gráfico de tiempo](#) el p. 338.



El nodo Evaluación ayuda a evaluar y comparar modelos predictivos. El diagrama de evaluación muestra la calidad con que los modelos pronostican resultados particulares. Ordena registros en función del valor pronosticado y la confianza del pronóstico. Divide el registro en grupos de igual tamaño (**cuantiles**) y, a continuación, representa el valor del criterio de negocios de cada cuantil de mayor a menor. El gráfico muestra múltiples modelos como líneas independientes. Si desea obtener más información, consulte el tema [Nodo Evaluación](#) el p. 341.

Una vez añadido un nodo de gráficos a una ruta, puede pulsar dos veces en el nodo para abrir un cuadro de diálogo para especificar opciones. La mayoría de los gráficos contienen un número de opciones únicas presentadas en una o varias pestañas. También existen varias opciones de pestaña comunes a todos los gráficos. Las siguientes secciones contienen más información acerca de estas opciones comunes.

Una vez configuradas las opciones de un nodo de gráficos, puede ejecutarlo desde el cuadro de diálogo o como parte de una ruta. En la ventana del gráfico que se genera, puede crear nodos Derivar (Conjunto y Marca) y Seleccionar basados en una selección o región de datos, estableciendo los datos en subconjuntos de forma eficaz. Por ejemplo, puede utilizar esta potente función para identificar y excluir valores atípicos.

## ***Aspecto estético, superposiciones, paneles y animación***

### ***Superposiciones y aspectos estéticos***

Los aspectos estéticos (y superposiciones) añaden dimensionalidad a una visualización. El efecto de un aspecto estético (agrupación, conglomeración o apilación) depende del tipo de visualización, el tipo de campo (variable) y el tipo y estadístico del elemento gráfico. Por ejemplo, puede utilizarse un campo categórico de color para agrupar puntos en un diagrama de dispersión o para crear las pilas de un gráfico de barras apiladas. También puede utilizarse un rango numérico continuo de color para indicar los valores del rango para cada punto de un diagrama de dispersión.

Debería experimentar con los aspectos estéticos y superposiciones para descubrir qué cubre sus necesidades. Las siguientes descripciones pueden ayudarle a elegir lo correcto.

*Nota:* No todos los aspectos estéticos o superposiciones están disponibles para todos los tipos de visualización.

- **Color.** Cuando un campo categórico define el color, divide la visualización basándose en las categorías individuales, correspondiendo cada color a una categoría distinta. Cuando el color es un rango numérico continuo, el color varía basándose en el valor del campo de rango. Si el elemento gráfico (por ejemplo, una barra o un cuadro) representa a más de un

registro o caso y se utiliza un campo de rango para el color, éste varía dependiendo de la *media* del campo de rango.

- **Forma.** La forma se define a través de un campo categórico que divide la visualización en elementos de formas diferentes, uno por cada categoría.
- **Transparencia.** Cuando un campo categórico define la transparencia, divide la visualización basándose en las categorías individuales, correspondiendo cada nivel de transparencia a cada categoría. Cuando la transparencia es un rango numérico continuo, la transparencia varía basándose en el valor del campo de rango. Si el elemento gráfico (por ejemplo, una barra o un cuadro) representa a más de un registro o caso y se utiliza un campo de rango para la transparencia, el color varía dependiendo de la *media* del campo de rango. En el valor más alto, los elementos gráficos son totalmente transparentes. En el valor más bajo, son totalmente opacos.
- **Etiqueta de datos.** Las etiquetas de datos se definen mediante cualquier tipo de campo cuyos valores se utilicen para crear etiquetas que se adjunten a los elementos gráficos.
- **Tamaño.** Cuando un campo categórico define el tamaño, divide la visualización basándose en las categorías individuales, correspondiendo cada tamaño a una categoría distinta. Cuando el tamaño es un rango numérico continuo, el tamaño varía basándose en el valor del campo de rango. Si el elemento gráfico (por ejemplo, una barra o un cuadro) representa a más de un registro o caso y se utiliza un campo de rango para el tamaño, éste varía dependiendo de la *media* del campo de rango.

Figura 5-1

Gráfico con un aspecto estético de superposición de colores

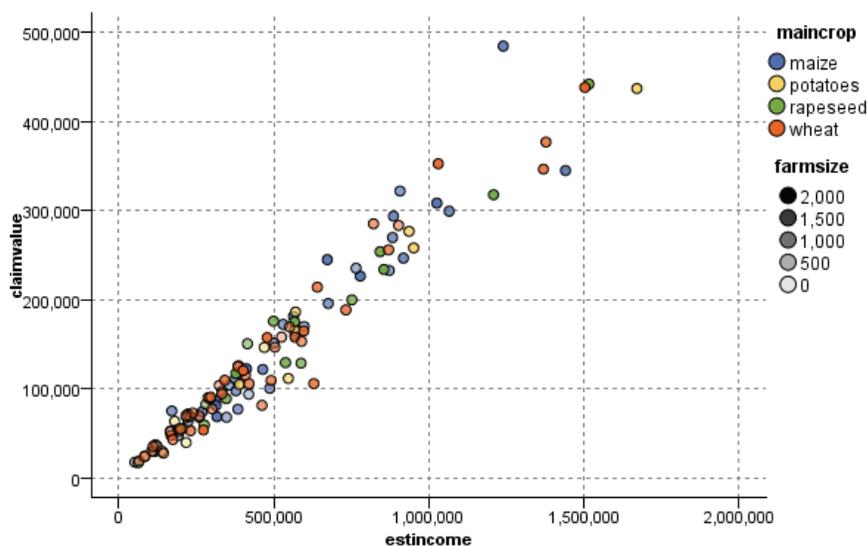
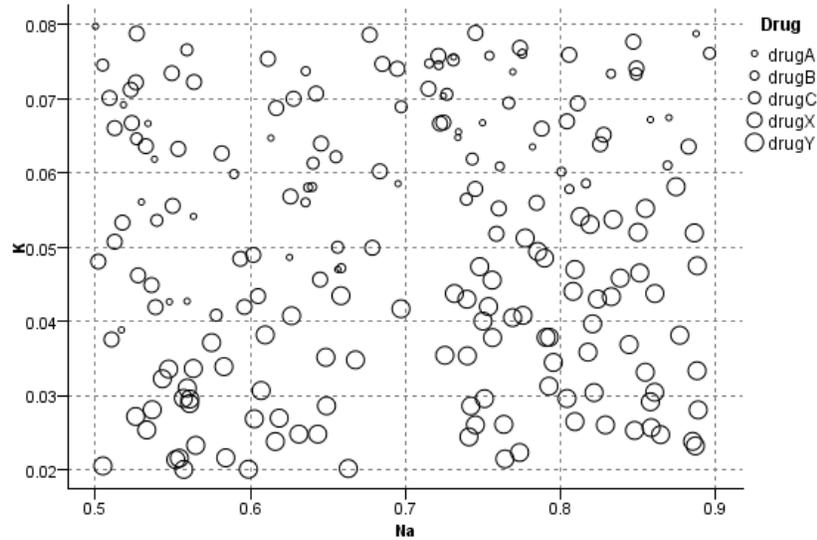


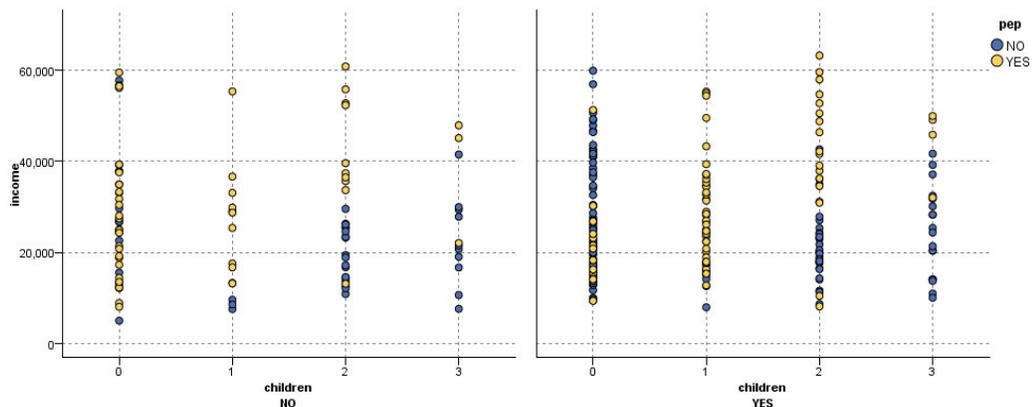
Figura 5-2  
Gráfico con un aspecto estético de superposición de tamaños



### Paneles y animación

**Adición de paneles.** La opción Paneles, también conocida como facetas, crea una tabla de gráficos. Se genera un gráfico por cada categoría en los campos de paneles, pero todos los paneles aparecerán simultáneamente. Paneles es de utilidad para comprobar si la visualización está sujeta a las condiciones de los campos de paneles. Por ejemplo, puede incluir un histograma en un panel según género para determinar si las distribuciones de frecuencia son iguales entre hombres y mujeres. Es decir, puede comprobar si el salario está sujeto a diferencias de sexo. Seleccione un campo categórico para paneles.

Figura 5-3  
Gráfico con paneles según estado civil (SÍ/NO)



**Animación.** La opción Animación se parece a la opción Paneles en cuanto a que se crean varios gráficos a partir de los valores del campo de animación, pero estos gráficos no se muestran juntos. Más bien, utiliza los controles en el modo de exploración para animar el resultado y desplazarse por una secuencia de gráficos individuales. Además, a diferencia de la opción Paneles, Animación no necesita un campo categórico. Puede especificar un campo continuo cuyos valores estén divididos en rangos automáticamente. Puede variar el tamaño del rango con los controles de animación en el modo de exploración. No todas las visualizaciones ofrecen animación.

Figura 5-4

Gráfico animado con una variable con tres categorías, control deslizante en la tensión arterial baja

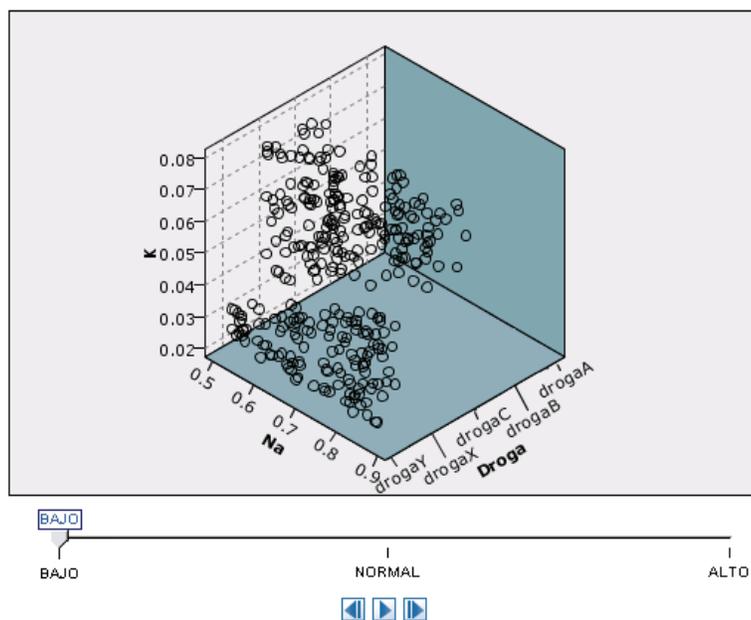


Figura 5-5  
Gráfico animado con una variable con tres categorías, control deslizante en la tensión arterial normal

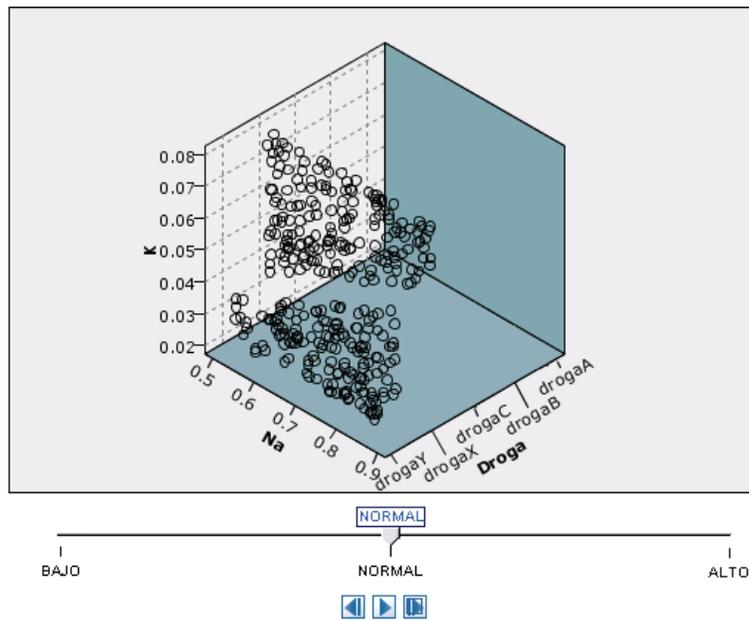
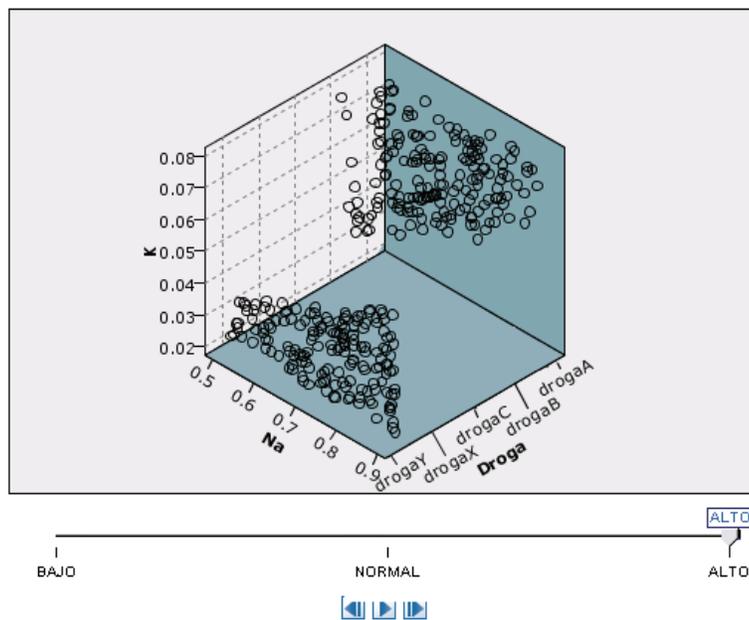


Figura 5-6  
Gráfico animado con una variable con tres categorías, control deslizante en la tensión arterial alta



## Uso de la pestaña Resultados

En el caso de todos los tipos de gráficos, puede especificar las siguientes opciones para el nombre de archivo y la visualización de los gráficos generados.

*Nota:* Los gráficos de nodo de distribución tienen una configuración adicional.

**Nombre del resultado.** Especifica el nombre del gráfico generado cuando se ejecuta el nodo. Automático selecciona un nombre según el nodo que genera el resultado. De forma opcional, puede seleccionar Personalizado para especificar un nombre diferente.

**Salida a pantalla.** Seleccione esta opción para generar y mostrar el gráfico en una nueva ventana.

**Salida a archivo.** Seleccione para guardar el resultado como un archivo.

- **Gráfico de resultados.** Seleccione para producir resultados en un formato de gráfico. Disponibles solamente en los nodos Distribución.
- **Tabla de resultados.** Seleccione para producir resultados en un formato de tabla. Disponibles solamente en los nodos Distribución.
- **Nombre de archivo.** Permite especificar un nombre de archivo para el gráfico o tabla generados. Utilice el botón de puntos suspensivos (...) para especificar una ubicación y un archivo específicos.
- **Tipo de archivo.** Especifique el tipo de archivo de la lista desplegable. Para todos los nodos de gráfico, excepto el nodo Distribución con una opción Tabla de resultados, los tipos de archivo de gráficos disponibles son los siguientes:
  - Mapa de bits (*.bmp*)
  - PNG (*.png*)
  - Objeto Resultado (*.cou*)
  - JPEG (*.jpg*)
  - HTML (*.html*)
  - Documento ViZml (*.xml*) para utilizarlo en otras aplicaciones de IBM® SPSS® Statistics.

Para la opción Tabla de resultados en el nodo Distribución, los tipos de archivo disponibles son los siguientes:

- Datos delimitados por tabulaciones (*.tab*)
- Datos delimitados por comas (*.csv*)
- HTML (*.html*)
- Objeto Resultado (*.cou*)

**Paginar resultados.** Al guardar la salida como HTML, esta opción se activa para permitir controlar el tamaño de cada página HTML. (Se aplica solamente al nodo Distribución.)

**Líneas por página.** Cuando se selecciona Pagar resultados, esta opción se activa para permitir determinar la longitud de cada página HTML. La configuración por defecto es de 400 líneas. (Se aplica solamente al nodo Distribución.)

## Uso de la pestaña Anotaciones

Se utiliza para todos los nodos, esta pestaña proporciona opciones para cambiar el nombre de los nodos, ofrecer herramientas personalizadas y almacenar una anotación más extensa.

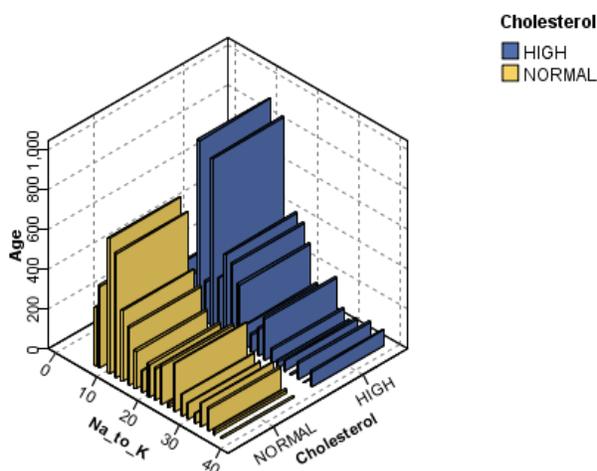
## Gráficos 3D

Los gráficos y los gráficos de colección de IBM® SPSS® Modeler tienen la capacidad de mostrar información en un tercer eje. Así se dispone de una mayor flexibilidad para visualizar los datos para seleccionar subconjuntos o derivar campos nuevos para el modelado.

Una vez haya creado un gráfico 3D, puede pulsar en él y arrastrar el ratón para girarlo y poder verlo desde cualquier ángulo.

Figura 5-7

Gráfico de colección con ejes x, y, z



Hay dos formas de crear gráficos 3D en SPSS Modeler: representar información en un tercer eje (gráficos 3D verdaderos) y mostrar gráficos con efectos 3D. Ambos métodos están disponibles para gráficos y colecciones.

### Para representar información en un tercer eje

- ▶ En el cuadro de diálogo del nodo de gráficos, pulse en la pestaña Gráfico.
- ▶ Pulse en el botón 3D para activar las opciones del eje z.
- ▶ Utilice el botón de elección de campos para seleccionar un campo para el eje z. En algunos casos solamente se permiten los campos simbólicos. El selector de campos mostrará los campos adecuados.

**Para añadir efectos 3D a un gráfico**

- ▶ Una vez creado un gráfico, pulse en la pestaña Gráfico en la ventana de resultados.
- ▶ Pulse en el botón 3D para cambiar la vista a un gráfico tridimensional.

## **Nodo Tablero**

El nodo Tablero permite seleccionar distintos resultados de gráficos (gráficos de barras, gráfico de sectores, histogramas, diagramas de dispersión, mapas de calor, etc.) en un único nodo. Empieza en la primera pestaña seleccionando los campos de datos que desea explorar y, a continuación, el nodo le proporciona una selección de tipos de gráficos que se adaptan a sus datos. El nodo filtra automáticamente cualquier tipo de gráfico que no sea compatible con las selecciones de campo. Puede definir opciones de gráfico más avanzadas o detalladas en la pestaña Detallado.

*Nota:* Debe conectar el nodo Tablero a una ruta con los datos para poder editar el nodo o seleccionar tipos de gráficos.

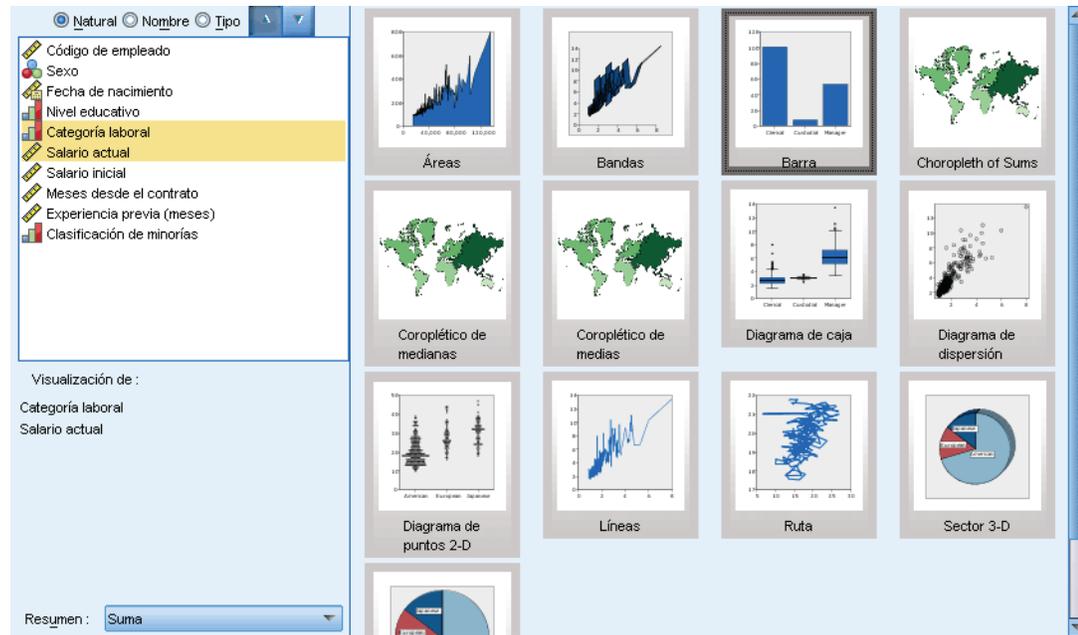
Existen dos botones que permiten controlar las plantillas de visualización (y hojas de estilo y mapas) disponibles:

**Administrar.** Administra plantillas, hojas de estilo y mapas de visualización en su ordenador. Puede importar, exportar, cambiar el nombre y eliminar plantillas, hojas de estilo y mapas de visualización en su máquina local. Si desea obtener más información, consulte el tema [Administración de plantillas, hojas de estilo y archivos de mapas](#) el p. 289.

**Ubicación.** Cambia la ubicación donde están almacenadas las plantillas, hojas de estilo y mapas de visualización. La ubicación actual se enumera a la derecha del botón. Si desea obtener más información, consulte el tema [Configuración de la ubicación de las plantillas, hojas de estilo y mapas](#) el p. 287.

## Tablero Ficha Básico

Figura 5-8  
Ficha Básico



Si no está seguro de qué tipo de visualización representará mejor sus datos, utilice la ficha Básico. Cuando seleccione sus datos, aparecerá un subconjunto de tipos de visualizaciones adecuados para los datos. Para obtener ejemplos, consulte [Ejemplos de Tableros](#) el p. 267.

- Seleccione uno o más campos (variables) de la lista. Mantenga pulsada la tecla Ctrl para seleccionar varios campos.

Tenga en cuenta que el nivel de medida del campo determina el tipo de visualizaciones disponibles. Puede cambiar el nivel de medida pulsando con el botón derecho en el campo de la lista y seleccionando una opción. Para obtener más información sobre los tipos de niveles de medición, consulte [Tipos de campos \(variables\)](#) el p. 254.

- Seleccione un tipo de visualización. Para obtener descripciones de los tipos disponibles, consulte [Tipos de visualización de Graphboard preincorporado disponibles](#) el p. 259.
- En el caso de ciertas visualizaciones, puede seleccionar un estadístico de resumen. Hay diferentes subconjuntos de estadísticos disponibles dependiendo de si el estadístico está basado en recuentos o si se ha calculado a partir de un campo continuo. Los estadísticos disponibles también dependen de la misma plantilla. Una lista completa de los estadísticos que pueden estar disponibles sigue al paso siguiente.
- Si desea definir más opciones, como los campos de aspectos estéticos y paneles opcionales, pulse en Detallado. Si desea obtener más información, consulte el tema [Tablero Pestaña Detallado](#) el p. 257.

**Estadísticos de resumen calculados a partir de un campo continuo**

- **Media.** Una medida de tendencia central. El promedio aritmético, la suma dividida por el número de casos.
- **Mediana.** Es el valor por encima y por debajo del cual se encuentran la mitad de los casos, el percentil 50. Si hay un número par de casos, la mediana es la media de los dos valores centrales, cuando los casos se ordenan en orden ascendente o descendente. La mediana es una medida de tendencia central que no es sensible a los valores atípicos (a diferencia de la media, que puede resultar afectada por unos pocos valores extremadamente altos o bajos).
- **Moda.** El valor que ocurre con mayor frecuencia. Si varios valores comparten la mayor frecuencia de aparición, cada uno de ellos es un modo.
- **Mínimo.** Valor más pequeño de una variable numérica.
- **Máximo.** El mayor valor de una variable numérica.
- **Rango.** La diferencia entre los valores mínimos y máximos.
- **Medio intervalo.** La mitad del rango, es decir, el valor cuya diferencia con el mínimo es igual a la diferencia con el máximo.
- **Sum.** Suma o total de todos los valores, a lo largo de todos los casos que no tengan valores perdidos.
- **Suma acumulada.** La suma acumulada de los valores. Cada elemento del gráfico muestra la suma de un subgrupo más la suma total de los grupos anteriores.
- **Porcentaje de la suma.** El porcentaje en cada subgrupo basado en un campo sumado en comparación con la suma en todos los grupos.
- **Suma del porcentaje acumulado.** El porcentaje acumulativo en cada subgrupo basado en un campo sumado en comparación con la suma en todos los grupos. Cada elemento del gráfico muestra el porcentaje de un subgrupo más el porcentaje total de los grupos anteriores.
- **Variance.** Es una medida de dispersión en torno a la media, igual a la suma de las desviaciones al cuadrado respecto a la media, dividida por el número de casos menos 1. La varianza se mide en unidades que son el cuadrado de las de la variable en cuestión.
- **Desviación típica.** Es una medida de la dispersión en torno a la media. En una distribución normal, el 68% de los casos se encuentra dentro de una desviación típica de la media y el 95% queda entre dos desviaciones típicas. Por ejemplo, si la edad media es de 45 años, con una desviación típica de 10, el 95% de los casos estaría entre los 25 y 65 en una distribución normal.
- **Error típico.** Medida de cuánto puede variar el valor de un estadístico de contraste de muestra en muestra. Es la desviación típica de la distribución muestral de un estadístico. Por ejemplo, el error típico de la media es la desviación típica de las medias muestrales.
- **Kurtosis.** Medida del grado en que las observaciones están agrupadas en torno al punto central. Para una distribución normal, el valor del estadístico de kurtosis es 0. Una kurtosis positiva indica que, con respecto a una distribución normal, las observaciones se concentran más en el centro de la distribución y presentan colas más estrechas hasta los valores extremos de la distribución, en cuyo punto las colas de la distribución leptocúrtica son más gruesas con respecto a una distribución normal. Una kurtosis negativa indica que, con respecto a una distribución normal, las observaciones se concentran menos y presentan colas más gruesas

hasta los valores extremos de la distribución, en cuyo punto las colas de la distribución platicúrtica son más estrechas con respecto a una distribución normal.

- **Asimetría.** Medida de la asimetría de una distribución. La distribución normal es simétrica y tiene un valor de asimetría igual a 0. Una distribución que tenga una asimetría positiva significativa tiene una cola derecha larga. Una distribución que tenga una asimetría negativa significativa tiene una cola izquierda larga. Como regla aproximada, un valor de la asimetría mayor que el doble de su error típico se asume que indica una desviación de la simetría.

Los siguientes estadísticos de región pueden resultar en más de un elemento gráfico por subgrupo. Si utiliza los elementos de intervalo, área, o gráficos de borde, un estadístico de región da como resultado un elemento gráfico que muestra el rango. El resto de elementos gráficos dan como resultado dos elementos diferentes, uno que muestra el inicio del rango y el otro que muestra el final del rango.

- **Región: Rango.** El rango entre los valores mínimo y máximo.
- **Región: 95% de intervalo de confianza para la media.** Un rango de valores con un 95% de opciones de incluir la media poblacional.
- **Región: 95% de intervalo de confianza individual.** Un rango de valores con un 95% de opciones de incluir el valor predictivo especificando el caso individual.
- **Región: 1 Desviación típica por encima/debajo de la media.** Un rango de valores entre 1 desviación típica por encima y por debajo de la media.
- **Región: 1 Error típico por encima/debajo de la media.** Un rango de valores entre 1 error típico por encima y por debajo de la media.

#### **Estadísticos de resumen basados en recuentos**

- **Recuento.** Número de filas/casos.
- **Recuento acumulado.** El número acumulado de filas/casos. Cada elemento del gráfico muestra el recuento de un subgrupo más el recuento total de los grupos anteriores.
- **Porcentaje de recuento.** El porcentaje de filas/casos de cada subgrupo en comparación con el número total de filas/casos.
- **Porcentaje acumulado de recuento.** El porcentaje acumulado de filas/casos de cada subgrupo en comparación con el número total de filas/casos. Cada elemento del gráfico muestra el porcentaje de un subgrupo más el porcentaje total de los grupos anteriores.

#### **Tipos de campos (variables)**

Los iconos que aparecen junto a los campos en las listas de campos para indicar el tipo de campo y los tipos de datos. Los iconos también identifican múltiples conjuntos de respuestas.

Nivel de medida	Tipo de datos			
	Numeric	Cadena	Date	Time
Continuo		n/a		

Conjunto ordenado				
Set				
Conjuntos de respuestas múltiples, categorías múltiples				
Conjunto de respuestas múltiples, dicotomías múltiples				

### **Nivel de medida**

Un nivel de medida de campo es importante cuando crea una visualización. A continuación se incluye una descripción de los niveles de medida. Puede cambiar temporalmente el nivel de medida pulsando con el botón derecho en un campo en una lista de campos y seleccionando una opción. En la mayoría de los casos, debe considerar únicamente las dos clasificaciones más amplias de los campos, categóricas y continuas:

**Categóricas.** Los datos con un número limitado de valores o categorías distintos (por ejemplo, sexo o religión). Los campos categóricos pueden ser campos de cadena (alfanuméricos) o campos numéricos que utilizan códigos numéricos para representar categorías (por ejemplo, 0 = *hombre* y 1 = *mujer*). También se denominan datos cualitativos. Los conjuntos, conjuntos ordenados y marcas son campos categóricos.

- 
- 
- 

**Continua.** Datos medidos en un intervalo o escala de cociente, donde los datos indican el orden de valores y la distancia entre valores. Por ejemplo, un salario de 72.195 \$ es superior que un salario de 52.398 \$ y la distancia entre ambos valores es de 19.797 \$. También se denominan datos cuantitativos, de escala o de rango numérico.

Los campos categóricos definen categorías en la visualización, normalmente para dibujar diferentes elementos gráficos o para agrupar elementos gráficos. Los campos continuos se suelen resumir en categorías de campos categóricos. Por ejemplo, una visualización por defecto de ingresos de categorías de género mostrará los ingresos medios de hombres y los ingresos medios de mujeres. Los valores brutos de los campos continuos también se pueden representar, como en un diagrama de dispersión. Por ejemplo, un diagrama de dispersión puede mostrar el salario actual y el salario inicial de cada caso. Un campo categórico se puede usar para agrupar los casos por género.

### ***Tipos de datos***

El nivel de medición no es la única propiedad de un campo que determina su tipo. Un campo también se guarda como un tipo de datos específico. Los tipos de datos posibles son cadenas (datos no numéricos como letras), valores numéricos (números reales) y fechas. Al contrario que el nivel de medición, un tipo de datos de campo no se puede modificar temporalmente. Debe cambiar la forma en la que se guardan los datos en el conjunto de datos originales.

### ***Conjuntos de resp. múltiples***

Algunos archivos de datos admiten un tipo especial de “campo” llamado **conjunto de respuestas múltiples**. Los conjuntos de respuestas múltiples no son “campos” realmente en el sentido normal. Los conjuntos de respuestas múltiples utilizan múltiples campos para registrar respuestas a preguntas donde el encuestado puede ofrecer más de una respuesta. Los conjuntos de respuestas múltiples se tratan como campos categóricos y la mayoría de las acciones que puede realizar con los campos categóricos, también las puede realizar con los conjuntos de respuestas múltiples.

Los conjuntos de respuestas múltiples pueden ser conjuntos de dicotomía múltiple o conjuntos de categoría múltiple.

**conjuntos de dicotomías múltiples.** Un conjunto de dicotomía múltiple suele estar formado por múltiples campos dicotómicos: campos con sólo dos valores posibles de naturaleza sí/no, presente/ausente, marcado/no marcado. Aunque los campos no puedan ser estrictamente dicotómicos, todos los campos del conjunto están codificados de la misma forma.

Por ejemplo, una encuesta proporciona cinco respuestas posibles para la pregunta, “¿En cuál de los siguientes orígenes confía para las noticias?” El encuestado puede indicar varias opciones seleccionando una casilla de verificación junto a cada opción. Las cinco respuestas serán cinco campos en los archivos de datos, con el código 0 para *No* (no marcado) y 1 para *Sí* (marcado).

**Conjuntos de categoría múltiple.** Un conjunto de categoría múltiple está formado por múltiples campos, todos ellos codificados de la misma forma, con frecuencia con varias categorías de respuesta posibles. Por ejemplo, una pregunta de la encuesta es, “Nombre hasta tres nacionalidades que describan su herencia étnica”. Puede haber hasta trescientas respuestas posibles, pero para la codificación, la lista se limita a las 40 nacionalidades más comunes, con el resto de categorías definidas como “otras”. En un archivo de datos, las tres opciones son tres campos, cada uno con 41 categorías (40 nacionalidades codificadas y una categoría “otras”).

## Tablero Pestaña Detallado

Figura 5-9  
Pestaña Detallado

Utilice la pestaña Detallado cuando sepa el tipo de visualización que desea crear o cuando quiera añadir aspectos estéticos, paneles y/o animaciones opcionales a una visualización. Para obtener ejemplos, consulte [Ejemplos de Tableros](#) el p. 267.

- ▶ Si había seleccionado un tipo de visualización en la pestaña Básico, se mostrará aquí. Si no es así, seleccione uno de la lista desplegable. Para obtener información sobre los tipos de visualización, consulte [Tipos de visualización de Graphboard preincorporado disponibles](#) el p. 259.
- ▶ Justo a la derecha de la imagen en miniatura de la visualización encontrará controles para especificar los campos (variables) necesarios para el tipo de visualización. Debe especificar todos estos campos.
- ▶ En el caso de ciertas visualizaciones, puede seleccionar un estadístico de resumen. En algunos casos (como los gráficos de barras), puede utilizar una de estas opciones de resumen para el aspecto estético de la transparencia. Para obtener descripciones de los estadísticos de resumen, consulte [Tablero Ficha Básico](#) el p. 252.
- ▶ Puede seleccionar uno o más de los aspectos estéticos opcionales. Estos aspectos pueden aumentar las dimensiones permitiéndole incluir otros campos en la visualización. Por ejemplo, puede utilizar un campo para modificar el tamaño de los puntos de un diagrama de dispersión. Para obtener más información sobre los aspectos estéticos opcionales, consulte [Aspecto estético, superposiciones, paneles y animación](#) el p. 244. Recuerde que el aspecto estético de la transparencia no se admite durante el procesamiento.
- ▶ Si está creando una visualización de mapa, el grupo Archivos de mapa muestra el archivo o archivos de mapas que se usarán. Si hay un archivo de mapa por defecto se muestra este archivo.

Para cambiar el archivo de mapa, pulse **Seleccionar archivo de mapa** para que se muestre el cuadro de diálogo **Seleccionar mapas**. También puede especificar el archivo de mapa por defecto en este cuadro de diálogo. Si desea obtener más información, consulte el tema [Seleccionar archivos de mapas para visualizaciones de mapas](#) el p. 258.

- Puede seleccionar una o más opciones de paneles o animación. Para obtener más información sobre las opciones de paneles y animación, consulte [Aspecto estético, superposiciones, paneles y animación](#) el p. 244.

### ***Seleccionar archivos de mapas para visualizaciones de mapas***

Si selecciona una plantilla de visualización de mapa, necesita un archivo de mapa que defina la información geográfica para dibujar el mapa. Si hay un archivo de mapa por defecto se usará para la visualización de mapas. Para elegir otro archivo de mapa, pulse **Seleccionar archivo de mapa** en la pestaña **Detallado** para que se muestre el cuadro de diálogo **Seleccionar mapas**.

El cuadro de diálogo **Seleccionar mapas** le permite elegir un archivo de mapa primario y uno de referencia. Los archivos de mapas definen la información geográfica para dibujar el mapa. Su aplicación se instala con un conjunto de archivos de mapas estándar. Si tiene otros archivos ESRI shapefile y quiere usarlos, primero tiene que convertirlos a archivos SMZ. Si desea obtener más información, consulte el tema [Conversión y distribución de mapas en formato shapefile](#) el p. 290. Tras convertir el mapa, pulse **Administrar...** en el cuadro de diálogo **Selector de plantillas** para importar el mapa en el Sistema de administración de modo que esté disponible en el cuadro de diálogo **Seleccionar mapas**.

A continuación se muestran algunos puntos que hay que considerar al especificar archivos de mapas:

- Todas las plantillas de mapas necesitan al menos un archivo de mapas.
- Normalmente el archivo de mapa enlaza un atributo de la leyenda con la clave de datos.
- Si la plantilla no necesita una leyenda que enlace con una clave de datos, necesita un archivo de mapa de referencia y campos que especifiquen las coordenadas (tales como longitud y latitud) para dibujar los elementos en el mapa de referencia.
- Las plantillas de mapas superpuestas necesitan dos mapas: un archivo de mapa primario y uno de referencia. El mapa de referencia se dibuja primero para que quede detrás del mapa primario.

Para obtener más información sobre términos de mapas, tales como atributos y características, consulte [Conceptos claves de los mapas](#) el p. 291.

**Archivo de mapa.** Puede seleccionar cualquier archivo de mapa que se encuentre en el Sistema de administración. Aquí se incluyen los archivos de mapa instalados previamente y los importados por usted. Si desea obtener más información sobre la administración de archivos de mapas, consulte [Administración de plantillas, hojas de estilo y archivos de mapas](#) el p. 289.

**Leyenda.** Especifique el atributo que quiere usar como clave para enlazar el archivo de mapa a la clave de datos.

**Guarde como predeterminados este archivo y configuración de mapa.** Seleccione esta casilla de verificación si quiere usar el archivo de mapa seleccionado como predeterminado. Si especifica un archivo de mapa como predeterminado no tendrá que especificar un archivo de mapa cada vez que cree una visualización de mapa.

**Clave de datos.** Este control muestra el mismo valor tal y como aparece en Selector de plantilla de la pestaña Detallado. Se incluye aquí por si necesita cambiar la clave si ha seleccionado un archivo de mapa específico.

**Mostrar todas las características del mapa en la visualización.** Cuando esta opción está activada, todas las características del mapa están representadas en la visualización, incluso si no coincide ningún valor de clave de datos. Si quiere ver tan sólo las características de las que hay datos, desactive esta opción. Las funciones identificadas en las leyendas de la lista Claves de mapa no coincidentes no se mostrarán en la visualización.

**Comparar valores de datos y mapa.** La leyenda y la clave de datos se enlazan respectivamente para crear la visualización del mapa. Los valores de la clave y la leyenda deben coincidir. De lo contrario no podrá crear la visualización del mapa. Pulse Comparar para comprobar que coinciden los valores de la leyenda y la clave de datos. El icono que se muestra le informa del estado de la comparación. Estos iconos se describen a continuación. Si se ha hecho una comparación y hay valores de la clave de datos que no coinciden con los de la leyenda, los valores de la clave de datos aparecen en la lista Claves de datos no coincidentes. En la lista Claves de mapa no coincidentes, también puede ver los valores de leyenda que no tienen valores de la clave de datos coincidentes. Si Mostrar todas las características del mapa en la visualización no está activada, las funciones identificadas por estos valores de leyenda no se mostrarán.

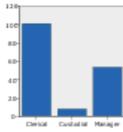
Tabla 5-1  
Iconos de comparación

Icono	Descripción
	No se ha realizado ninguna comparación. Éste es el estado por defecto antes de pulsar en Comparar. Debe tener cuidado ya que no sabe si los valores de la clave de datos y la leyenda coinciden.
	Se ha realizado una comparación y los valores de la clave de datos y la leyenda coinciden por completo. Para cada valor de la clave de datos hay una característica coincidente identificada por la leyenda.
	Se ha realizado una comparación y algunos valores de la clave de datos y de la leyenda no coinciden. Para algunos valores de la clave de datos, no hay ninguna característica coincidente identificada por la leyenda. Debe tener cuidado. Si continúa, la visualización del mapa no incluirá todos los valores.
	Se ha realizado una comparación y no hay valores de la clave de datos y de la leyenda que coincidan. Debe elegir otra clave de datos u otra leyenda, ya que si continúa no se representará ningún mapa.

## ***Tipos de visualización de Graphboard preincorporado disponibles***

Puede crear diferentes tipos de vistas. Los siguientes tipos preincorporados están disponibles en las pestañas Básico y Detallado. Algunas de las descripciones para las plantillas (especialmente las plantillas de mapas) identifican los campos (variables) especificados en la pestaña Detallado utilizando texto especial.

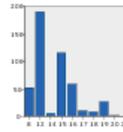
**Tabla 5-2**  
Tipos de gráficos disponibles



### Barras

Calcula un estadístico de resumen de un campo numérico continuo y muestra los resultados en forma de barras de cada categoría de un campo categórico.

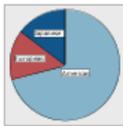
*Requiere:* Un campo categórico y un campo continuo.



### Gráfico de recuentos

Muestra la proporción de filas/casos en cada categoría de un campo categórico como barras. También puede utilizar el nodo de gráfico Distribución para crear este gráfico. Este nodo ofrece algunas opciones adicionales. Si desea obtener más información, consulte el tema [Nodo Distribución](#) el p. 308.

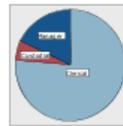
*Requiere:* Un único campo categórico.



### Sectores

Calcula la suma de un campo numérico continuo y muestra la proporción de esa suma distribuida en cada categoría de un campo categórico como sectores de un gráfico de sectores.

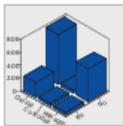
*Requiere:* Un campo categórico y un campo continuo.



### Sectores de recuentos

Muestra la proporción de filas o casos en cada categoría de un campo categórico como sectores de un gráfico de sectores.

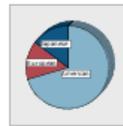
*Requiere:* Un único campo categórico.



### Barras 3-D

Calcula un estadístico de resumen de un campo numérico continuo y muestra los resultados del cruce de categorías entre dos campos categóricos.

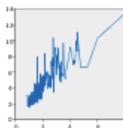
*Requiere:* Un par de campos categóricos y un campo continuo.



### 3-D

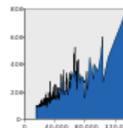
Es igual que el gráfico Sectores excepto por el efecto en 3D adicional.

*Requiere:* Un campo categórico y un campo continuo.



### Líneas

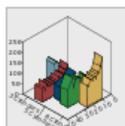
Calcula un estadístico de resumen de un campo para cada valor de otro campo y dibuja una línea que conecta dichos valores. También puede utilizar el nodo Gráfico para crear un gráfico de línea. Este nodo ofrece algunas opciones adicionales. Si desea obtener más información, consulte el tema [Nodo Gráfico](#) el p. 299.



### Áreas

Calcula un estadístico de resumen de un campo para cada valor de otro campo y dibuja un área que conecta dichos valores. La diferencia entre un gráfico de línea y uno de áreas es mínima, ya que el de áreas presenta una línea con el espacio que aparece debajo coloreado. Sin embargo, si utiliza un aspecto estético de color, el resultado será una

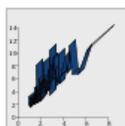
*Requiere:* Un par de campos de cualquier tipo.



### Área 3D

Muestra los valores de un campo representado en contraposición a los valores de otro y divididos según un campo categórico. Se obtiene un elemento de área para cada categoría.

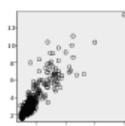
*Requiere:* Un campo categórico y un par de campos de cualquier tipo.



### Bandas

Calcula un estadístico de resumen de un campo para cada valor de otro campo y dibuja una banda que conecta dichos valores. Una barra acumulada es básicamente una línea con efectos en 3D. Realmente no es un gráfico en 3D.

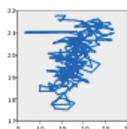
*Requiere:* Un par de campos de cualquier tipo.



### Diagrama de dispersión

Muestra los valores de un campo representados en contraposición a los valores de otro. Este gráfico puede destacar la relación entre los campos (si existe). También puede utilizar el nodo Gráfico para crear un diagrama de dispersión. Este nodo ofrece algunas opciones adicionales. Si desea obtener más información, consulte el tema [Nodo Gráfico](#) el p. 299.

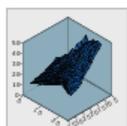
*Requiere:* Un par de campos de cualquier tipo.



### Ruta

Muestra los valores de un campo en contraposición a los valores de otro, con una línea que conecta dichos valores en el orden en el que aparecen en el conjunto de datos original. El orden es la principal diferencia entre una ruta y una línea.

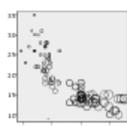
*Requiere:* Un par de campos de cualquier tipo.



### Superficie

Muestra los valores de tres campos representados en contraposición a los valores de otro, con una superficie que conecta dichos valores.

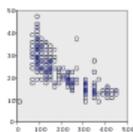
*Requiere:* Tres campos de cualquier tipo.



### Gráfico de burbujas

Igual que el diagrama de dispersión básico, muestra los valores de un campo representado en contraposición a los valores de otro. La diferencia es que los valores de un tercer campo se utilizan para modificar el tamaño de los puntos individuales.

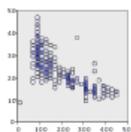
*Requiere:* Tres campos de cualquier tipo.



### Diagrama de dispersión en intervalos

Igual que el diagrama de dispersión básico, muestra los valores de un campo representado en contraposición a los valores de otro. La diferencia es que los valores similares se agrupan en intervalos y que se utiliza el color o tamaño para indicar el número de casos de cada intervalo.

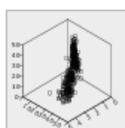
*Requiere:* Un par de campos continuos.



### Diagrama de dispersión en intervalos hexagonales

Consulte la descripción de Diagrama de dispersión en intervalos. La diferencia es la forma de los intervalos subyacentes, que son hexágonos en lugar de círculos. El aspecto del diagrama de dispersión en intervalos hexagonales es parecido al del diagrama de dispersión en intervalos. Sin embargo, el número de valores de cada intervalo variará dependiendo del gráfico debido a la forma de los intervalos subyacentes.

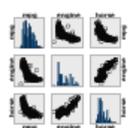
*Requiere:* Un par de campos continuos.



### Diagrama de dispersión 3-D

Muestra los valores de tres campos representados en contraposición a los valores de cada uno de ellos. Este gráfico puede destacar la relación entre los campos (si existe). También puede utilizar el nodo Gráfico para crear un diagrama de dispersión 3-D. Este nodo ofrece algunas opciones adicionales. Si desea obtener más información, consulte el tema [Nodo Gráfico](#) el p. 299.

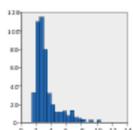
*Requiere:* Tres campos de cualquier tipo.



### Matriz de diagramas de dispersión (SPLOM)

Muestra los valores de un campo representado en contraposición a los valores de otro para cada campo. Una SPLOM es igual que una tabla de diagramas de dispersión. La SPLOM también incluye un histograma de cada campo.

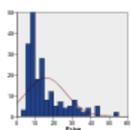
*Requiere:* Dos o más campos continuos.



### Histograma

Muestra la distribución de la frecuencia de un campo. Un histograma puede ayudarle a determinar el tipo de distribución y comprobar si ésta es asimétrica. También puede utilizar el nodo de gráfico Histograma para crear este gráfico. Este nodo ofrece algunas opciones adicionales. Si desea obtener más información, consulte el tema [Pestaña Gráfico de Histograma](#) el p. 314.

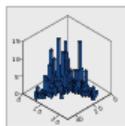
*Requiere:* Un único campo o de cualquier tipo.



### Histograma con distribución normal

Muestra la distribución de la frecuencia de un campo continuo con una curva superpuesta de la distribución normal.

*Requiere:* Un único campo continuo.



### Histograma 3-D

Muestra la distribución de la frecuencia de un par de campos continuos.

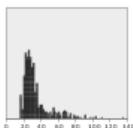
*Requiere:* Un par de campos continuos.



### Densidad 3-D

Muestra la distribución de la frecuencia de un par de campos continuos. Es similar a un histograma en 3D, con la única diferencia de que se utiliza una superficie en lugar de barras para mostrar la distribución.

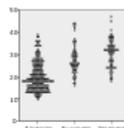
*Requiere:* Un par de campos continuos.



### Diagrama de puntos

Muestra los casos o filas individuales y los apila en los puntos de datos diferentes del eje  $x$ . Este gráfico es parecido a un histograma en cuanto a que muestra la distribución de los datos; sin embargo, éste presenta cada caso o fila en lugar de un recuento total de un intervalo específico (rango de valores).

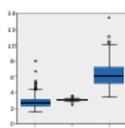
*Requiere:* Un único campo de cualquier tipo.



### Gráfico de puntos 2-D

Muestra los casos o filas individuales y los apila en los distintos puntos de datos del eje  $y$  y para cada categoría de un campo categórico.

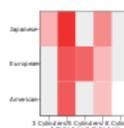
*Requiere:* Un campo categórico y un campo continuo.



### Diagramas de caja

Calcula los cinco estadísticos (mínimo, primer cuartil, mediana, tercer cuartil y máximo) de un campo continuo para cada categoría de un campo categórico. Los resultados se muestran como elementos de un gráfico de caja o esquema. Los gráficos de caja pueden ayudarle a ver cómo varía la distribución de los datos continuos entre las categorías.

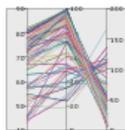
*Requiere:* Un campo categórico y un campo continuo.



### Mapa de calor

Calcula la media de un campo continuo del cruce de categorías entre dos campos categóricos.

*Requiere:* Un par de campos categóricos y un campo continuo.



### Paralelo

Crea ejes paralelos de cada campo y dibuja una línea a través del valor de campo para cada fila o caso de los datos.

*Requiere:* Dos o más campos continuos.



### Coropleta de recuentos

Calcula el recuento de cada categoría de un campo categórico (Clave de datos) y dibuja un mapa que utiliza la saturación de color para representar los recuentos del mapa correspondientes a las categorías.

*Requiere:* Un campo categórico. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).



### Coropleta de medias/medianas/sumas

Calcula la media, la mediana o la suma de un campo continuo (Color) para cada categoría de un campo categórico (Clave de datos) y dibuja un mapa que utiliza la saturación de color para representar los estadísticos calculados en el mapa correspondientes a las categorías.

*Requiere:* Un campo categórico y un campo continuo. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).



### Coropleta de valores

Dibuja un mapa que utiliza colores para representar los valores de un campo categórico (Color) del mapa correspondientes a los valores definidos por otro campo categórico (Clave de datos). Si existen varios valores categóricos del campo Color para cada característica, se utiliza el valor modal.

*Requiere:* Un par de campos categóricos. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).



### Coordenadas de una coropleta de recuentos

Es similar a la coropleta de recuentos con la diferencia de que hay dos campos continuos adicionales (Longitud y Latitud) que identifican las coordenadas para dibujar los puntos del mapa de coropletas.

*Requiere:* Un campo categórico y un par de campos continuos. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).



### Coordenadas de una coropleta de medias/medianas/sumas

Es similar a la coropleta de medias/medianas/sumas con la diferencia de que hay dos campos continuos adicionales (Longitud y Latitud) que identifican las coordenadas para dibujar los puntos del mapa de coropletas.

*Requiere:* Un campo categórico y tres campos continuos. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).



### Coordenadas de una coropleta de valores

Es similar a la coropleta de valores con la diferencia de que hay dos campos continuos adicionales (Longitud y Latitud) que identifican las coordenadas para dibujar los puntos del mapa de coropletas.

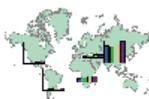
*Requiere:* Un par de campos categóricos y un par de campos continuos. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).



### Barras de recuentos de un mapa

Calcula la proporción de filas o casos en cada categoría de un campo categórico (Categorías) para cada característica del mapa (Clave de datos) y dibuja un mapa y los gráficos de barras en el centro de cada mapa.

*Requiere:* Un par de campos categóricos. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).



### Barras de un mapa

Calcula un estadístico de resumen de un campo continuo (Valores) y muestra los resultados de cada categoría de un campo categórico (Categorías) para cada característica del mapa (Clave de datos); los gráficos de barras se posicionan en el centro de cada mapa.

*Requiere:* Un par de campos categóricos y un campo continuo. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).



### Sectores de recuentos de un mapa

Muestra la proporción de filas o casos en cada categoría de un campo categórico (Categorías) para cada característica del mapa (Clave de datos) y dibuja un mapa y las proporciones como sectores de un gráfico de sectores en el centro de cada mapa.

*Requiere:* Un par de campos categóricos. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).



### Sectores de un mapa

Calcula la suma de un campo continuo (Valores) en cada categoría de un campo categórico (Categorías) para cada característica del mapa (Clave de datos) y dibuja un mapa y las sumas como sectores de un gráfico de sectores en el centro de cada mapa.

*Requiere:* Un par de campos categóricos y un campo continuo. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).



### Gráfico de líneas de un mapa

Calcula un estadístico de resumen de un campo continuo (Y) para cada valor de otro campo (X) para cada característica del mapa (Clave de datos) y dibuja un mapa y los gráficos de líneas que conectan los valores en el centro de cada mapa.

*Requiere:* Un campo categórico y un par de campos de cualquier tipo. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).



#### Coordenadas de un mapa de referencia

Dibuja un mapa y los puntos utilizando campos continuos (Longitud y Latitud) que identifican las coordenadas para los puntos.

*Requiere:* Un par de campos de rango. Un archivo de mapas.



#### Flechas de un mapa de referencia

Dibuja un mapa y las flechas utilizando campos continuos que identifican los puntos de inicio (Longitud de inicio y Latitud de inicio) y los puntos finales (Longitud final y Latitud final) de cada flecha. Cada registro o caso en los resultados de los datos en una flecha del mapa.

*Requiere:* Cuatro campos continuos. Un archivo de mapas.



#### Mapa de superposición de puntos

Dibuja un mapa de referencia y superpone sobre él otro mapa de puntos con puntos coloreadas por un campo categórico (Color).

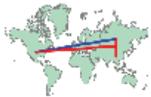
*Requiere:* Un par de campos categóricos. Un archivo de mapas de puntos cuya clave coincide con las categorías de la clave de datos. Un archivo de mapas de referencia.



#### Mapa de superposición de polígonos

Dibuja un mapa de referencia y superpone sobre él otro mapa de polígonos con polígonos coloreados por un campo categórico (Color).

*Requiere:* Un par de campos categóricos. Un archivo de mapas de polígonos cuya clave coincide con las categorías de la clave de datos. Un archivo de mapas de referencia.



#### Mapa de superposición de líneas

Dibuja un mapa de referencia y superpone sobre él otro mapa de líneas con líneas coloreadas por un campo categórico (Color).

*Requiere:* Un par de campos categóricos. Un archivo de mapas de líneas cuya clave coincide con las categorías de la clave de datos. Un archivo de mapas de referencia.

## Creación de visualizaciones de mapas

Para algunas visualizaciones, sólo debe realizar dos elecciones: campos (variables) de interés y una plantilla para visualizar esos campos. No se requieren elecciones o acciones adicionales. Las visualizaciones de mapas requieren al menos un paso adicional: Seleccione un archivo de mapas que defina la información geográfica para la visualización de mapas.

Los pasos básicos para la creación de un mapa simple son los siguientes:

- ▶ Seleccione los campos de interés en la pestaña Básico. Para obtener información acerca del tipo y número de campos necesarios para varias visualizaciones de mapas, consulte [Tipos de visualización de Graphboard preincorporado disponibles](#) el p. 259.
- ▶ Seleccione una plantilla del mapa.
- ▶ Pulse en la pestaña Detallado.
- ▶ Compruebe que tanto la Clave de datos como otras listas desplegadas obligatorias se ha establecido en los campos correctos.
- ▶ En el grupo Archivos de mapas, pulse Seleccionar un archivo de mapas.
- ▶ Seleccione el cuadro de diálogo Seleccionar mapas para seleccionar la leyenda así como el archivo de mapas. Los valores de la leyenda deben coincidir con los valores del campo que especifican las claves de datos. Puede utilizar el botón Comparar para comparar esos valores. Si selecciona una superposición de plantillas de mapas, deberá seleccionar un mapa de referencia. El mapa de referencia no tiene clave de los datos. Se utiliza como fondo para el mapa principal. Si desea obtener más información acerca del cuadro de diálogo Seleccionar mapas, consulte [Seleccionar archivos de mapas para visualizaciones de mapas](#) el p. 258.
- ▶ Pulse Aceptar para cerrar el cuadro de diálogo Seleccionar mapas.
- ▶ En el Selector de plantillas de tablero, pulse Ejecutar para crear la visualización del mapa.

## Ejemplos de Tableros

Esta sección incluye varios ejemplos diferentes para mostrar las opciones disponibles. Los ejemplos también ofrecen información para interpretar las visualizaciones resultantes.

Estos ejemplos usan la ruta denominada *graphboard.str*, que hace referencia a los archivos de datos, denominados *employee\_data.sav*, *customer\_subset.sav* y *worldsales.sav*. Estos archivos están disponibles en la carpeta *Demos* de cualquier IBM® SPSS® Modeler instalación Client. Puede acceder desde el grupo de programas SPSS Modeler en el menú Inicio de Windows. El archivo *graphboard.str* se encuentra en la carpeta *streams*.

Se recomienda que lea los ejemplos en el orden presentado. Los siguientes ejemplos se basan en los anteriores.

### **Ejemplo: Gráfico de barras con un estadístico de resumen**

Crearemos un gráfico de barras que resume un campo o variable numérico continuo para cada categoría de una variable de conjunto o categórica. De forma específica, crearemos un gráfico de barras que muestre el salario medio de hombres y mujeres.

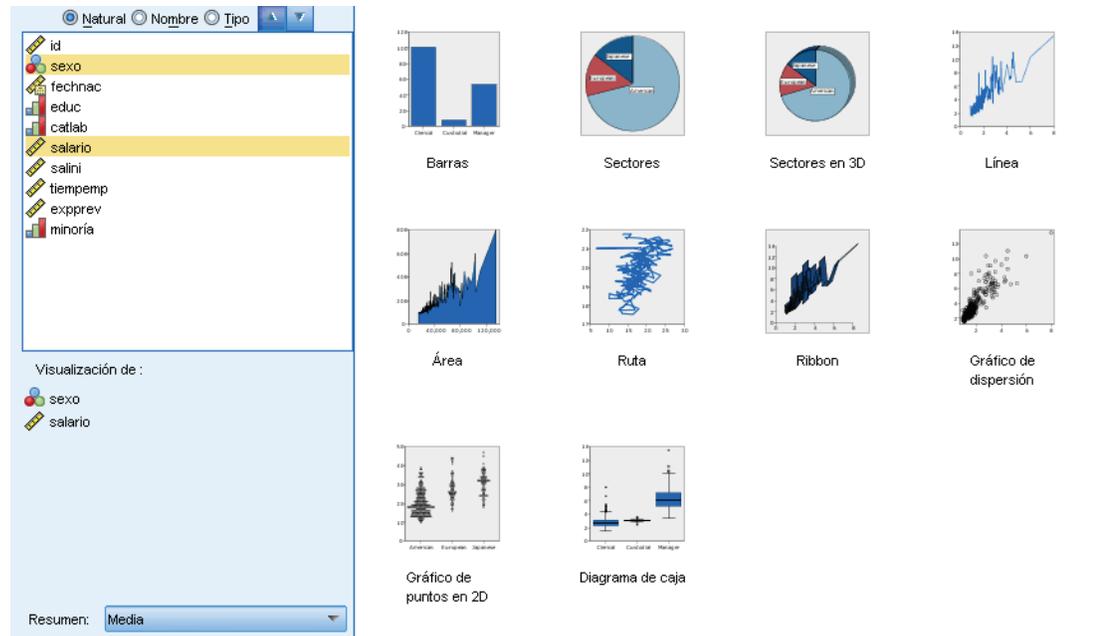
Éste y varios de los ejemplos siguientes utilizan *Employee data*, que es un conjunto de datos hipotético con información sobre los trabajadores de una empresa.

- ▶ Añada un nodo de origen Archivo de Statistics que haga referencia a *employee\_data.sav*.
- ▶ Añada un nodo Tablero y ábralo para editarlo.

- ▶ En la pestaña Básico, seleccione *Sexo* y *Salario actual*. (Mantenga pulsada la tecla Ctrl mientras pulsa para seleccionar varios campos o variables.)
- ▶ Seleccione Barras.
- ▶ En la lista desplegable Resumen, seleccione Media.

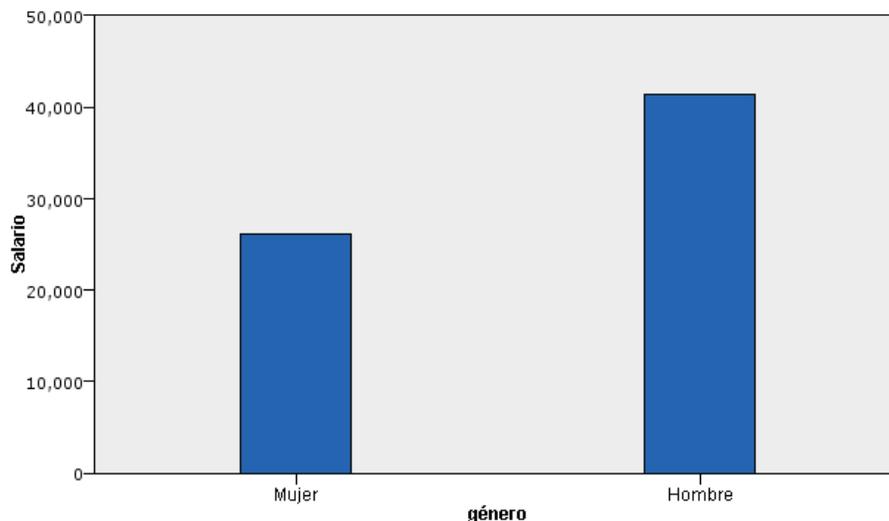
Figura 5-11

Opciones de la pestaña Básico, gráfico de barras con un estadístico de resumen



- ▶ Pulse en Ejecutar.
- ▶ En la pantalla resultante, pulse el botón de la barra de herramientas “Mostrar campo y etiquetas de valores” (el segundo del grupo de dos en el centro de la barra de herramientas).

Figura 5-12  
Gráfico de barras con un estadístico de resumen



Podemos observar lo siguiente:

- Basándonos en la altura de las barras, está claro que el salario medio de los hombres es mayor que el salario medio de las mujeres.

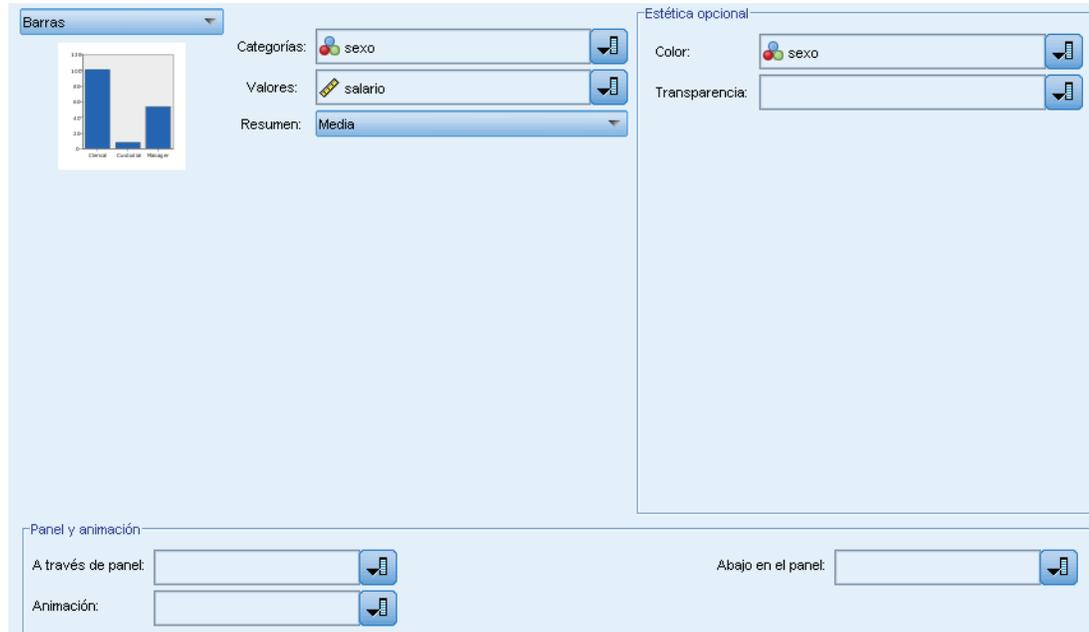
### **Ejemplo: Gráfico de barras conglomerado con un estadístico de resumen**

Ahora crearemos un gráfico de barras conglomerado para ver si la diferencia en el salario medio entre hombres y mujeres está relacionada con el tipo de trabajo. Tal vez las mujeres, en promedio, se dedican a ciertos trabajos más que los hombres.

*Nota:* Este ejemplo usa *Employee data*.

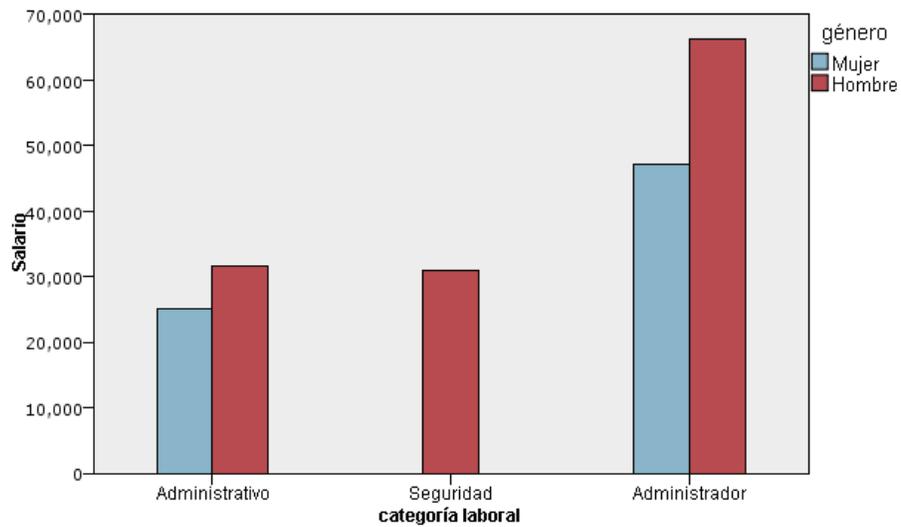
- ▶ Añada un nodo Tablero y ábralo para editarlo.
- ▶ En la pestaña Básico, seleccione *Categoría laboral* y *Salario actual*. (Mantenga pulsada la tecla Ctrl mientras pulsa para seleccionar varios campos o variables.)
- ▶ Seleccione Barras.
- ▶ En la lista Resumen, seleccione Media.
- ▶ Pulse en la pestaña Detallado. Recuerde que sus selecciones en la pestaña anterior se reflejarán aquí.
- ▶ En el grupo Aspectos estéticos opcionales, seleccione *sexo* en la lista desplegable Color.

**Figura 5-13**  
Opciones de la pestaña Detallado, gráfico de barras conglomerado



- Pulse en Ejecutar.

**Figura 5-14**  
Gráfico de barras agrupadas



Podemos observar lo siguiente:

- La diferencia en los salarios medios de cada tipo de trabajo no parece ser tan grande como lo era en el gráfico de barras que comparaba los salarios medios de todos los hombres y mujeres. Quizás haya un número variable de hombres y mujeres en cada grupo. Puede comprobarlo creando un gráfico de barras de recuentos.
- Independientemente del tipo de trabajo, el salario medio de los hombres siempre es mayor que el de las mujeres.

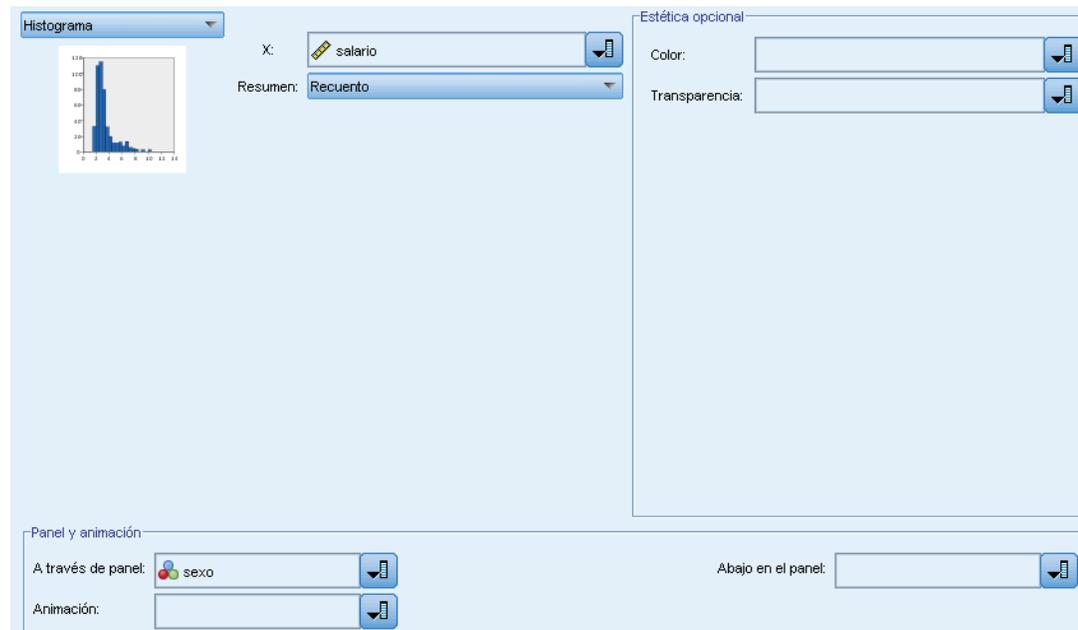
### **Ejemplo: Histograma panelado**

Crearemos un histograma panelado por sexos para poder comparar las distribuciones de frecuencia del salario de hombres y mujeres. La distribución de frecuencia muestra cuántos casos o filas se encuentran dentro de rangos de salario específicos. El histograma panelado puede ayudarnos a analizar aun más la diferencia de salarios entre los dos sexos.

*Nota:* Este ejemplo usa *Employee data*.

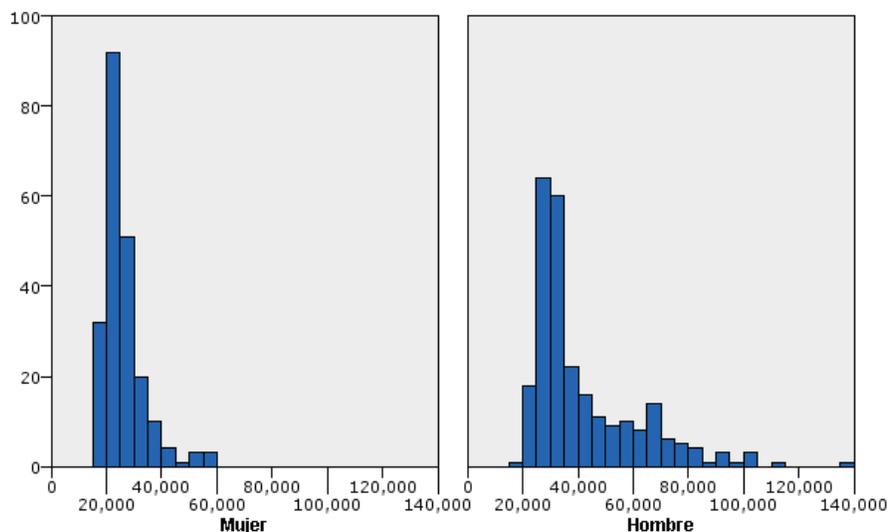
- ▶ Añada un nodo Tablero y ábralo para editarlo.
- ▶ En la pestaña Básico, seleccione *Salario actual*.
- ▶ Seleccione Histograma.
- ▶ Pulse en la pestaña Detallado.
- ▶ En el grupo Paneles y animación, seleccione *sexo* en la lista desplegable A través del panel.

**Figura 5-15**  
Opciones de la pestaña Detallado, histograma panelado



- ▶ Pulse en Ejecutar.

Figura 5-16  
Histograma panelado



Podemos observar lo siguiente:

- Ninguna de las distribuciones de frecuencia es normal. Es decir, los histogramas no parecen curvas de campana, como debería ser si los datos estuvieran distribuidos con normalidad.
- Las barras más altas están en el lado izquierdo de cada gráfico. Por lo tanto, tanto en hombres como en mujeres, hay un mayor número de personas con salarios más bajos.
- Las distribuciones de frecuencia de salarios entre hombres y mujeres no son iguales. Observe la forma de los histogramas. Hay más hombres que mujeres con salarios altos.

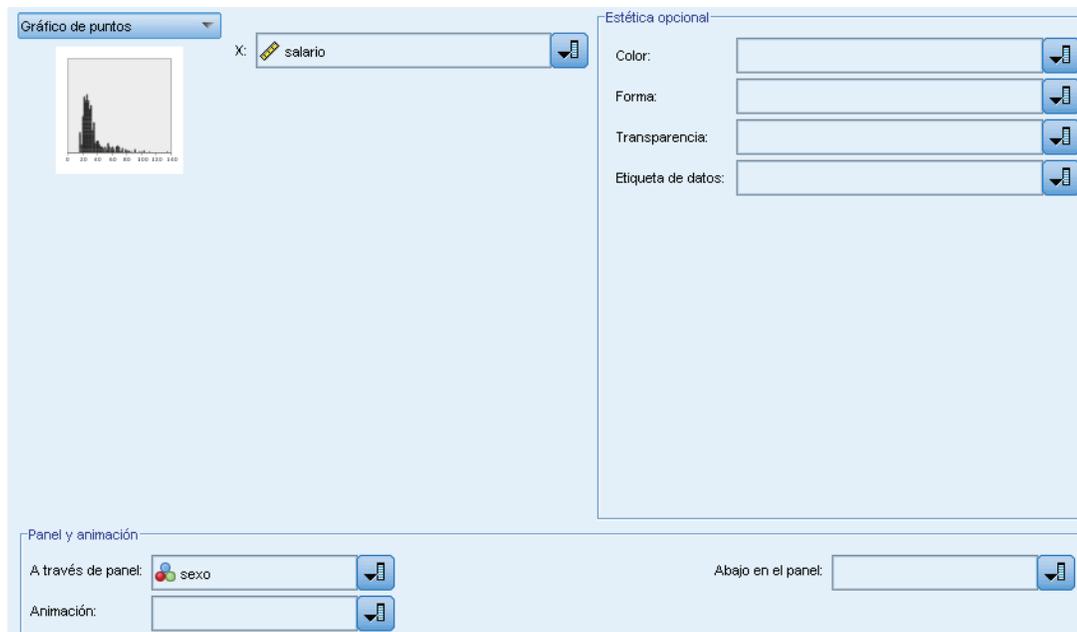
### **Ejemplo: Gráfico de puntos panelado**

Como un histograma, un gráfico de puntos muestra la distribución de un rango numérico continuo. A diferencia de un histograma, que muestra los recuentos de rangos de datos en intervalos, un gráfico de puntos muestra todos los casos o filas de los datos. Por lo tanto, un gráfico de puntos ofrece una mayor cantidad de detalles en comparación con el histograma. De hecho, el uso de un gráfico de puntos puede ser el punto de inicio preferido cuando se analizan distribuciones de frecuencia.

*Nota:* Este ejemplo usa *Employee data*.

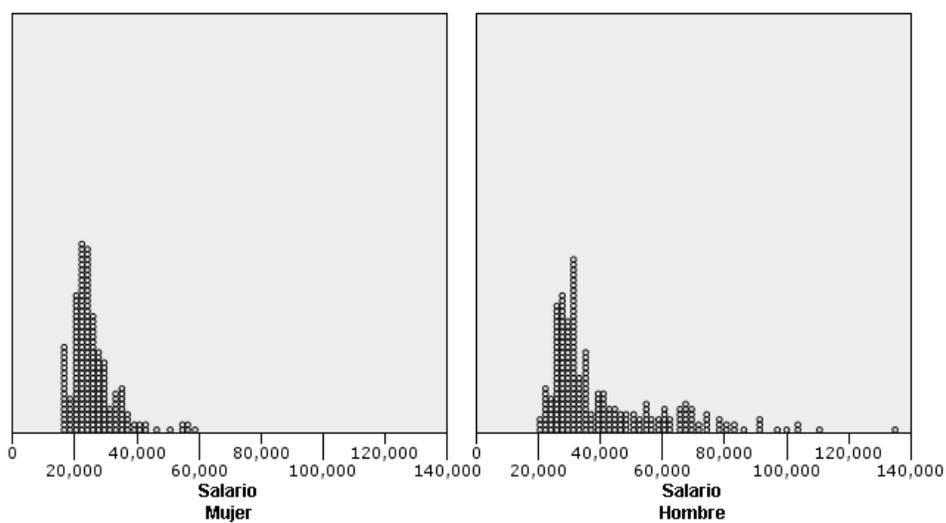
- ▶ Añada un nodo Tablero y ábralo para editarlo.
- ▶ En la pestaña Básico, seleccione *Salario actual*.
- ▶ Seleccione Gráfico de puntos.
- ▶ Pulse en la pestaña Detallado.
- ▶ En el grupo Paneles y animación, seleccione *sexo* en la lista desplegable A través del panel.

Figura 5-17  
Opciones de la pestaña Detallado, gráfico de puntos panelado



- ▶ Pulse en Ejecutar.
- ▶ Maximice la ventana de resultados para ver el gráfico con mayor claridad.

Figura 5-18  
Gráfico de puntos panelado



En comparación con el histograma (consulte [Ejemplo: Histograma panelado](#) el p. 271), podemos observar lo siguiente:

- El pico de los 20.000 que apareció en el histograma de mujeres es menos drástico en el gráfico de puntos. Hay muchos casos o filas concentrados alrededor de dicho valor, pero la mayoría de los valores están más cerca de los 25.000. Este nivel de detalle no es aparente en el histograma.
- Aunque el histograma de los hombres sugiere que el salario medio de los hombres desciende gradualmente tras los 40.000, el gráfico de puntos muestra que la distribución es bastante uniforme tras este valor, hasta los 80.000. En cualquier valor de salario dentro de ese rango, hay tres o más hombres que obtienen dicho salario en particular.

### **Ejemplo: Diagramas de caja**

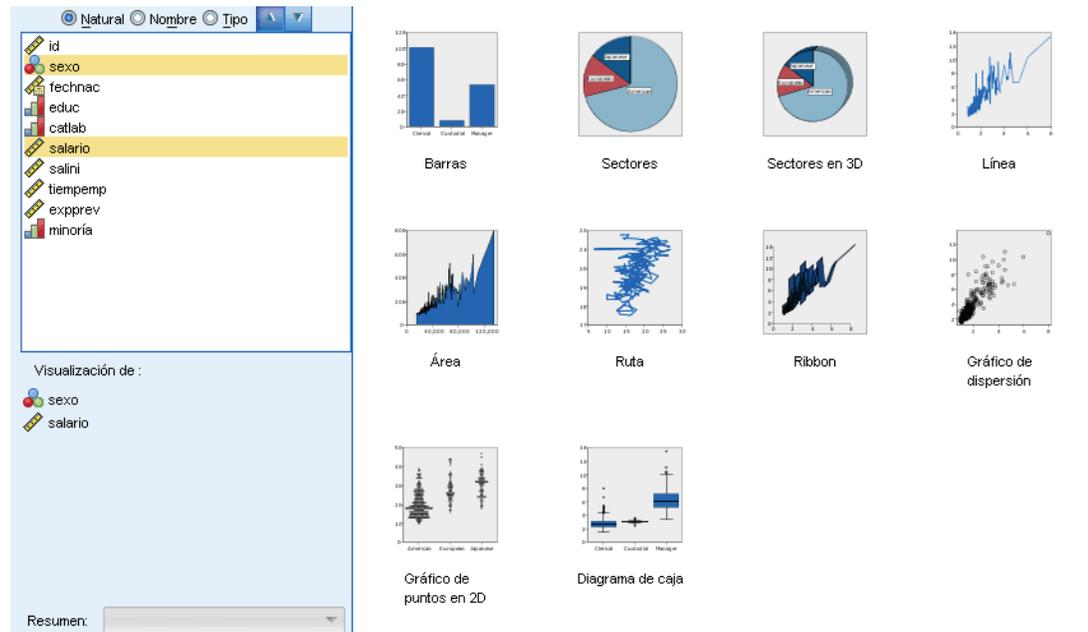
Un diagrama de cajas es otra visualización que resultar útil para saber cómo se distribuyen los datos. Un diagrama de cajas contiene varias medidas estadísticas que estudiaremos tras crear la visualización.

*Nota:* Este ejemplo usa *Employee data*.

- ▶ Añada un nodo Tablero y ábralo para editarlo.
- ▶ En la pestaña Básico, seleccione *Sexo* y *Salario actual*. (Mantenga pulsada la tecla Ctrl mientras pulsa para seleccionar varios campos o variables.)
- ▶ Seleccione Gráfico de caja.

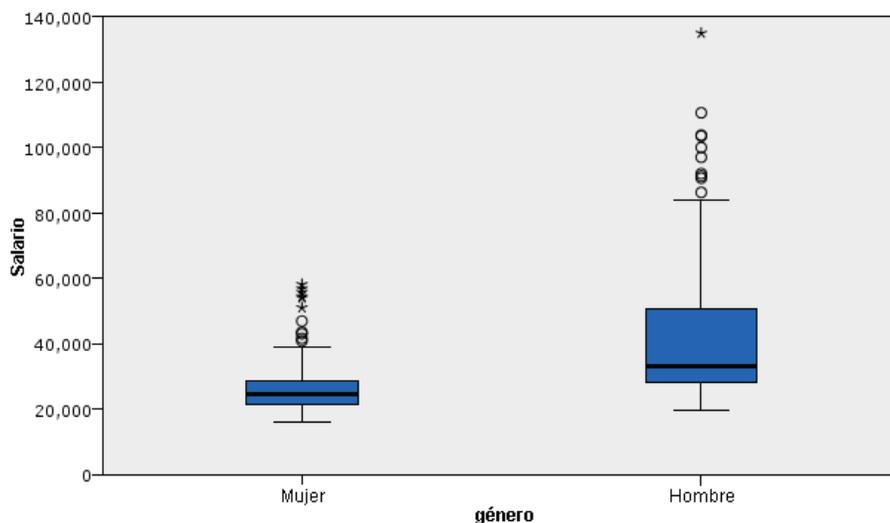
Figura 5-19

Opciones de la pestaña Básico, gráfico de caja



- ▶ Pulse en Ejecutar.

Figura 5-20  
Diagramas de caja



Veamos las distintas partes del gráfico de caja:

- La línea oscura que se encuentra en la mitad de las cajas es la mediana del *salario*. La mitad de los casos o filas tiene un valor superior a la mediana y la otra mitad tiene un valor inferior. Igual que la media, la mediana es una medida de tendencia central. A diferencia de la media, se ve menos influida por los casos o filas con valores extremos. En este ejemplo, la mediana es inferior que la media (compárelo con [Ejemplo: Gráfico de barras con un estadístico de resumen](#) el p. 267). La diferencia entre la media y la mediana indica que hay menos casos o filas con valores extremos que elevan la media. Es decir, hay un par de trabajadores con grandes salarios.
- La parte inferior de la caja indica el percentil 25. El veinticinco por ciento de los casos o filas tienen valores por debajo del percentil 25. La parte superior de la caja representa el percentil 75. El veinticinco por ciento de los casos o filas tienen valores por encima del percentil 75. Esto significa que el 50 % de los casos o filas se encuentran dentro de la caja. La caja es mucho más pequeña para mujeres que para hombres. Es una pista que indica que el *salario* varía menos entre las mujeres que entre los hombres. Las partes superior e inferior de la caja suelen denominarse **bisagras**.
- Las barras en forma de T que salen de las cajas se denominan **cercas internas** o **patillas** o **bigotes**. Tienen una extensión de 1,5 veces la altura de la caja o, si no hay ningún caso o fila con valor en dicho rango, hasta los valores mínimo y máximo. Si los datos se distribuyen con normalidad, se espera que aproximadamente el 95 % de los datos se encuentre entre las cercas internas. En este ejemplo, las cercas internas tienen una extensión menor para las mujeres que para los hombres, otro indicativo de que el *salario* varía menos entre las mujeres que entre los hombres.
- Los puntos son **valores atípicos**. Se definen como valores que no están dentro de las cercas internas. Los valores atípicos son valores extremos. Los asteriscos o estrellas son **valores atípicos extremos**. Representan casos o filas con valores superiores a la altura de las cajas multiplicada por tres. Existen varios valores atípicos tanto para mujeres como para hombres.

Recuerde que la media es superior que la mediana. La media de mayor tamaño es el resultado de estos valores atípicos.

### **Ejemplo: Gráfico de sectores**

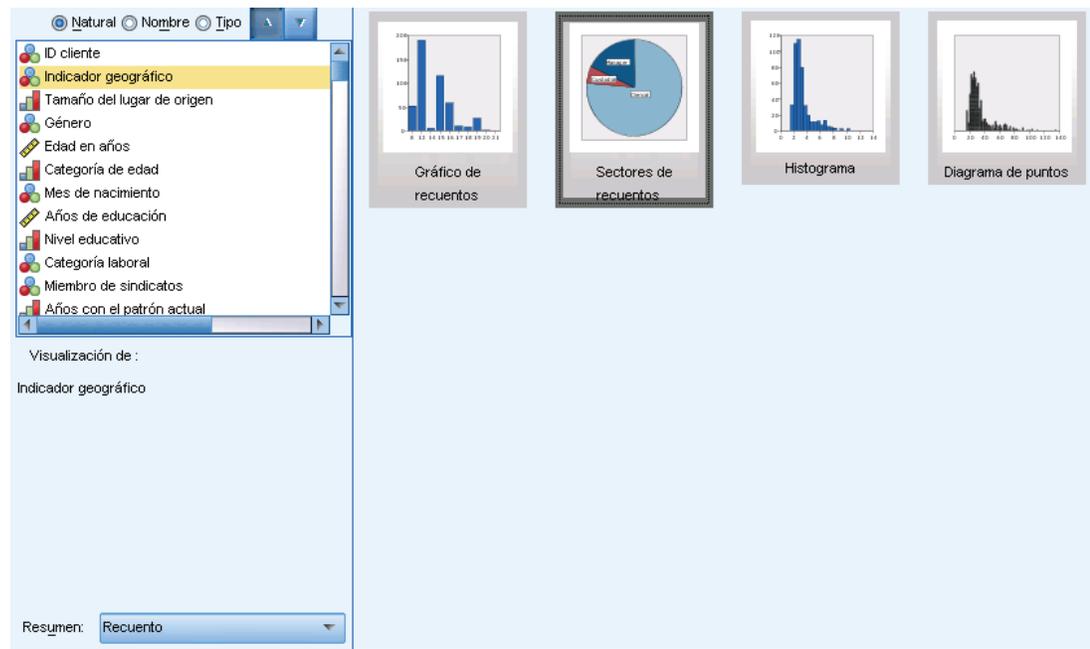
Ahora utilizaremos un conjunto de datos diferente para explorar otros tipos de visualización. El conjunto de datos es *customer\_subset*, que es un archivo de datos hipotéticos con información sobre clientes.

Primero crearemos un gráfico de sectores para comprobar la proporción de clientes en diferentes regiones geográficas.

- ▶ Añada un nodo de origen Archivo de Statistics que haga referencia a *customer\_subset.sav*.
- ▶ Añada un nodo Tablero y ábralo para editarlo.
- ▶ En la pestaña Básico, seleccione *Indicador geográfico*.
- ▶ Seleccione Sectores de recuentos.

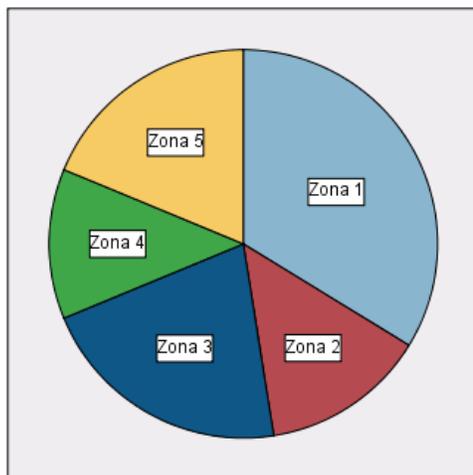
Figura 5-21

Opciones de la pestaña Básico, gráfico de sectores



- ▶ Pulse en Ejecutar.

Figura 5-22  
Gráfico de sectores



Podemos observar lo siguiente:

- La zona 1 tiene más clientes que el resto de zonas.
- Los clientes se distribuyen equitativamente entre el resto de zonas.

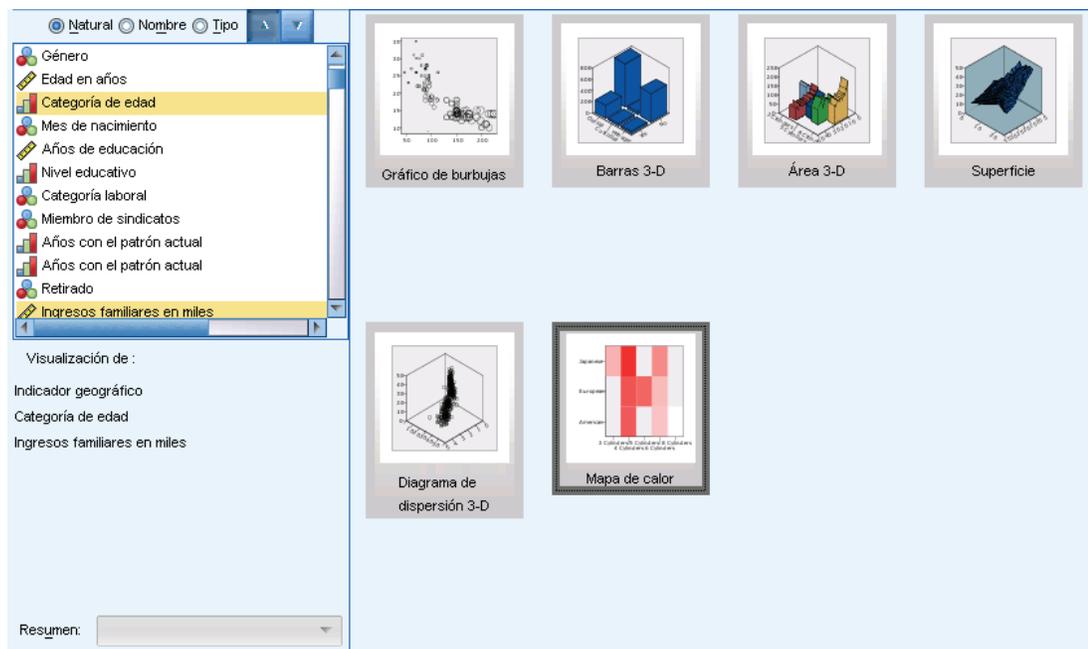
### **Ejemplo: Mapa de calor**

Ahora crearemos un mapa de calor categórico para comprobar la media de ingresos de los clientes en diferentes regiones geográficas y grupos de edad.

*Nota:* Este ejemplo utiliza `customer_subset`.

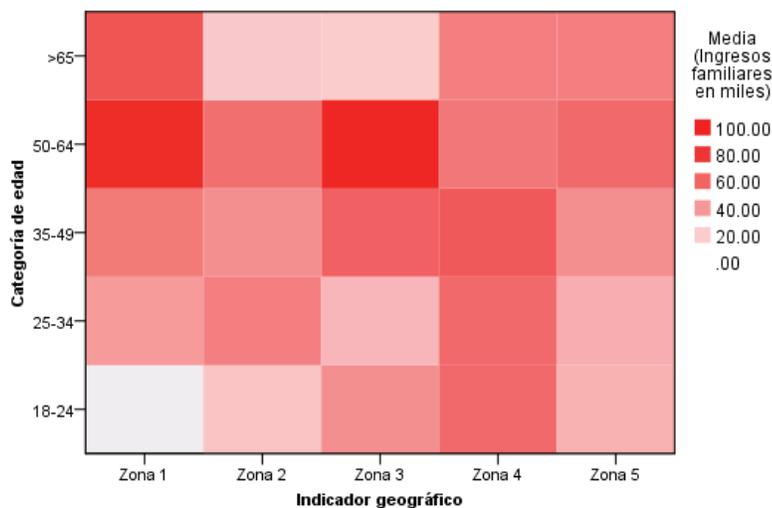
- ▶ Añada un nodo Tablero y ábralo para editarlo.
- ▶ En la pestaña Básico, seleccione *Indicador geográfico*, *Categoría de edad* e *Ingresos del hogar en miles*, en ese orden. (Mantenga pulsada la tecla Ctrl mientras pulsa para seleccionar varios campos o variables.)
- ▶ Seleccione Mapa de calor.

Figura 5-23  
Opciones de la pestaña Básico, mapa de calor



- ▶ Pulse en Ejecutar.
- ▶ En la ventana de resultados, pulse el botón de la barra de herramientas “Mostrar campo y etiquetas de valores” (el botón de la derecha de los dos en el centro de la barra de herramientas).

Figura 5-24  
Mapa de calor categórico



Podemos observar lo siguiente:

- Un mapa de calor es como una tabla que utiliza colores en lugar de números para representar los valores de las casillas. El color rojo intenso y brillante indica el valor más alto, mientras que el gris indica un valor bajo. El valor de cada casilla es la media del campo o variable continuos de cada par de categorías.
- Excepto en las zonas 2 y 5, el grupo de clientes cuya edad está entre 50 y 64 tiene una media de ingresos por hogar superior a la de otros grupos.
- No hay clientes con una edad entre 25 y 34 en la zona 4.

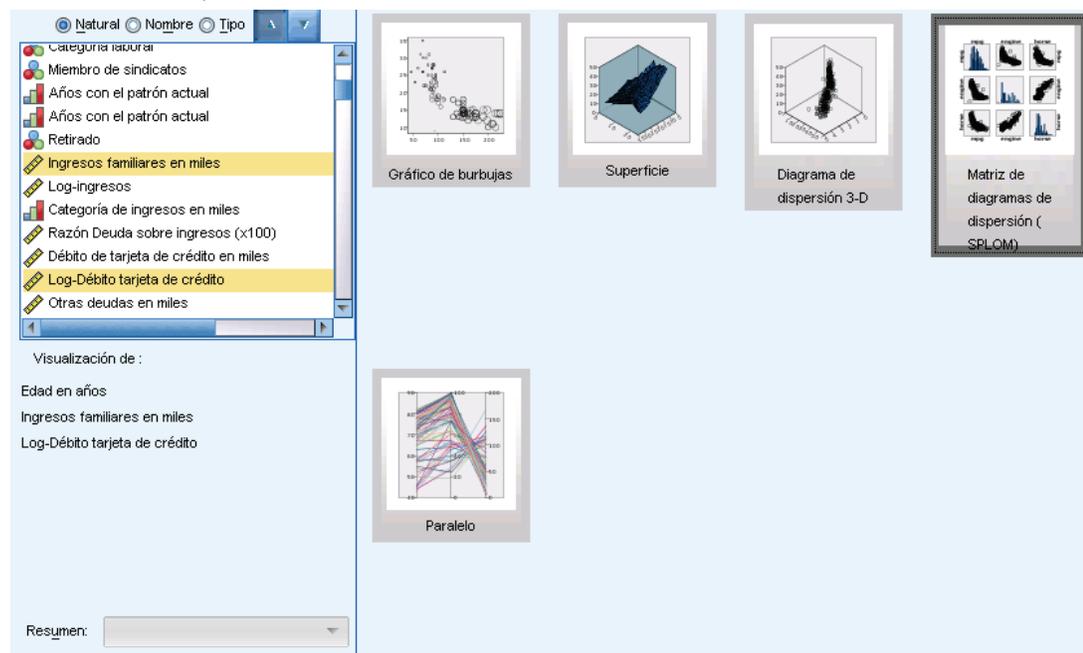
### **Ejemplo: Matriz de diagramas de dispersión (SPLOM)**

Crearemos una matriz de diagramas de dispersión de distintas variables para determinar si existen relaciones entre las variables del conjunto de datos.

*Nota:* Este ejemplo utiliza *customer\_subset*.

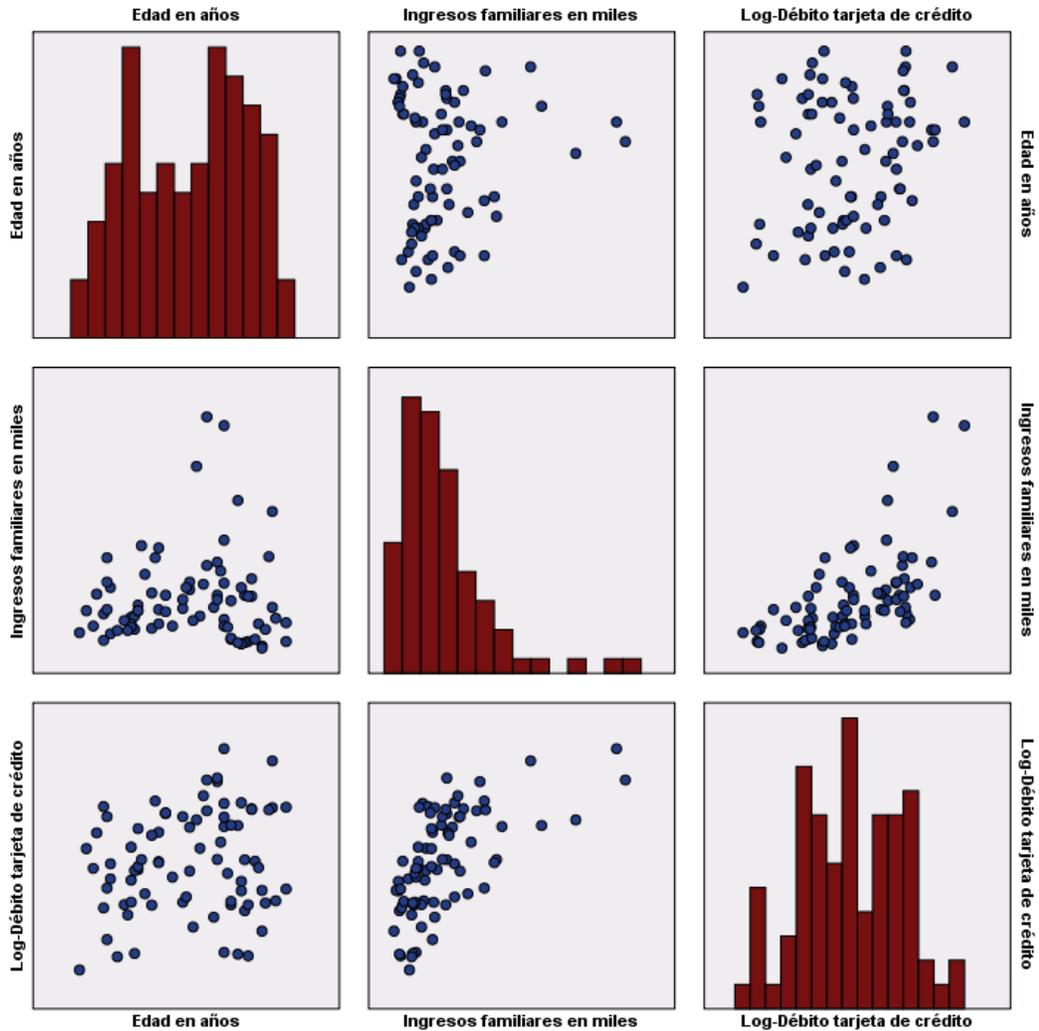
- ▶ Añada un nodo Tablero y ábralo para editarlo.
- ▶ En la pestaña Básico, seleccione *Edad en años*, *Ingresos del hogar en miles* y *Deuda de tarjeta de crédito en miles*. (Mantenga pulsada la tecla Ctrl mientras pulsa para seleccionar varios campos o variables.)
- ▶ Seleccione SPLOM.

Figura 5-25  
Selecciones de la pestaña Básico, SPLOM



- ▶ Pulse en Ejecutar.
- ▶ Maximice la ventana de resultados para ver la matriz con mayor claridad.

Figura 5-26  
Matriz de diagramas de dispersión (SPLOM)



Podemos observar lo siguiente:

- Los histogramas que aparecen en la diagonal muestran la distribución de cada variable de la SPLOM. El histograma de *edad* aparece en la casilla superior izquierda, el de *ingresos* en la casilla central y el de *deudacred* en la casilla inferior derecha. Ninguna de las variables aparece distribuida con normalidad. Es decir, ninguno de los histogramas parece una curva de campana. Además, tenga en cuenta que los histogramas de *ingresos* y *deudacred* son positivamente asimétricos.
- No parece que exista ninguna relación entre *edad* y el resto de variables.
- Hay una relación lineal entre *ingresos* y *deudacred*. Es decir, el valor de *deudacred* aumenta a la par que el valor de *ingresos*. Puede que desee crear diagramas de dispersión individuales de estas variables y las otras variables relacionadas para ampliar las relaciones.

**Ejemplo: Coropleta (mapa de color) de sumas**

Ahora crearemos una visualización de mapas. Luego, en el siguiente ejemplo, crearemos una variación de esta visualización. El conjunto de datos *worldsales*, que es un archivo de datos hipotéticos que contiene ingresos por ventas clasificados por continentes y productos.

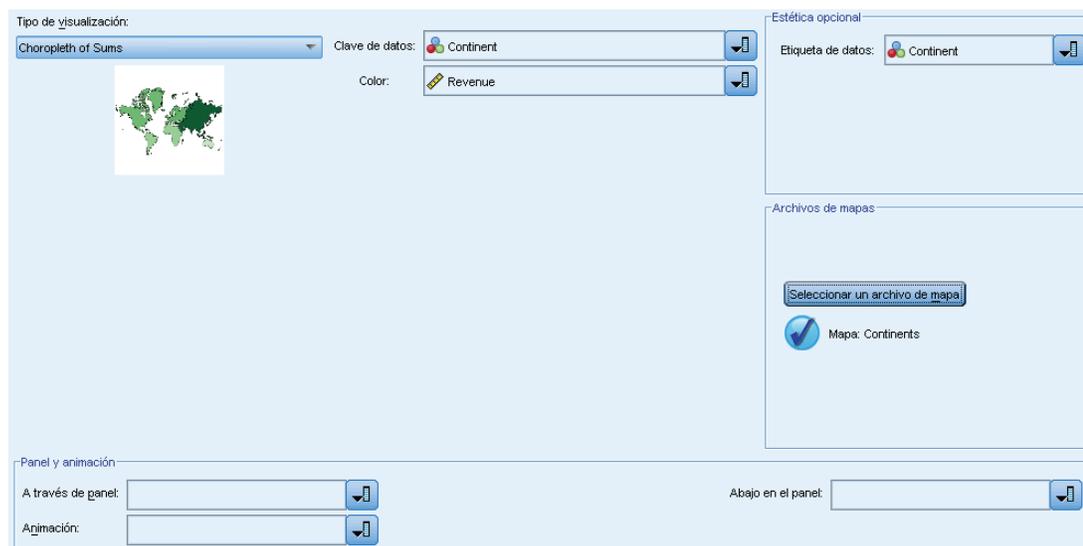
- ▶ Añada un nodo Tablero y ábralo para editarlo.
- ▶ En la pestaña Básico, seleccione *Continente e Ingreso*. (Mantenga pulsada la tecla Ctrl mientras pulsa para seleccionar varios campos o variables.)
- ▶ Seleccione Coropleta de sumas.
- ▶ Pulse en la pestaña Detallado.
- ▶ En el grupo Aspectos estéticos opcionales, seleccione *Continente* en la lista desplegable Etiquetas de datos.
- ▶ En el grupo Archivos de mapas, pulse Seleccionar un archivo de mapas.
- ▶ En el cuadro de diálogo Seleccionar mapas, compruebe que Mapa está configurada en *Continentes* y Leyenda en *CONTINENTE*.
- ▶ En los grupos Comparar mapa y Valores de datos, haga clic en Comparar para asegurarse de que las leyendas coinciden con las claves de datos. En este ejemplo, todos los valores de las claves de datos han coincidido con las leyendas y las características. También podemos ver que ningún dato coincide con Oceanía.

Figura 5-27  
 Seleccione el cuadro de diálogo Mapas



- En el cuadro de diálogo Seleccionar mapas, haga clic en Aceptar.

Figura 5-28  
 Selecciones de la pestaña Básico, Coropleta de sumas



- Pulse en Ejecutar.

Figura 5-29  
Coropleta de sumas



Desde la visualización de mapas, podemos ver fácilmente que los ingresos son superiores en Norteamérica e inferiores en Sudamérica y África. Cada continente está etiquetado porque utilizamos *Continente* para la estética de las etiquetas de datos.

### **Ejemplo: Gráficos de barras de un Mapa**

Este ejemplo muestra cómo se dividen los ingresos por producto en cada continente.

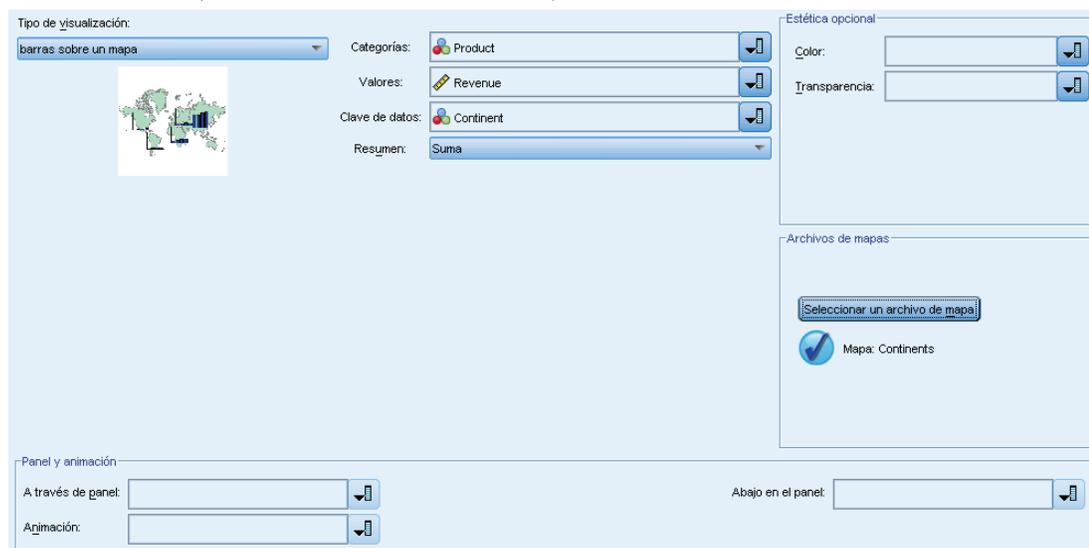
*Nota:* Este ejemplo utiliza *worldsales*.

- Añada un nodo Tablero y ábralo para editarlo.
- En la pestaña Básico, seleccione *Continente*, *Producto*, e *Ingreso*. (Mantenga pulsada la tecla Ctrl mientras pulsa para seleccionar varios campos o variables.)
- Seleccione Barras de un mapa.
- Pulse en la pestaña Detallado.

Si utiliza más de un campo de un tipo específico, es importante comprobar que cada campo está asignado al bloque correcto.

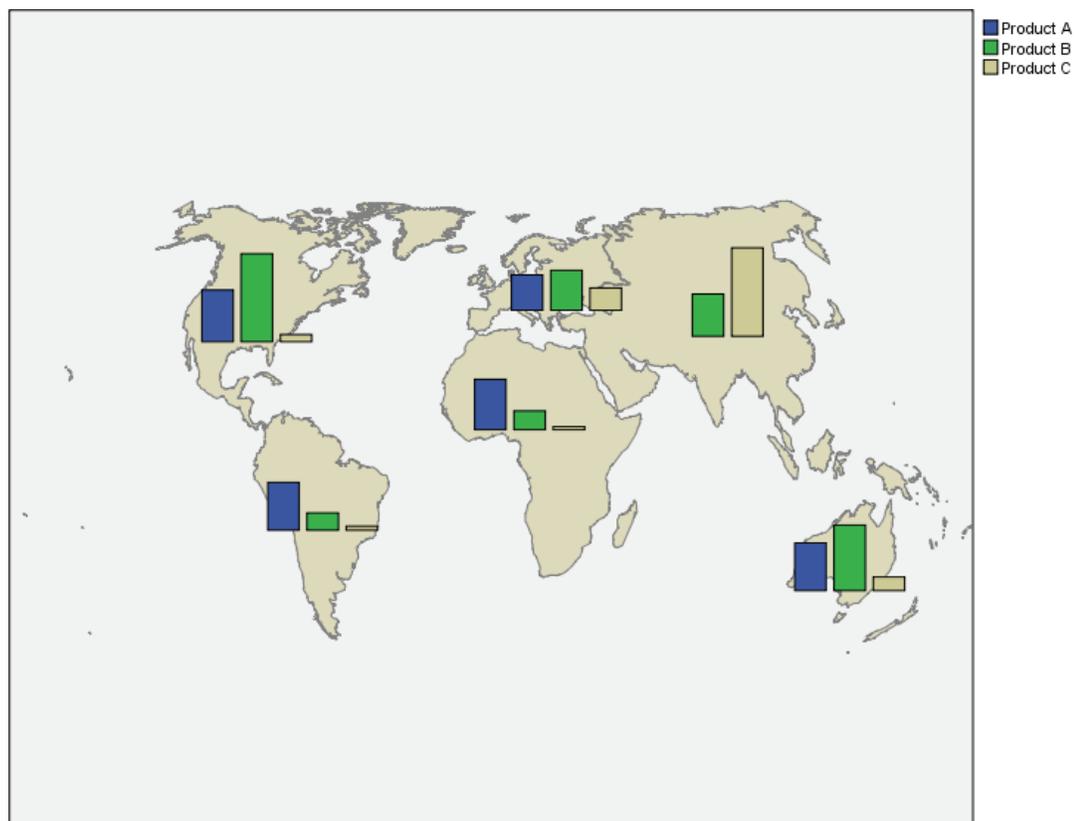
- ▶ En la lista desplegable Categorías, seleccione *Producto*.
- ▶ En la lista desplegable Valores, seleccione *Ingreso*.
- ▶ En la lista desplegable Clave de datos, seleccione *Continente*.
- ▶ En la lista desplegable Resumen, seleccione *Suma*.
- ▶ En el grupo Archivos de mapas, pulse Seleccionar un archivo de mapas.
- ▶ En el cuadro de diálogo Seleccionar mapas, compruebe que Mapa está configurada en *Continentes* y Leyenda en *CONTINENTE*.
- ▶ En los grupos Comparar mapa y Valores de datos, haga clic en Comparar para asegurarse de que las leyendas coinciden con las claves de datos. En este ejemplo, todos los valores de las claves de datos han coincidido con las leyendas y las características. También podemos ver que ningún dato coincide con Oceanía.
- ▶ En el cuadro de diálogo Seleccionar mapas, haga clic en Aceptar.

Figura 5-30  
Selecciones de la pestaña Básico, Barras en un mapa



- ▶ Pulse en Ejecutar.
- ▶ Maximice la ventana de resultados para ver la pantalla con mayor claridad.

Figura 5-31  
Gráficos de barras de un Mapa



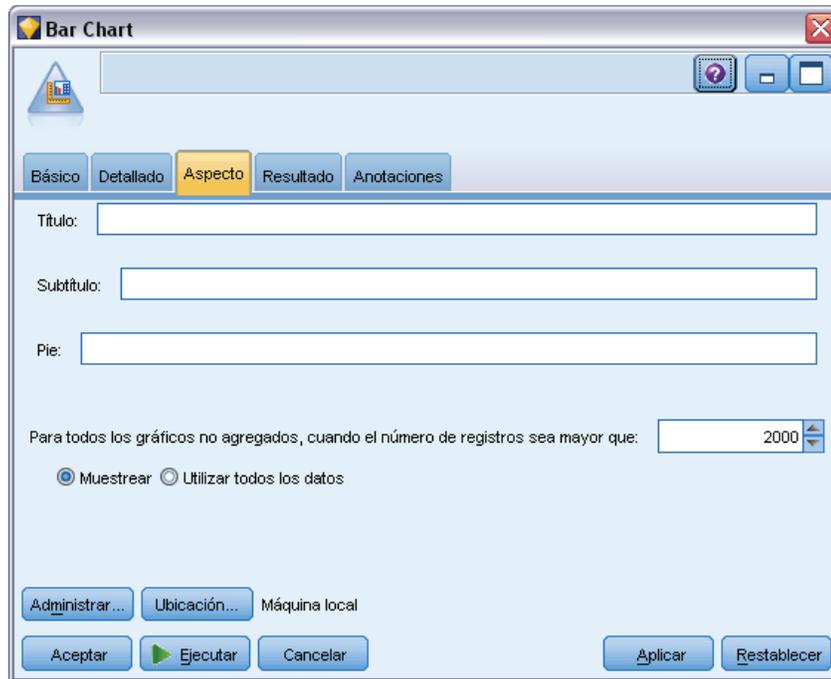
Podemos observar lo siguiente:

- La distribución del total de ingresos en productos es muy similar en Sudamérica y África.
- *Producto C* genera los menores ingresos en todo el mundo excepto en Asia.
- No hay ingresos o ingresos mínimos del *Producto A* en Asia.

### ***Pestaña Aspecto de Tablero***

Puede especificar las opciones de aspecto antes de crear gráficos.

Figura 5-32  
Configuración de la pestaña Aspecto de un nodo Tablero



### **Opciones generales de aspecto**

**Título.** Introduzca el texto que desea para el título del gráfico.

**Subtítulo.** Introduzca el texto que desea para el subtítulo del gráfico.

**Pie.** Introduzca el texto que desea para el pie del gráfico.

**Muestreo.** Especifique un método para conjuntos de datos grandes. Puede especificar un tamaño de conjunto de datos máximo o utilizar el número por defecto de registros. El rendimiento mejora en el caso de conjuntos de datos grandes al seleccionar la opción Muestra. Si lo prefiere, puede trazar todos los puntos de datos seleccionando Utilizar todos los datos, pero debe tener en cuenta que esto puede hacer que el rendimiento del software descienda radicalmente.

### **Opciones de aspecto de la hoja de estilo**

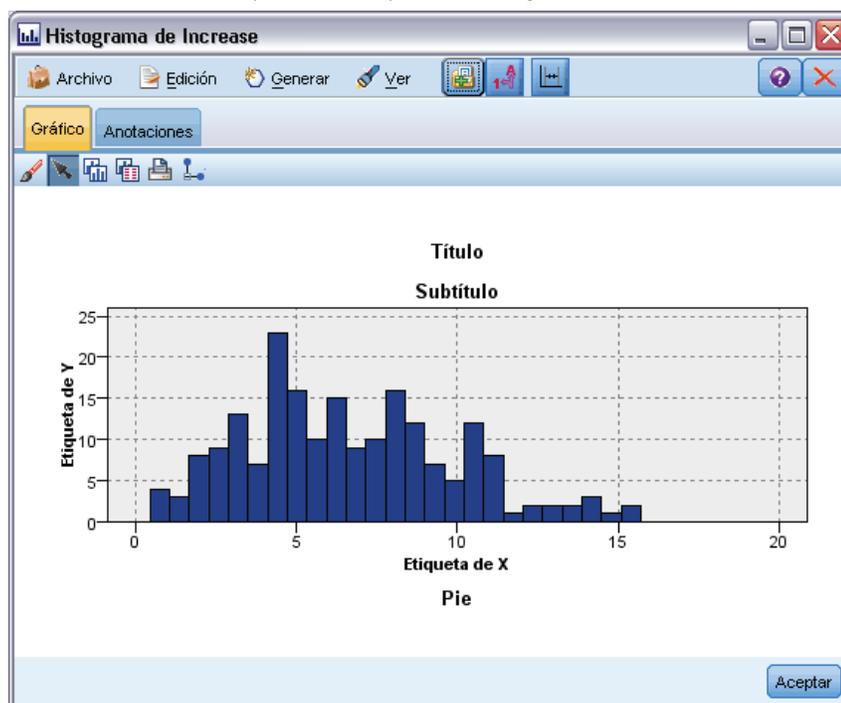
Existen dos botones que permiten controlar las plantillas de visualización (y hojas de estilo y mapas) disponibles:

**Administrar.** Administra plantillas, hojas de estilo y mapas de visualización en su ordenador. Puede importar, exportar, cambiar el nombre y eliminar plantillas, hojas de estilo y mapas de visualización en su máquina local. Si desea obtener más información, consulte el tema [Administración de plantillas, hojas de estilo y archivos de mapas](#) el p. 289.

**Ubicación.** Cambia la ubicación donde están almacenadas las plantillas, hojas de estilo y mapas de visualización. La ubicación actual se enumera a la derecha del botón. Si desea obtener más información, consulte el tema [Configuración de la ubicación de las plantillas, hojas de estilo y mapas](#) el p. 287.

El siguiente ejemplo muestra dónde se colocan las opciones de apariencia en el gráfico. (*Nota:* No todos los gráficos utilizan estas opciones.)

Figura 5-33  
Posición de las varias opciones de apariencia de gráficos



### **Configuración de la ubicación de las plantillas, hojas de estilo y mapas**

Las plantillas y hojas de estilo de visualización y los archivos de mapas se guardan en una carpeta local específica o en el IBM® SPSS® Collaboration and Deployment Services Repository. Al seleccionar plantillas, hojas de estilo y mapas, sólo se muestran los incorporados en esta ubicación. Al mantener todas las plantillas, hojas de estilo y archivos de mapas en una ubicación, las aplicaciones de IBM SPSS pueden acceder a ellas fácilmente. Si desea obtener más información sobre cómo agregar más plantillas, hojas de estilo y archivos de mapas a esta ubicación, consulte [Administración de plantillas, hojas de estilo y archivos de mapas](#) el p. 289.

#### **Cómo se configura la ubicación de las plantillas, hojas de estilo y archivos de mapas**

- En el cuadro de diálogo de plantillas o de hojas de estilo, pulse Ubicación... para abrir el cuadro de diálogo Plantillas, hojas de estilo y mapas.

- ▶ Seleccione una opción para la ubicación por defecto de las plantillas, hojas de estilo y mapas:

**Máquina local.** Las plantillas, hojas de estilo y archivos de mapas se guardan en una carpeta concreta o en su máquina local. En Windows XP, esta carpeta es *C:\Documents and Settings\<user>\Application Data\SPSSInc\Graphboard*. La carpeta no se puede cambiar.

**IBM® SPSS® Collaboration and Deployment Services Repository.** Las plantillas, hojas de estilo y archivos de mapas se guardan en una carpeta del IBM SPSS Collaboration and Deployment Services Repository especificada por el usuario. Para identificar la carpeta específica, pulse Carpeta. Si desea obtener más información, consulte [Uso del IBM SPSS Collaboration and Deployment Services Repository como ubicación de las plantillas, hojas de estilo y mapas](#) el p. 288.

- ▶ Pulse en Aceptar.

### ***Uso del IBM SPSS Collaboration and Deployment Services Repository como ubicación de las plantillas, hojas de estilo y mapas***

Las plantillas y hojas de estilo de visualización se pueden almacenar en el IBM® SPSS® Collaboration and Deployment Services Repository. Esta ubicación es una carpeta específica en el IBM SPSS Collaboration and Deployment Services Repository. Si se define como la ubicación predeterminada, todas las plantillas, hojas de estilo y archivos de mapas de esta ubicación estarán disponibles para su selección.

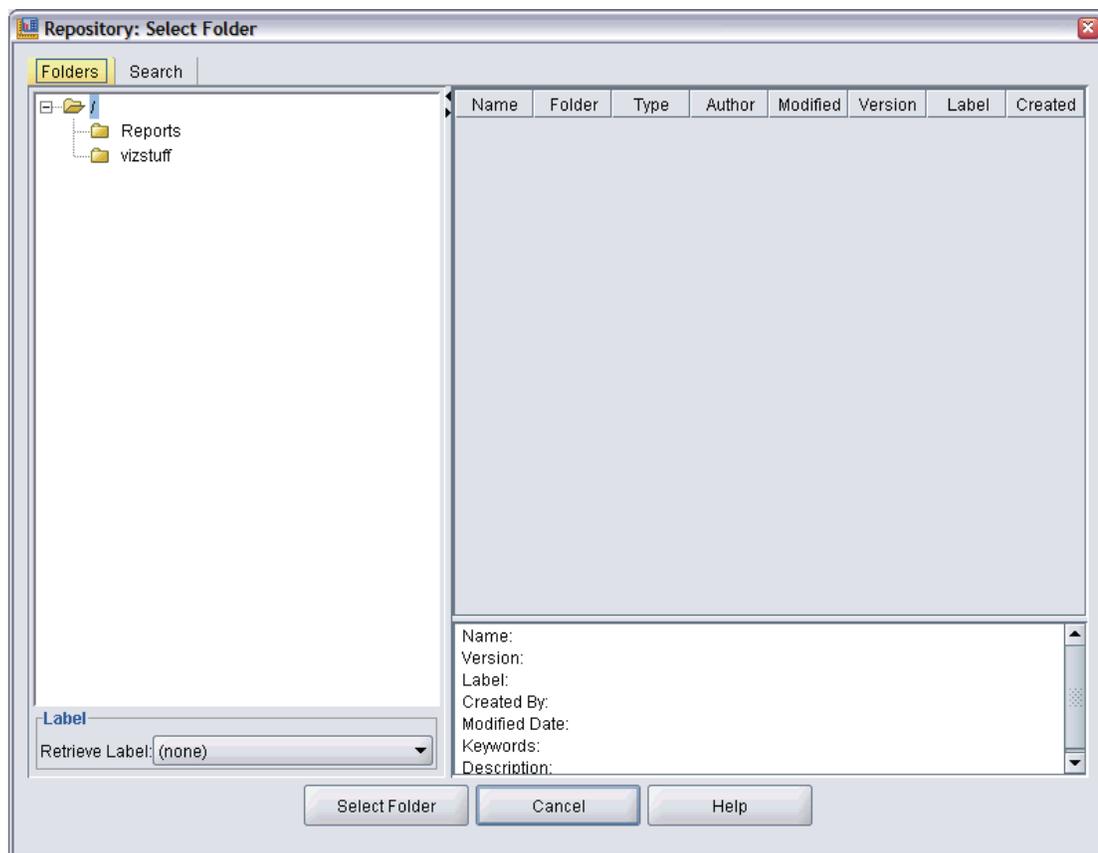
#### ***Cómo se configura una carpeta en el IBM SPSS Collaboration and Deployment Services Repository como la ubicación de las plantillas, hojas de estilo y archivos de mapas***

- ▶ En un cuadro de diálogo con un botón Ubicación, pulse Ubicación....
- ▶ Seleccione IBM® SPSS® Collaboration and Deployment Services Repository.
- ▶ Pulse en Carpeta.

*Nota:* Si aún no está conectado al IBM SPSS Collaboration and Deployment Services Repository, se le solicitará la información de conexión.

- ▶ En el cuadro de diálogo Seleccionar carpeta, seleccione la carpeta en la que se guardarán las plantillas, hojas de estilo y archivos de mapas.

Figura 5-34  
Cuadro de diálogo Seleccionar carpeta



- ▶ Si lo desea, seleccione una etiqueta en Recuperar etiqueta. Sólo se mostrarán las plantillas, hojas de estilo y archivos de mapas con esa etiqueta.
- ▶ Si busca una carpeta que contiene una plantilla u hoja de estilo concreta, es posible que desee buscar la plantilla, hoja de estilo o archivo de mapa en la pestaña Buscar. El cuadro de diálogo Seleccionar carpeta selecciona automáticamente la carpeta en la que se encuentra la plantilla, hoja de estilo o archivo de mapa.
- ▶ Pulse Seleccionar carpeta.

### ***Administración de plantillas, hojas de estilo y archivos de mapas***

Puede administrar las plantillas, hojas de estilo y archivos de mapas en la ubicación local de su equipo utilizando el cuadro de diálogo Administrar plantillas, hojas de estilo y mapas. Este cuadro de diálogo le ayuda a importar, exportar, cambiar el nombre y eliminar plantillas, hojas de estilo y archivos de mapas de visualización en la ubicación local de su equipo.

- ▶ Pulse Administrar... en uno de los cuadros de diálogo de selección de las plantillas, hojas de estilo o mapas.

**Cuadro de diálogo Administrar plantillas, hojas de estilo y mapas...**

La pestaña Plantilla muestra todas las plantillas locales. La pestaña Hoja de estilo muestra todas las hojas de estilo locales, así como visualizaciones de ejemplo con datos de muestra. Puede seleccionar una de las hojas de estilo que aplicarán sus estilos a las visualizaciones de ejemplo. Si desea obtener más información, consulte el tema [Aplicación de hojas de estilo](#) el p. 387. La pestaña Mapa muestra todos los archivos de mapas locales. Esta pestaña también muestra las leyendas incluyendo valores de muestra, comentarios si se hizo alguno al crear el mapa y una vista previa del mapa.

Los botones siguientes funcionan en cualquier pestaña que esté activada actualmente.

**Importar.** Importa una plantilla, hoja de estilo o archivo de mapas de visualización desde el sistema de archivos. Si importa una plantilla, hoja de estilo o archivo de mapas, estará disponible para la aplicación de IBM SPSS. Si otro usuario ha enviado una plantilla, hoja de estilo o archivo de mapas, se importa antes de utilizarlo en su aplicación.

**Exportar.** Exporta una plantilla, hoja de estilo o archivo de mapas de visualización al sistema de archivos. Exporte una plantilla, hoja de estilo o archivo de mapas si desea enviarlo a otro usuario.

**Cambiar nombre.** Cambia el nombre de la plantilla, hoja de estilo o archivo de mapas de visualización seleccionado. No puede cambiar un nombre por uno que ya esté en uso.

**Exportar leyenda.** Exporta las leyendas como un archivo de valores separados por comas (CSV). Este botón solo está activado en la pestaña Mapa.

**Eliminar.** Elimina la plantilla, hoja de estilo o archivo de mapas de visualización seleccionado. Puede seleccionar varias plantillas, hojas de estilo o archivos de mapas dejando pulsado Ctrl . No puede deshacer esta acción, por lo que debe proceder con cuidado.

**Conversión y distribución de mapas en formato shapefile**

El selector de plantillas de tablero le permite crear visualizaciones de mapas a partir de la combinación de una plantilla de visualización y de un archivo SMZ. Los archivos SMZ se parecen a los archivos con formato shapefile ESRI (formato de archivo SHP) en los que se contiene la información geográfica para dibujar un mapa (por ejemplo, las fronteras de los países), pero están optimizados para la visualización de mapas. El selector de plantillas de tablero tiene preinstalados varios archivos SMZ. Si ya cuenta con un formato ESRI shapefile y desea utilizarlo para visualizar mapas, primero debe convertir el formato shapefile a un archivo con formato SMZ mediante la utilidad de conversión de mapas. La utilidad de conversión de mapas admite formatos ESRI shapefile con puntos, polilíneas o polígonos (tipos de formas 1, 3 y 5) que contengan una única capa.

Además de convertir formatos ESRI shapefile, la utilidad de conversión de mapas le permite modificar el nivel de detalle de un mapa, cambiar las etiquetas de funciones, combinar y mover funciones, entre muchos otros cambios opcionales. También puede utilizar la utilidad de conversión de mapas para modificar un archivo SMZ existente (incluyendo los ya instalados).

### **Edición de archivos SMZ preinstalados**

- ▶ Exporte el archivo SMZ del sistema de gestión. Si desea obtener más información, consulte el tema [Administración de plantillas, hojas de estilo y archivos de mapas](#) el p. 289.
- ▶ Utilice la utilidad de conversión de mapas para abrir y editar el archivo SMZ exportado. Se recomienda que guarde el archivo con un nombre diferente. Si desea obtener más información, consulte el tema [Uso de la utilidad de conversión de mapas](#) el p. 292.
- ▶ Importe el archivo SMZ modificado en el sistema de gestión. Si desea obtener más información, consulte el tema [Administración de plantillas, hojas de estilo y archivos de mapas](#) el p. 289.

### **Recursos adicionales para archivos de mapas**

Los datos geoespaciales en el formato de archivo SHP, que se deben utilizar para cubrir sus necesidades de asignación desde varias fuentes tanto privadas como públicas. Consulte con los sitios Web de administraciones públicas si busca datos gratuitos. En este producto, muchas de las plantillas están basadas en datos disponibles para el público y que se obtienen en GeoCommons (<http://www.geocommons.com>) y el Bureau of the Census de EE.UU. (<http://www.census.gov>). Otra fuente de EE.UU. de datos geoespaciales locales, federales y estatales es el Servicio geológico de EE.UU. (<http://www.geodata.gov>).

**NOTA IMPORTANTE:** Se ha obtenido información acerca de productos que no son de IBM de los proveedores de esos productos, de sus publicaciones anunciadas o de otras fuentes disponibles públicamente. IBM no ha probado estos productos y no puede confirmar la precisión de su rendimiento, su compatibilidad o cualquier otra reclamación relacionada con productos que no sean de IBM. Las preguntas acerca de las aptitudes de productos que no sean de IBM deben dirigirse a los proveedores de dichos productos. Cualquier referencia a sitios Web que no sean de IBM en esta información solamente es ofrecida por comodidad y de ningún modo sirve como aprobación de esos sitios Web. Los materiales en esos sitios Web no son parte de los materiales de este programa de IBM, a menos que esté indicado en un archivo de avisos junto a este programa de IBM y el uso de los sitios de materiales será bajo su propia responsabilidad.

## **Conceptos claves de los mapas**

Comprender algunos conceptos clave relacionados con el formato shapefile le ayudará a utilizar de manera eficaz la utilidad de conversión de mapas.

Un **shapefile** proporciona la información geográfica para dibujar un mapa. Existen tres tipos de formatos shapefile compatibles con la utilidad de conversión de mapas:

- **Punto.** El formato shapefile identifica las localizaciones de los puntos como, por ejemplo, las ciudades.
- **Polilínea.** El formato shapefile identifica las rutas y sus localizaciones como, por ejemplo, los ríos.
- **Polígono.** El formato shapefile identifica las regiones limítrofes y sus localizaciones, como los países.

Por lo general utilizará un formato shapefile de polígonos. Los mapas de coropletas se crean a partir de los formatos shapefile de polígonos. Los mapas de coropletas utilizan colores para representar un valor dentro de polígonos individuales (regiones). Los formatos shapefile de puntos y polilíneas se superponen normalmente en un formato shapefile de polígonos. Un ejemplo de ello es un formato shapeline de puntos de las ciudades de EE.UU. superpuesto sobre un formato shapeline con polígonos de los estados de EE.UU.

Un shapefile se compone de **funciones**. Las funciones son las entidades geográficas individuales. Por ejemplo, las funciones pueden ser países, estados, ciudades, etc. El shapefile también contiene datos acerca de las funciones. Estos datos se almacenan en **atributos**. Los atributos son similares a los campos o variables en un archivo de datos. Hay al menos un atributo que es la **leyenda del mapa** de la función. La leyenda del mapa puede ser una etiqueta como, por ejemplo, el nombre de un país o de un estado. La leyenda del mapa es lo que vinculará a una variable o campo en un archivo de datos para crear la visualización de un mapa.

Tenga en cuenta que solamente podrá retener el atributo o los atributos clave en el archivo SMZ. La utilidad de conversión de mapas no puede guardar atributos adicionales. Esto significa que necesitará crear varios archivos SMZ si quiere agregarlos a distintos niveles. Por ejemplo, si quiere agregar estados y regiones de EE.UU., necesitará distintos archivos SMZ: uno que tenga una clave para identificar estados y otro que tenga una clave para identificar regiones.

## ***Uso de la utilidad de conversión de mapas***

### ***Uso de la utilidad de conversión de mapas***

- ▶ Seleccione en los menús:  
Herramientas > Utilidad de conversión de mapas

La utilidad de conversión de datos consta de cuatro escenas (pasos) principales. Uno de los pasos también consta de pasos secundarios para un control más detallado de la edición del archivo de mapas.

### ***Paso 1: Elegir los archivos de destino y de fuente***

Primero tiene que seleccionar un archivo de mapas de fuente y uno de destino del archivo del mapa convertido. Necesitará los archivos *.shp* y *.dbf* del formato shapefile.

**Seleccionar un archivo *.shp* (ESRI) o *.smz* para la conversión.** Vaya a un archivo de mapas existente de su equipo. Este es el archivo que convertirá y guardará como un archivo SMZ. El archivo *.dbf* del formato shapefile *debe* almacenarse en la misma ubicación con un nombre base de archivo que coincida con el archivo *.shp*. El archivo *.dbf* es necesario porque contiene la información del atributo del archivo *.shp*.

**Definir un destino y un nombre de archivo para el archivo del mapa convertido.** Introduzca una ruta y un nombre de archivo para el archivo SMZ que se creará a partir de la fuente del mapa original.

## **Paso 2: Elegir la leyenda del mapa**

Ahora elegirá las leyendas del mapa que va a incluir en el archivo SMZ. A continuación, puede cambiar algunas opciones que afectarán a la presentación del mapa. Los pasos posteriores en la utilidad de conversión de mapas incluyen una presentación preliminar del mapa. Las opciones de presentación que elija se utilizarán para generar la presentación preliminar del mapa.

**Elegir la leyenda principal del mapa.** Seleccione el atributo que es la clave principal para la identificación y el etiquetado de las funciones del mapa. Por ejemplo, la leyenda principal de un mapa del mundo puede ser el atributo que identifica los nombres de los países. La clave principal también vinculará sus datos a las funciones del mapa, así que asegúrese de que los valores (etiquetas) del atributo que elija coinciden con los valores de sus datos. Al elegir un atributo, se muestran las etiquetas a modo de ejemplo. Si necesita cambiar estas etiquetas, podrá hacerlo en el último paso.

**Elegir las leyendas adicionales que vaya a incluir.** Además de la leyenda principal del mapa, elija otros atributos clave que desee incluir en el archivo SMZ generado. Por ejemplo, algunos atributos pueden contener etiquetas traducidas. Si quiere que los datos se codifiquen en otros idiomas, puede que desee preservar estos atributos. Tenga en cuenta que solamente puede elegir aquellas claves adicionales que representen las mismas funciones que la clave principal. Por ejemplo, si la clave principal fuese los nombres completos de los estados de EE.UU., puede seleccionar solamente aquellas claves alternativas que representen los estados de EE.UU., como las abreviaturas de los estados.

**Suavizar automáticamente el mapa.** Los shapefiles con polígonos contienen normalmente demasiados puntos de datos y demasiados detalles para las visualizaciones estadísticas de los mapas. El exceso de detalles puede distraer e impactar de forma negativa en el rendimiento. Puede reducir el nivel de detalles y generalizar el mapa con el suavizado. El mapa será más nítido y se presentará más rápidamente. Cuando se suaviza el mapa automáticamente, el ángulo máximo es de 15 grados y el porcentaje que hay que mantener es 99. Para obtener información acerca de esta configuración, consulte [Suavizar el mapa](#) el p. 294. Tenga en cuenta que tiene la oportunidad de aplicar el suavizado adicional posteriormente en otro paso.

**Eliminar los límites entre los polígonos en la misma función.** Algunas funciones pueden incluir funciones secundarias con límites internos en las funciones principales de interés. Por ejemplo, un mapa del mundo con los continentes puede contener límites internos para los países dentro de cada continente. Si selecciona esta opción, los límites internos no aparecerán en el mapa. En el ejemplo del mapa del mundo con los continentes, si selecciona esta opción se eliminarían los límites de los países y se mantendrían las fronteras de los continentes.

## **Paso 3: Editar el mapa**

Ahora que ha especificado las opciones básicas del mapa, puede editar opciones más específicas. Estas modificaciones son opcionales. Este paso de la utilidad de conversión de mapas le guía a través de las tareas asociadas y muestra una presentación preliminar del mapa para que pueda verificar sus cambios. En función del tipo de formato shapefile (punto, polilínea o polígono) y del sistema de coordenadas, puede que algunas tareas no estén disponibles.

Cada tarea tiene los siguientes controles comunes en la parte izquierda de la utilidad de conversión de mapas.

**Mostrar las etiquetas del mapa.** Por defecto, las etiquetas de funciones no se muestran en la presentación preliminar. Puede elegir mostrar las etiquetas. Aunque las etiquetas son útiles para identificar funciones, pueden interferir con la selección directa del mapa de presentación preliminar. Active esta opción cuando lo necesite; por ejemplo, al editar las etiquetas de funciones.

**Aplicar color a la presentación preliminar del mapa.** Por defecto, el mapa de presentación preliminar muestra las áreas con un único color. Todas las funciones son del mismo color. Puede disponer de una gama de colores asignados a las funciones individuales del mapa. Esta opción puede ser útil para distinguir distintas funciones en el mapa. Será de especial ayuda si combina funciones y desea ver cómo se representan las nuevas funciones en la presentación preliminar.

Cada tarea tiene el siguiente control común en la parte derecha de la utilidad de conversión de mapas.

**Deshacer.** Si introduce un cambio no deseado, pulse en Deshacer para volver al estado anterior. Puede deshacer un máximo de 100 cambios.

### ***Suavizar el mapa***

Los shapefiles con polígonos contienen normalmente demasiados puntos de datos y demasiados detalles para las visualizaciones estadísticas de los mapas. El exceso de detalles puede distraer e impactar de forma negativa en el rendimiento. Puede reducir el nivel de detalles y generalizar el mapa con el suavizado. El mapa será más nítido y se presentará más rápidamente. Esta opción no está disponible para los mapas de puntos y de polilíneas.

**Ángulo máximo.** El ángulo máximo, que debe ser un valor comprendido entre 1 y 20, especifica la tolerancia al suavizado de conjuntos de puntos casi lineales. Un valor más alto permite más tolerancia al suavizado lineal y mostrará posteriormente más puntos, dando como resultado un mapa más generalizado. Para aplicar el suavizado lineal, la utilidad de conversión de mapas comprueba el ángulo interior formado por cada conjunto de tres puntos en el mapa. Si 180 menos el ángulo da un resultado inferior al valor especificado, la utilidad de conversión de mapas mostrará el punto medio. Es decir, la utilidad de conversión de mapas comprueba si la línea formada por los tres puntos es una línea casi recta. Si es así, la utilidad de conversión de mapas considerará la línea como una línea recta entre los extremos y mostrará el punto medio.

**Porcentaje que hay que mantener.** El porcentaje que hay que mantener, que debe ser un valor comprendido entre 90 y 100, determina la cantidad de terreno que hay que mantener cuando se suaviza el mapa. Esta opción afecta solamente a aquellas funciones que tienen varios polígonos, como podría ser el caso de una función que incluye islas. Si el área total de una función menos un polígono es superior al porcentaje especificado del área original, la utilidad de conversión de mapas mostrará el polígono del mapa. La utilidad de conversión de mapas nunca eliminará todos los polígonos de una función. Es decir, siempre habrá al menos un polígono de la función, con independencia de la cantidad de suavizado aplicada.

Cuando seleccione un ángulo máximo y un porcentaje, pulse en Aplicar. Se actualizará la presentación preliminar con los cambios de suavizado. Si necesita suavizar de nuevo el mapa, repita el proceso hasta obtener el nivel deseado de homogeneidad. Tenga en cuenta que existe un límite para aplicar el suavizado. Si suaviza repetidamente, llegará a un punto en el que no podrá aplicar más suavizado al mapa.

### **Editar las etiquetas de funciones**

Puede editar las etiquetas de funciones según sea necesario (quizás para igualar sus datos) y también puede volver a posicionar las etiquetas en el mapa. Incluso si piensa que no necesita modificar las etiquetas, debería revisarlas antes de crear visualizaciones del mapa. Debido a que las etiquetas no se muestran por defecto en la presentación preliminar, puede que también desee seleccionar Mostrar las etiquetas en el mapa para mostrarlas.

**Claves.** Seleccione la clave que contenga las etiquetas de funciones que desee revisar y/o editar.

**Funciones.** Esta lista muestra las etiquetas de funciones presentes en la clave seleccionada. Para editar la etiqueta, pulse dos veces sobre ella en la lista. Si se muestran las etiquetas del mapa, también puede pulsar dos veces en las etiquetas de funciones directamente en la presentación preliminar del mapa. Si desea comparar las etiquetas con un archivo de datos real, pulse en Comparar.

**X/Y.** Estos cuadros de texto incluyen el punto central actual de la etiqueta seleccionada de la función del mapa. Las unidades se muestran en las coordenadas del mapa. Éstas pueden ser coordenadas locales, coordenadas cartesianas (por ejemplo, el sistema estatal de coordenadas planas) o coordenadas geográficas (donde X es la longitud e Y la latitud). Introducir las coordenadas para la nueva posición de la etiqueta. Si se muestran las etiquetas, también puede pulsar y arrastrar una etiqueta del mapa para moverla. Se actualizarán los cuadros de texto con la nueva posición.

**Comparar.** Si tiene un archivo de datos que contenga los valores de los datos que se pretenden igualar a las etiquetas de funciones de una clave determinada, pulse en Comparar para mostrar el cuadro de diálogo Comparar con una fuente de datos externa. En este cuadro de diálogo podrá abrir el archivo de datos y comparar sus valores directamente con aquellos valores de las etiquetas de funciones de la leyenda del mapa.

### **Cuadro de diálogo Comparar con una fuente de datos externa**

El cuadro de diálogo Comparar con una fuente de datos externa, le permite abrir un archivo de valores separados por tabuladores (con una extensión *.txt*) o un archivo de valores separados por comas (con una extensión *.csv*). Cuando se abre el archivo, puede seleccionar un campo en el archivo de datos para compararlo con las etiquetas de funciones de una leyenda específica del mapa. A continuación, puede corregir todas las discrepancias en el archivo del mapa.

**Campos en el archivo de datos.** Seleccione el campo cuyos valores desee comparar con las etiquetas de funciones. Si la primera fila del archivo *.txt* o *.csv* contiene etiquetas descriptivas de cada campo, marque Utilizar primera fila como etiquetas de columna. De lo contrario, se identificará cada campo según su posición en el archivo de datos (por ejemplo, “Columna 1”, “Columna 2”, etc.).

**Clave para comparar.** Seleccione la leyenda del mapa cuyas etiquetas de funciones desee comparar con los valores del campo del archivo de datos.

**Comparar.** Pulse cuando esté preparado para comparar los valores.

**Resultados comparativos.** Por defecto, la tabla de resultados comparativos solamente muestra los valores del campos no coincidentes en el archivo de datos. La aplicación intenta buscar una etiqueta de funciones relacionada, normalmente comprobando los espacios insertados o inexistentes. Pulse en la lista desplegable en la columna *Etiqueta del mapa* para hacer coincidir la etiqueta de funciones en el archivo del mapa con el valor del campo mostrado. Si no hay ninguna etiqueta de funciones correspondiente en su archivo de mapas, seleccione *Dejar sin marcar*. Si desea ver todos los valores del campo, incluso aquellos que ya coinciden con una etiqueta de funciones, cancele la selección de *Mostrar solamente los casos no coincidentes*. Puede que desee realizar esta acción para anular una o más coincidencias.

Puede utilizar cada función una sola vez para igualarla a un valor de campo. Si desea igualar varias funciones con un único valor de campo, puede combinar las funciones y, a continuación, igualar la nueva función combinada con el valor del campo. Si desea obtener más información acerca de la combinación de funciones, consulte [Combinar funciones](#) el p. 296.

### ***Combinar funciones***

La combinación de funciones es útil para crear regiones más grandes en un mapa. Por ejemplo, si estuviese convirtiendo un mapa de estados, podría combinar los estados (las funciones en este ejemplo) con regiones más grandes del Norte, Sur, Este y Oeste.

**Claves.** Seleccione la leyenda del mapa que contenga las etiquetas de funciones que le ayudarán a identificar las funciones que desea combinar.

**Funciones.** Pulse en la primera función que desee combinar. Utilice la tecla Ctrl para pulsar en las otras funciones que desee combinar. Tenga en cuenta que también se seleccionarán las funciones en la presentación preliminar del mapa. Puede pulsar utilizando la tecla Ctrl directamente en las funciones de la presentación preliminar del mapa además de seleccionarlas de la lista.

Después de seleccionar las funciones que desee combinar, pulse en *Combinar* para mostrar el cuadro de diálogo *Poner nombre a la función combinada*, donde podrá aplicar una etiqueta a la nueva función. Puede que desee marcar *Colorear* la presentación preliminar del mapa después de combinar las funciones para asegurarse de que los resultados obtenidos son los que esperaba.

Después de combinar las funciones, puede que también quiera mover la etiqueta de la nueva función. Puede realizar esa acción en la tarea *Editar las etiquetas de funciones*. Si desea obtener más información, consulte el tema [Editar las etiquetas de funciones](#) el p. 295.

### ***Cuadro de diálogo Poner nombre a la función combinada***

El cuadro de diálogo *Poner nombre a la función combinada*, le permite asignar etiquetas a la nueva función combinada.

La tabla *Etiquetas* muestra información de cada clave en el archivo de mapas y le permite asignar una etiqueta para cada clave.

**Nueva etiqueta.** Introduzca una nueva etiqueta para la función combinada para asignarla a la leyenda específica del mapa.

**Clave.** La leyenda del mapa a la que va a asignar la nueva etiqueta.

**Etiquetas anteriores.** Las etiquetas de las funciones que se combinarán en una nueva función.

**Eliminar los límites entre los polígonos de la misma función.** Compruebe esta opción para eliminar los límites de las funciones que se combinaron de forma conjunta. Por ejemplo, si ha combinado estados con regiones geográficas, esta opción eliminaría los límites en torno a los estados individuales.

### ***Mover funciones***

Puede mover las funciones en el mapa. Esto puede ser útil cuando desea unir funciones como, por ejemplo, una península e islas alejadas.

**Claves.** Seleccione la leyenda del mapa que contenga las etiquetas de funciones que le ayudarán a identificar las funciones que desea mover.

**Funciones.** Pulse en la función que desee mover. Tenga en cuenta que se seleccionará la función en la presentación preliminar del mapa. También puede pulsar directamente en la función en la presentación preliminar del mapa.

**X/Y.** Estos cuadros de texto incluyen el punto central actual de la función del mapa. Las unidades se muestran en las coordenadas del mapa. Éstas pueden ser coordenadas locales, coordenadas cartesianas (por ejemplo, el sistema estatal de coordenadas planas) o coordenadas geográficas (donde X es la longitud e Y la latitud). Introducir las coordenadas para la nueva posición de la función. También puede pulsar y arrastrar una función del mapa para moverla. Se actualizarán los cuadros de texto con la nueva posición.

### ***Eliminar funciones***

Puede eliminar las funciones no deseadas del mapa. Esto puede resultarle útil cuando desee eliminar algunas perturbaciones eliminando las funciones que no sean de interés en la visualización del mapa.

**Claves.** Seleccione la leyenda del mapa que contenga las etiquetas de funciones que le ayudarán a identificar las funciones que desea eliminar.

**Funciones.** Pulse en la función que desee eliminar. Si desea eliminar varias funciones a la vez, pulse con la tecla Ctrl en las funciones adicionales. Tenga en cuenta que también se seleccionarán las funciones en la presentación preliminar del mapa. Puede pulsar utilizando la tecla Ctrl en las funciones directamente en la presentación preliminar del mapa además de seleccionarlas de la lista.

### ***Eliminar elementos individuales***

Además de eliminar funciones completas, puede eliminar algunos de los elementos individuales que las componen como, por ejemplo, lagos y pequeñas islas. Esta opción no está disponible para los mapas de puntos.

**Elementos.** Pulse en los elementos que desee eliminar. Si desea eliminar varios elementos a la vez, pulse con la tecla Ctrl en los elementos adicionales. Tenga en cuenta que también se seleccionarán los elementos en la presentación preliminar del mapa. Puede pulsar utilizando la tecla Ctrl en los

elementos directamente en la presentación preliminar del mapa además de seleccionarlos de la lista. Debido a que la lista de nombres de elementos no es descriptiva (cada elemento se asigna a un número dentro de la función), debería comprobar la selección en la presentación preliminar del mapa para asegurarse de que ha seleccionado los elementos que quería.

### **Definir la proyección**

La proyección del mapa especifica la forma en la que se representa La Tierra tridimensional en dos dimensiones. Todas las proyecciones causan distorsiones. Sin embargo, algunas proyecciones son más apropiadas dependiendo de si está visualizando un mapa global o un mapa más local. Además, algunas proyecciones preservan la forma de las funciones originales. Las proyecciones que preservan la forma son proyecciones conformes. Esta opción está disponible solamente para los mapas con coordenadas geográficas (longitud y latitud).

A diferencia de otras opciones de la utilidad de conversión de mapas, puede cambiarse la proyección después de haber creado una visualización del mapa.

**Proyección.** Seleccione una proyección del mapa. Si va a crear un mapa global o hemisférico, utilice la proyección *local*, la proyección de *Mercator* o la proyección de *Winkel Tripel*. Para las áreas más pequeñas, utilice la proyección *local*, la proyección *cónica conforme de Lambert* o la proyección *transversa de Mercator*. Todas las proyecciones utilizan el elipsoide WGS83 de los datos.

- La proyección **local** se utiliza siempre que se haya creado el mapa con un sistema de coordenadas local como, por ejemplo, el State Plane Coordinate System (sistema estatal de coordenadas planas). Estos sistemas de coordenadas se definen por las coordenadas cartesianas en lugar de las coordenadas geográficas (longitud y latitud). En la proyección local, las líneas horizontales y verticales se distribuyen por igual en un sistema de coordenadas cartesianas. La proyección local no es una proyección conforme.
- La proyección de **Mercator** es una proyección conforme para los mapas globales. Las líneas horizontales y verticales son rectas y siempre perpendiculares unas a otras. Tenga en cuenta que la proyección de Mercator se extiende hasta el infinito a medida que se va acercando a los polos Norte y Sur, de modo que no podrá utilizarla si su mapa incluye el Polo Norte o el Polo Sur. La distorsión es mayor cuando el mapa se aproxima a estos límites.
- La proyección de **Winkel Tripel** es una proyección no conforme para los mapas globales. Aunque no es una proyección conforme, ofrece un buen balance entre forma y tamaño. A excepción del Ecuador y del Meridiano de Greenwich, todas las líneas son curvas. Esta proyección es una buena opción si su mapa global incluye el Polo Norte o el Polo Sur.
- Como su nombre indica, la proyección **cónica conforme de Lambert** es una proyección conforme y se utiliza para mapas de masas continentales o masas de tierra más pequeñas que son más grandes por el Este y el Oeste que por el Norte y el Sur.
- La proyección **transversa de Mercator** es otra proyección conforme para los mapas de masas continentales y de masas de tierra más pequeñas. Utilice esta proyección para las masas de tierra que son más grandes por el Norte y el Sur en comparación con el Este y el Oeste.

#### **Paso 4: Finalizar**

En este punto puede añadir un comentario para describir el archivo de mapas y así como crear un archivo de mapas de muestra a partir de las leyendas del mapa.

**Leyendas del mapa.** Si hay varias claves en el archivo de mapas, seleccione una leyenda del mapa cuyas etiquetas de funciones desee mostrar en la presentación preliminar. Si crea un archivo de datos a partir del mapa, se utilizarán estas etiquetas para los valores de datos.

**Comentar.** Introduzca un comentario que describa el mapa o que ofrezca información adicional que pueda ser relevante para sus usuarios, como la fuente del formato shapefile original. El comentario aparecerá en el sistema de gestión del selector de plantillas de tablero.

**Crear un conjunto de datos a partir de las etiquetas de funciones.** Marque esta opción si desea crear un archivo de datos de texto a partir de las etiquetas de funciones que se muestran. Al hacer pulsar en Examinar..., podrá especificar una ubicación así como un nombre de archivo. Si agrega una extensión de *.txt*, se guardará el archivo como un archivo de valores separados por tabulaciones. Si agrega una extensión de *.csv*, se guardará el archivo como un archivo de valores separados por comas. CSV es el formato predeterminado si no se especifica ninguna extensión.

### **Distribución de archivos de mapas**

En el primer paso de la utilidad de conversión de mapas, se selecciona una ubicación en la que guardar el archivo SMZ convertido. Puede que también haya seleccionado añadir el mapa al sistema de gestión del selector de plantillas de tablero. Si elige guardar en el sistema de gestión, el mapa estará disponible en cualquier producto de IBM SPSS que ejecute en el mismo equipo.

Para distribuir el mapa a otros usuarios, necesitará enviarles el archivo SMZ. A continuación, esos usuarios pueden utilizar el sistema de gestión para importar el mapa. Simplemente puede enviar el archivo cuya ubicación ha especificado en el paso 1. Si desea enviar un archivo que se encuentra en el sistema de gestión, primero debe exportarlo:

- ▶ En el selector de plantillas, pulse en Gestionar...
- ▶ Pulse en la pestaña Mapa.
- ▶ Seleccione el mapa que desee distribuir.
- ▶ Pulse en Exportar... y elija una ubicación donde desee guardar el archivo.

Ahora puede enviar el archivo de mapas físico a otros usuarios. Los usuarios necesitarán invertir este proceso e importar el mapa al sistema de gestión.

### **Nodo Gráfico**

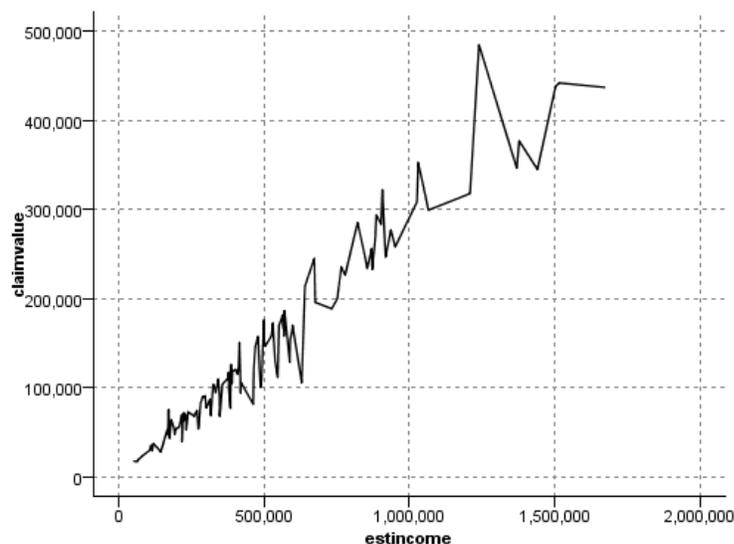
Los nodos Gráfico muestran la relación entre los campos numéricos. Puede crear un gráfico con puntos (también denominado diagrama de dispersión), o puede utilizar líneas. Puede crear tres tipos de gráficos de líneas especificando un valor de Modo para X en el cuadro de diálogo.

**Modo para X = Ordenar**

Al establecer el Modo para X en Ordenar, los datos se clasifican por valores en el campo representado por el eje  $x$ . Así se ejecuta una sola línea de izquierda a derecha en el gráfico. Al utilizar un campo nominal como una superposición, se producen varias líneas de diferentes tonos que se ejecutan de izquierda a derecha en el gráfico.

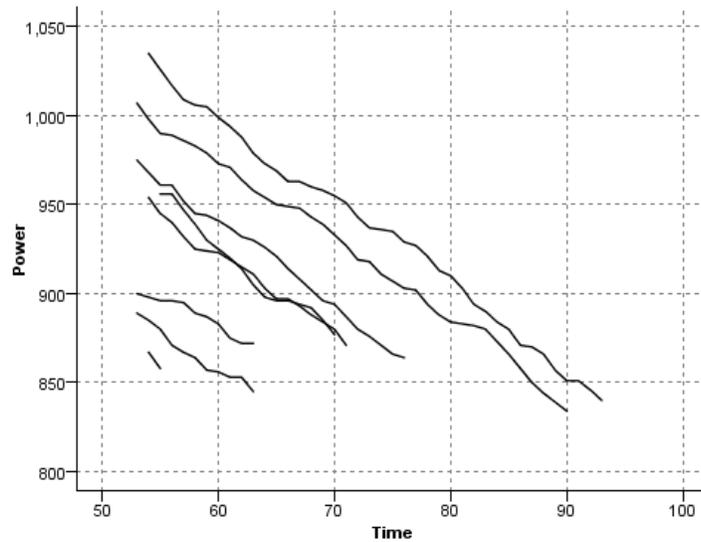
Figura 5-35

Gráfico de líneas con el Modo para X establecido en Ordenar

**Modo para X = Superponer**

Al establecer el Modo para X en Superponer, se crean varios gráficos de línea en el mismo gráfico. Los datos no se ordenan en el caso de un gráfico de superposición; siempre que los valores del eje  $x$  aumenten, los datos se representarán en una sola línea. Si los valores disminuyen, comienza una línea nueva. Por ejemplo, si  $x$  se mueve de 0 a 100, los valores de  $y$  se representarán en una sola línea. Cuando  $x$  cae por debajo de 100, se representa una línea nueva además de la primera. El gráfico terminado puede tener numerosos gráficos que resultan útiles para comparar varias series de valores de  $y$ . El tipo de gráfico es útil para los datos con un componente temporal periódico, como una demanda de electricidad en períodos sucesivos de 24 horas.

Figura 5-36  
Gráfico de líneas con el Modo para X establecido en Superponer

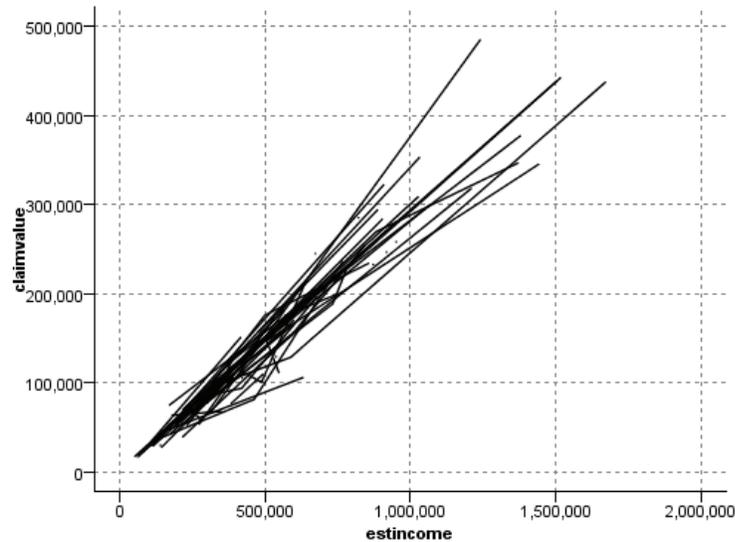


#### **Modo para X = Como se lee**

Al establecer el Modo para X en Como se lee, los valores de  $x$  e  $y$  se representan como se leen desde la fuente de datos. Esta opción es útil para los datos con un componente de serie temporal donde el interés recae sobre tendencias o patrones que dependen del orden de los datos. Puede que sea necesario ordenar los datos antes de crear este tipo de gráfico. También puede ser útil comparar dos gráficos similares con el Modo para X establecido en Ordenar y Como se lee para determinar hasta qué punto el patrón depende del orden.

Figura 5-37

Gráfico de línea mostrado anteriormente como Ordenar, ejecutado de nuevo con el Modo para X establecido en Como se lee

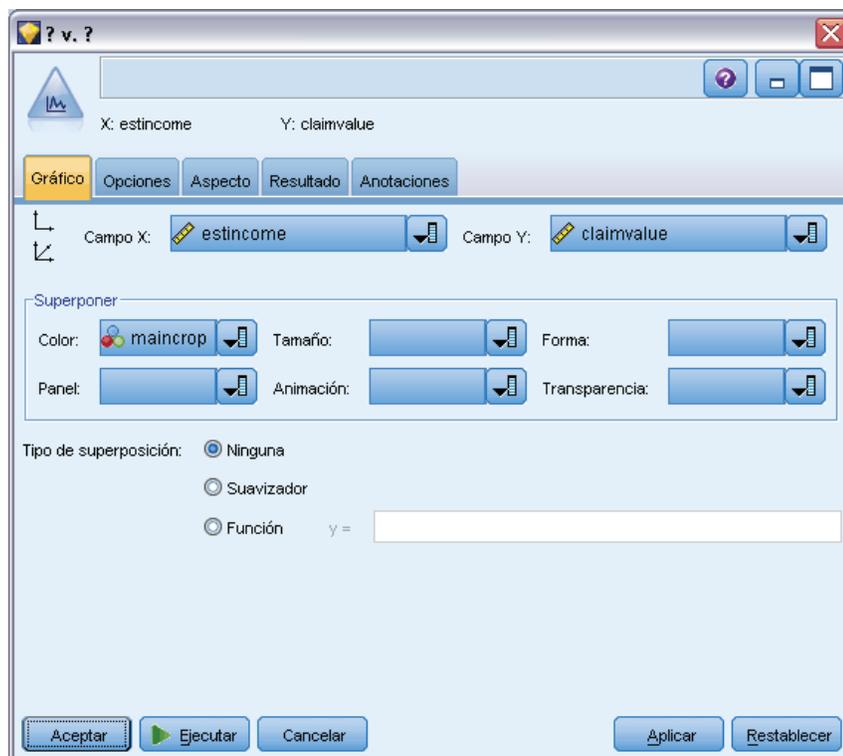


También puede utilizar el nodo Tablero para crear diagramas de dispersión y gráficos de línea. No obstante, tendrá más opciones para seleccionar en este nodo. Si desea obtener más información, consulte el tema [Tipos de visualización de Graphboard preincorporado disponibles](#) el p. 259.

### **Pestaña de nodo Gráfico**

Los gráficos muestran valores de un campo Y frente a los valores de un campo X. A menudo, estos campos corresponden a una variable dependiente y a una variable independiente respectivamente.

Figura 5-38  
Configuración de la pestaña Gráfico de un nodo Gráfico



**Campo X.** Desde la lista, seleccione el campo que desea visualizar en el eje horizontal  $x$ .

**Campo Y.** Desde la lista, seleccione el campo que desea visualizar en el eje vertical  $y$ .

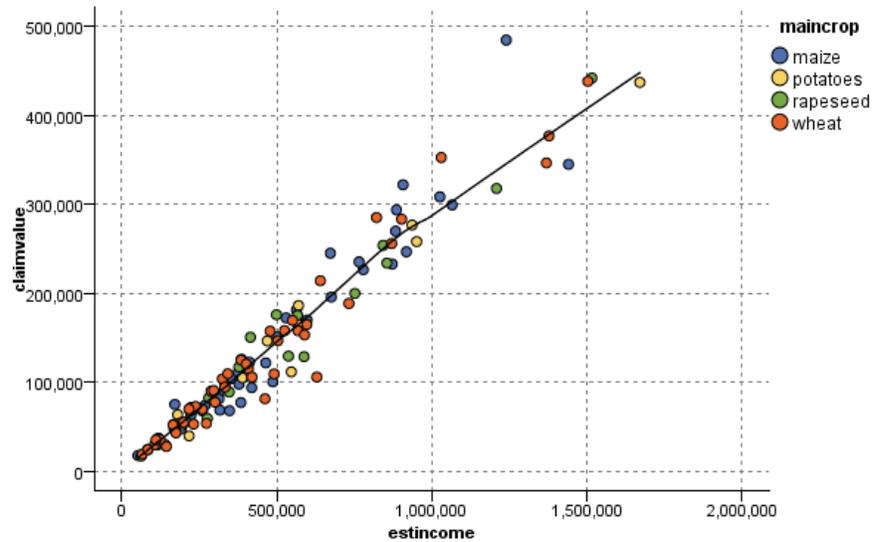
**Campo Z.** Al pulsar en el botón 3D del diagrama, puede seleccionar un campo de la lista que se muestra en el eje  $z$ .

**Superponer.** Existen varias formas de ilustrar las categorías de los valores de datos. Por ejemplo, puede utilizar *maincrop* (cosecha principal) como una superposición de colores para indicar los valores *estincome* y *claimvalue* para la cosecha principal cultivada por solicitantes de reclamaciones. Si desea obtener más información, consulte el tema [Aspecto estético, superposiciones, paneles y animación](#) el p. 244.

**Tipo de superposición.** Determina si se muestra una función superpuesta o suavizamiento. Las opciones suavizamiento y superpuesta se calculan siempre en función de  $y$ .

- **Ninguno.** No se muestra ningún campo de superposición.
- **Suavizamiento.** Muestra una línea de ajuste suavizada calculada con regresión de mínimos cuadrados ponderados localmente (LOESS). Este método calcula de forma eficaz una serie de regresiones, cada una enfocada a una pequeña zona del gráfico. Esto produce una serie de líneas de regresión “locales” que, a continuación, se unen para crear una curva suavizada.

Figura 5-39  
Gráfico con un campo de superposición de suavizamiento LOESS



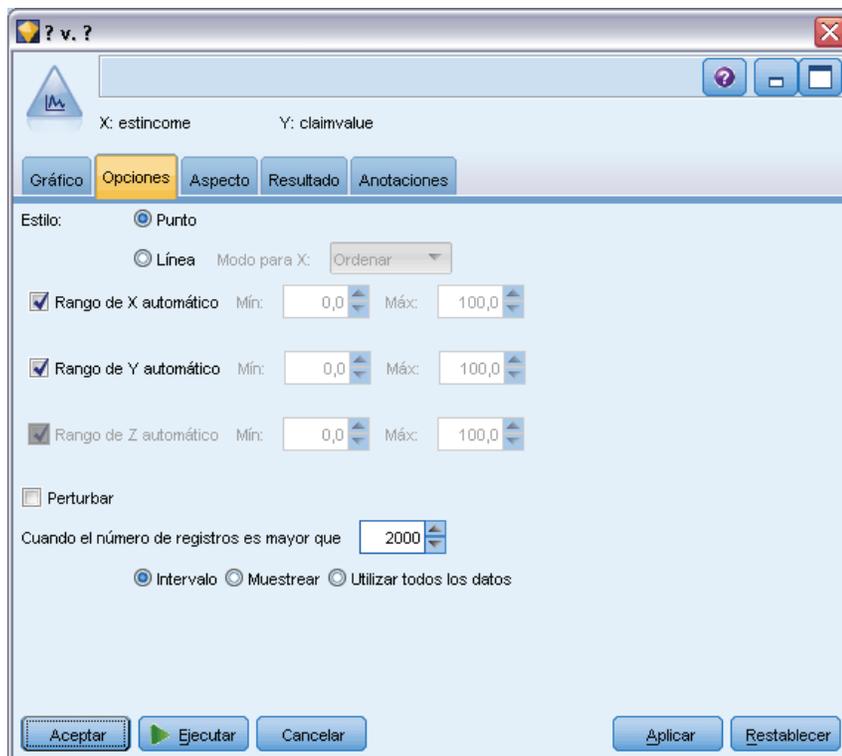
- **Función.** Seleccione esta opción para especificar una función conocida que comparar con los valores reales. Por ejemplo, para comparar valores reales con valores pronosticados, puede representar la función  $y = x$  como una superposición. Especifique una función para  $y =$  en el cuadro de texto. La función por defecto es  $y = x$ , pero puede especificar cualquiera, como una función cuadrática o una expresión arbitraria, en cuanto a  $x$ .

*Nota:* las funciones superpuestas no están disponibles para un panel o gráfico de animación.

Una vez establecidas las opciones de un gráfico, puede ejecutar el gráfico directamente desde el cuadro de diálogo pulsando en Ejecutar. Sin embargo, es posible que desee utilizar la pestaña Opciones para establecer especificaciones adicionales, como intervalos, Modo para X y estilo.

## Pestaña Opciones de Gráfico

Figura 5-40  
Configuración de la pestaña Opciones de un nodo Gráfico



**Estilo.** Permite seleccionar Punto o Línea para el estilo de gráfico. Al seleccionar Línea, se activa el control Modo para X. Si selecciona Punto utilizará un símbolo de adición (+) como forma de los puntos por defecto. Una vez creado el gráfico, podrá cambiar la forma del punto y alterar su tamaño.

**Modo para X.** En el caso de los gráficos de líneas, debe elegir un Modo para X que defina el estilo del gráfico de líneas. Seleccione Ordenar, Superponer o Como se lee. En el caso de Superponer o Como se lee, debe especificar un tamaño del conjunto de datos máximo utilizado para muestrear los primeros  $n$  registros. De lo contrario, se utilizará el valor por defecto de 2,000.

**Rango de X automático.** Seleccione esta opción para utilizar todo el rango de valores de los datos en el eje. Anule la selección para utilizar un subconjunto explícito de valores en función de los valores Mín. y Máx. que especificó. Introduzca los valores o utilice las flechas. Los rangos automáticos se seleccionan por defecto para activar la generación rápida de gráficos.

**Rango de Y automático.** Seleccione esta opción para utilizar todo el rango de valores de los datos en el eje. Anule la selección para utilizar un subconjunto explícito de valores en función de los valores Mín. y Máx. que especificó. Introduzca los valores o utilice las flechas. Los rangos automáticos se seleccionan por defecto para activar la generación rápida de gráficos.

**Rango de Z automático.** Solamente cuando se especifica un gráfico en 3D en la pestaña Gráfico. Seleccione esta opción para utilizar todo el rango de valores de los datos en el eje. Anule la selección para utilizar un subconjunto explícito de valores en función de los valores Mín. y Máx.

que especificó. Introduzca los valores o utilice las flechas. Los rangos automáticos se seleccionan por defecto para activar la generación rápida de gráficos.

**Perturbar.** Conocido también como **agitación**, la perturbación es útil para los gráficos de puntos de un conjunto de datos en el que se repiten muchos valores. Para ver una disposición clara de los valores, puede utilizar la perturbación para distribuir los puntos aleatoriamente alrededor del valor real.

*Nota para usuarios de versiones anteriores de SPSS Modeler:* el valor de perturbación utilizado en un gráfico emplea una métrica distinta en esta versión de IBM® SPSS® Modeler. En versiones anteriores el valor tiene un número real, pero ahora es una proporción del tamaño del marco. Esto significa que es bastante probable que los valores de perturbación de las rutas antiguas sean demasiado altos. Para esta versión, cualquier valor de perturbación distinto de cero se convertirá en el valor 0,2.

**Número máximo de registros que se van a dibujar.** Permite especificar un método para trazar conjuntos de datos grandes. Puede especificar un tamaño de conjunto de datos máximo o utilizar el valor por defecto de 2.000 registros. El rendimiento mejora en el caso de conjuntos de datos grandes al seleccionar las opciones Intervalo o Muestra. Si lo prefiere, puede trazar todos los puntos de datos seleccionando Utilizar todos los datos, pero debe tener en cuenta que esto puede hacer que el rendimiento del software descienda radicalmente.

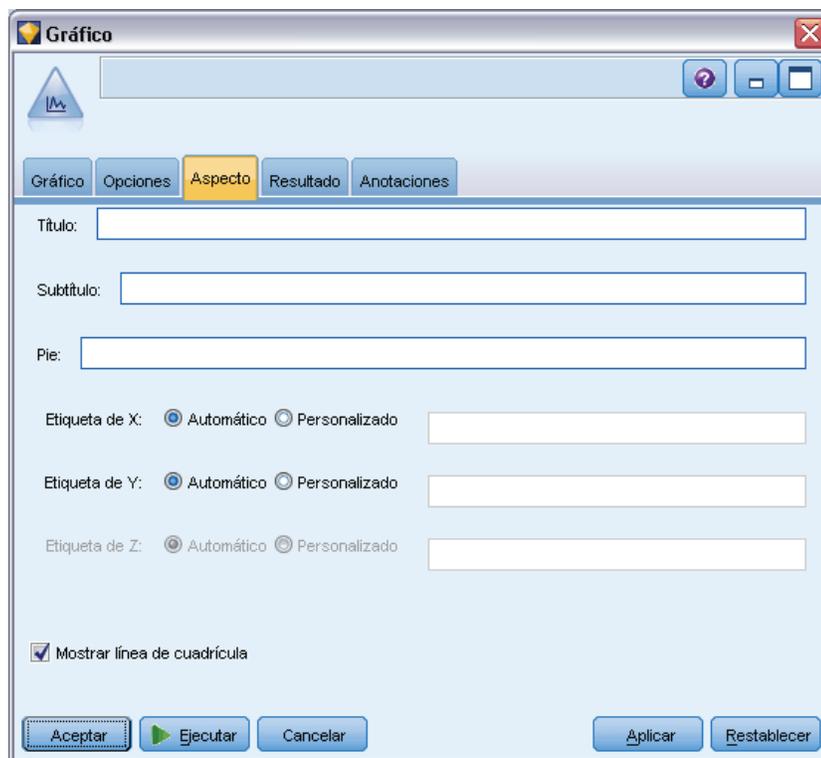
*Nota:* cuando la opción de Modo para X es Superponer o Como se lee, estas opciones están desactivadas y solamente se utilizan los primeros  $n$  registros.

- **Intervalo.** Seleccione esta opción para activar los intervalos cuando el conjunto de datos contiene más registros de los especificados. Los intervalos dividen el gráfico en cuadrículas finas antes del trazado en sí y realizan el recuento del número de puntos que aparecerán de cada una de las casillas de la cuadrícula. En el gráfico final se traza un punto por casilla en el centroide del intervalo (media de todas las ubicaciones de los puntos del intervalo). El tamaño de los símbolos trazados indica el número de puntos de esa región (a menos que haya utilizado el tamaño como una superposición). Al utilizar el centroide y el tamaño para representar el número de puntos, el gráfico en intervalos se convierte en una forma superior de representar conjuntos de datos grandes ya que previene un trazado excesivo en regiones densas (masas de color sin diferenciar) y reduce los artefactos de símbolos (patrones artificiales de densidad). Los artefactos de símbolos se producen cuando ciertos símbolos (en particular el símbolo de adición [+]) choca de tal manera que produce áreas densas que no existen en los datos sin procesar.
- **Muestra.** Seleccione esta opción para muestrear los datos de forma aleatoria en los registros introducidos en el campo de texto. El valor predeterminado es 2,000.

### ***Pestaña Aspecto de Gráfico***

Puede especificar las opciones de aspecto antes de crear gráficos.

Figura 5-41  
Configuración de la pestaña Aspecto de un nodo Gráfico



**Título.** Introduzca el texto que desea para el título del gráfico.

**Subtítulo.** Introduzca el texto que desea para el subtítulo del gráfico.

**Pie.** Introduzca el texto que desea para el pie del gráfico.

**Etiqueta de X.** Permite aceptar la etiqueta de eje  $x$  (horizontal) generada automáticamente o seleccionar Personalizado para especificar una etiqueta.

**Etiqueta de Y.** Permite aceptar la etiqueta del eje  $y$  (vertical) generada automáticamente o seleccionar Personalizado para especificar una etiqueta.

**Etiqueta de Z.** Disponible sólo para gráficos 3D, permite aceptar la etiqueta de eje  $z$  generada automáticamente o seleccionar Personalizado para especificar una etiqueta personalizada.

**Mostrar línea de cuadrícula.** Seleccionada por defecto, esta opción muestra una línea de cuadrícula detrás del gráfico o el diagrama que permite determinar con mayor facilidad los puntos de corte de la región y el tramo. Las líneas de cuadrícula siempre se muestran en blanco a menos que el fondo del gráfico sea blanco; en cuyo caso se muestran en gris.

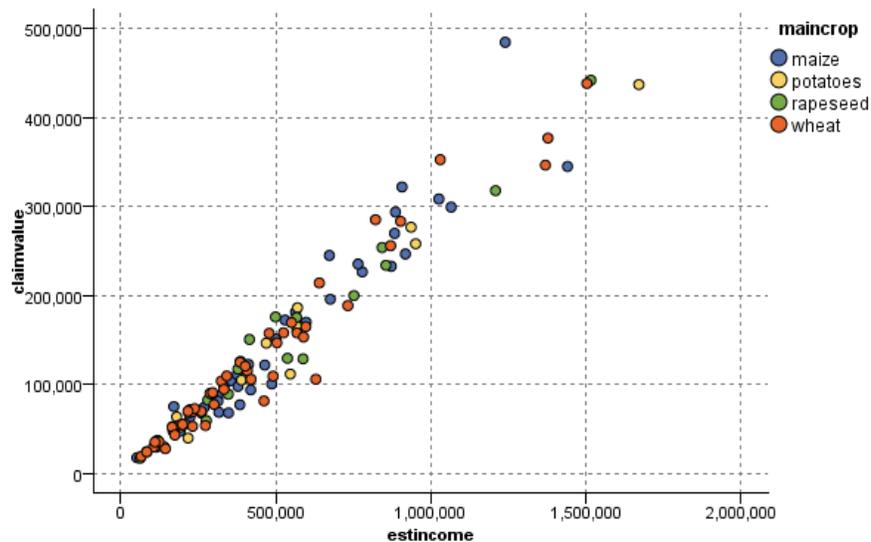
### Utilización de un gráfico

Los gráficos y los gráficos múltiples son esencialmente gráficos de  $X$  frente a  $Y$ . Por ejemplo, si explora un fraude potencial en las solicitudes de subvenciones para agricultura (tal y como se ilustra en el archivo *fraud.str* de la carpeta *Demos*, en el directorio de instalación de IBM®

SPSS® Modeler), puede que desee trazar el ingreso reclamado en la solicitud para compararlo con el ingreso estimado por una red neuronal. Al utilizar una superposición, como el tipo de cosecha, se ilustrará si existe una relación entre las reclamaciones (valor o número) y el tipo de cosecha.

Figura 5-42

Gráfico de la relación entre el ingreso estimado y el valor de reclamación con el tipo de cosecha principal como una superposición



Como los gráficos, los gráficos múltiples y los diagramas de evaluación son representaciones bidimensionales de  $Y$  frente a  $X$ , es fácil interactuar con ellos definiendo regiones, marcando elementos o incluso dibujando tramos. También puede generar nodos para los datos representados por dichos tramos, elemento o regiones. Si desea obtener más información, consulte el tema [Exploración de gráficos](#) el p. 355.

## Nodo Distribución

Los gráficos o tablas de distribución muestran las instancias de valores simbólicos (no numéricos), como un género o un tipo de hipoteca, en un conjunto de datos. El uso habitual del nodo de distribución consiste en mostrar los desequilibrios de los datos que pueden rectificarse mediante el nodo Equilibrar antes de crear un modelo. Puede generar automáticamente un nodo Equilibrar mediante el menú Generar en la ventana de gráfico o tabla de distribución.

También puede utilizar el nodo Tablero para crear gráficos de barras de recuentos. No obstante, tendrá más opciones para seleccionar en este nodo. Si desea obtener más información, consulte el tema [Tipos de visualización de Graphboard preincorporado disponibles](#) el p. 259.

*Nota:* para mostrar las instancias de valores numéricos, se debe utilizar un nodo Histograma.

## Pestaña Gráfico de Distribución

Figura 5-43

Configuración de la pestaña Gráfico de un nodo Distribución



**Gráfico.** Permite seleccionar el tipo de distribución. Seleccione Campos seleccionados para mostrar la distribución del campo seleccionado. Seleccione Todas las marcas (valores verdaderos) para mostrar la distribución de los valores verdaderos de los campos de marcas del conjunto de datos.

**Campo.** Permite seleccionar un campo nominal o marca para el que se va a mostrar la distribución de los valores. Solamente aparecen en la lista los campos que no se han establecido específicamente como numéricos.

**Superponer.** Permite seleccionar un campo nominal o marca para utilizarlo como superposición de colores, ilustrando la distribución de sus valores dentro de cada valor del campo seleccionado. Por ejemplo, puede utilizar la respuesta de la campaña de marketing (*pep*) como una superposición para el número de hijos (*hijos*) con objeto de ilustrar la respuesta por tamaño de familia. Si desea obtener más información, consulte el tema [Aspecto estético, superposiciones, paneles y animación](#) el p. 244.

**Normalizar por color.** Seleccione esta opción para escalar barras de tal forma que todas las barras ocupen el ancho total del gráfico. Los valores de superposición corresponden a una proporción de cada barra, lo que simplifica las comparaciones entre las categorías.

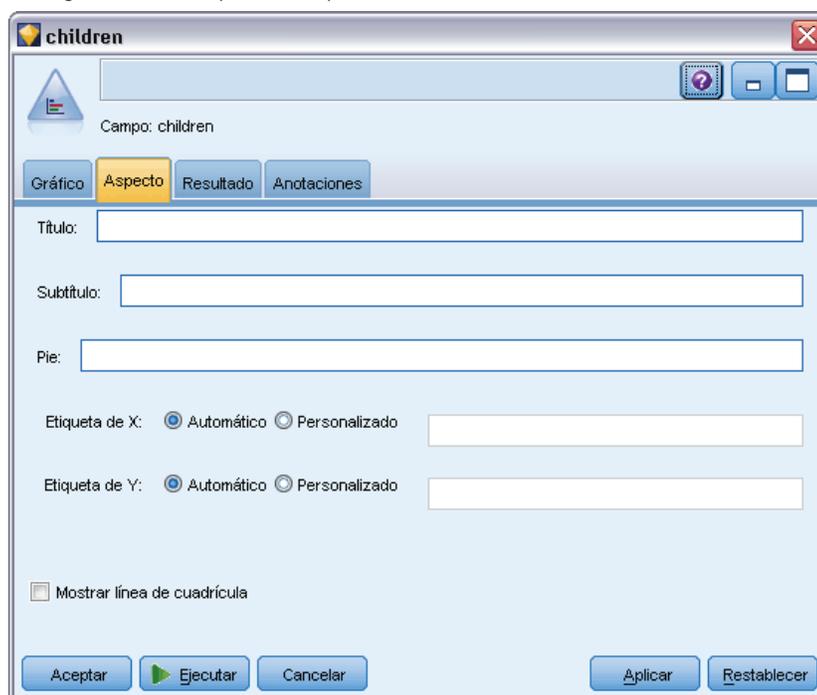
**Ordenar.** Permite seleccionar el método utilizado para mostrar valores en el gráfico de distribución. Seleccione Alfabético para utilizar el orden alfabético o Por recuento para enumerar los valores en orden de instancia decreciente.

**Escala proporcional.** Seleccione esta opción para escalar la distribución de valores de tal forma que el valor con el mayor recuento ocupe la anchura total del gráfico. El resto de barras se escalan en función de este valor. Al anular la selección de esta opción, las barras se escalan en función de los recuentos totales de cada valor.

## Pestaña Aspecto de Distribución

Puede especificar las opciones de aspecto antes de crear gráficos.

Figura 5-44  
Configuración de la pestaña Apariencia



**Título.** Introduzca el texto que desea para el título del gráfico.

**Subtítulo.** Introduzca el texto que desea para el subtítulo del gráfico.

**Pie.** Introduzca el texto que desea para el pie del gráfico.

**Etiqueta de X.** Permite aceptar la etiqueta de eje  $x$  (horizontal) generada automáticamente o seleccionar Personalizado para especificar una etiqueta.

**Etiqueta de Y.** Permite aceptar la etiqueta del eje  $y$  (vertical) generada automáticamente o seleccionar Personalizado para especificar una etiqueta.

**Mostrar línea de cuadrícula.** Seleccionada por defecto, esta opción muestra una línea de cuadrícula detrás del gráfico o el diagrama que permite determinar con mayor facilidad los puntos de corte de la región y el tramo. Las líneas de cuadrícula siempre se muestran en blanco a menos que el fondo del gráfico sea blanco; en cuyo caso se muestran en gris.

## Utilización de un nodo Distribución

Los nodos Distribución se utilizan para mostrar la distribución de valores simbólicos en un conjunto de datos. Se suelen utilizar antes que los nodos de manipulación para explorar los datos y corregir cualquier desequilibrio. Por ejemplo, si las instancias de los encuestados sin hijos se suceden con mucha más frecuencia que otros tipos de encuestados, tal vez desee reducir estas instancias para que se pueda generar una regla más útil en posteriores operaciones de minería de datos. Los nodos Distribución ayudan a examinar y tomar decisiones acerca de tales desequilibrios.

El nodo Distribución es extraño porque genera tanto un gráfico como una tabla para analizar los datos.

Figura 5-45

Gráficos de distribución con la proporción de número de personas con hijos con respuesta a una campaña de marketing

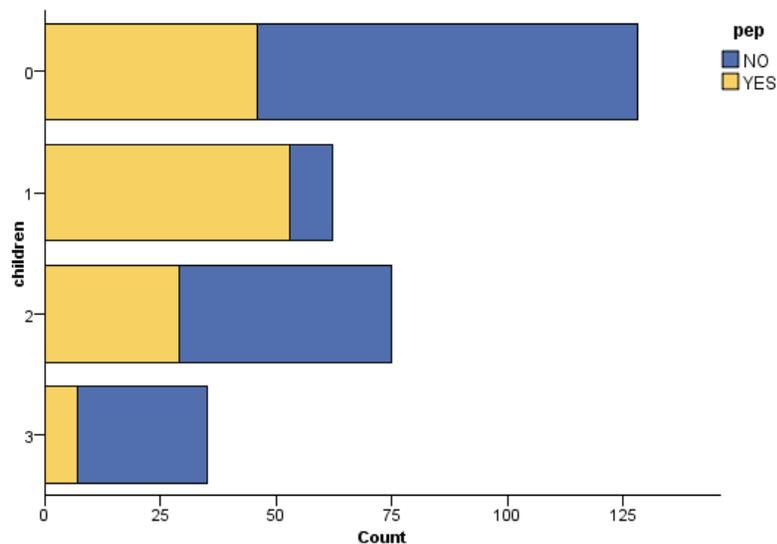
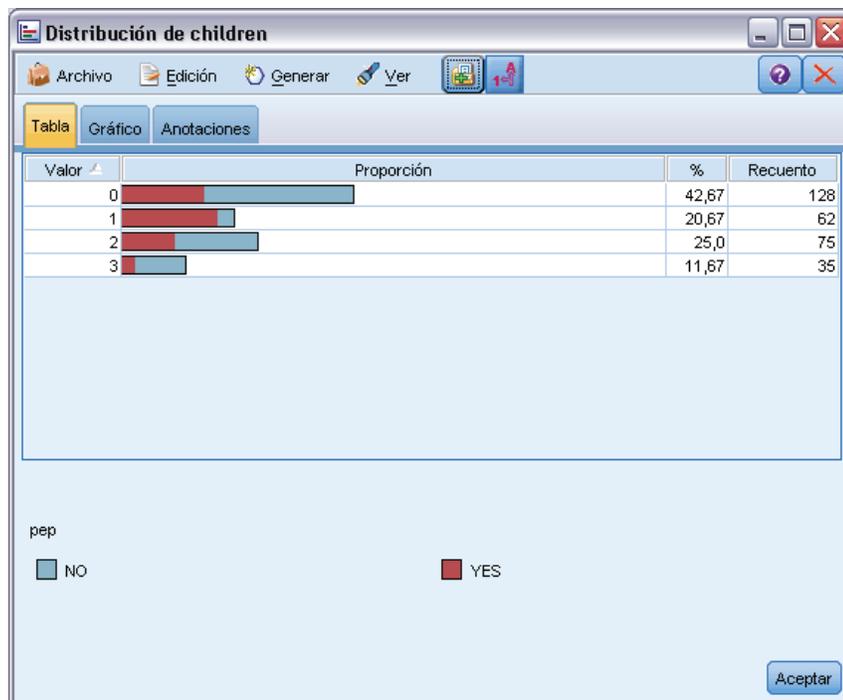


Figura 5-46

Tabla de distribución con la proporción de personas con o sin hijos con respuesta a una campaña de marketing



Una vez creado un gráfico y una tabla de distribución y examinados los resultados, puede utilizar las opciones de los menús para agrupar valores, copiar valores y generar un número de nodos para la preparación de datos. Además, puede copiar o exportar la información del gráfico y la tabla para utilizarla en otras aplicaciones como MS Word o MS PowerPoint. Si desea obtener más información, consulte el tema [Impresión, almacenamiento, copia y exportación de gráficos](#) el p. 389.

#### Para seleccionar y copiar valores desde una tabla de distribución

- ▶ Mantenga pulsado el botón del ratón mientras lo arrastra sobre las filas para seleccionar un conjunto de valores. También puede utilizar la opción Seleccionar todo del menú Edición.
- ▶ En el menú Edición, seleccione Copiar tabla o Copiar tabla (inc. nombres de campos).
- ▶ Péguelo en el portapapeles o en la aplicación deseada.

*Nota:* las barras no se copian directamente. En su lugar, se copian los valores de la tabla. Esto implica que los valores superpuestos no se mostrarán en la tabla copiada.

#### Para agrupar valores desde una tabla de distribución

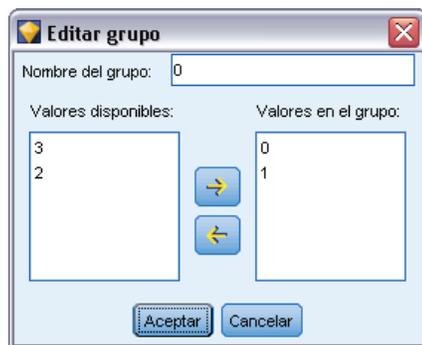
- ▶ Seleccione valores que desea agrupar manteniendo pulsada la tecla Ctrl mientras pulsa con el ratón en los valores.
- ▶ En el menú Edición, elija Agrupar.

*Nota:* Al agrupar y desagrupar valores, el gráfico de la pestaña Gráfico se vuelve a generar de forma automática para mostrar los cambios.

También puede:

- Desagrupar valores seleccionando el nombre del grupo en la lista de distribución y eligiendo Desagrupar en el menú Edición.
- Editar grupos seleccionando el nombre del grupo de la lista de distribución y eligiendo Editar grupo en el menú Edición. Se abrirá un cuadro de diálogo donde se pueden cambiar los valores hacia y desde el grupo.

Figura 5-47  
Cuadro de diálogo Editar grupo



### **Opciones del menú Generar**

Puede utilizar opciones del menú Generar para seleccionar un subconjunto de datos, derivar un campo de marcas, reagrupar valores, reclasificar valores o equilibrar los datos de un gráfico o una tabla. Estas operaciones generan un nodo de preparación de datos y lo colocan en el lienzo de rutas. Para utilizar el nodo de generación, conéctelo a una ruta existente. Si desea obtener más información, consulte el tema [Generación de nodos a partir de gráficos](#) el p. 364.

## **Nodo Histograma**

Los nodos Histograma muestran las instancias de valores de los campos numéricos. Se suelen utilizar para explorar los datos antes de las manipulaciones y la generación de modelos. Al igual que con el nodo Distribución, con frecuencia los nodos de histogramas se utilizan para detectar desequilibrios en los datos. Al mismo tiempo puede utilizar el nodo Tablero para producir un histograma, tiene más opciones para seleccionar a partir de este nodo. Si desea obtener más información, consulte el tema [Tipos de visualización de Graphboard preincorporado disponibles](#) el p. 259.

*Nota:* Para mostrar las instancias de valores para campos simbólicos, se debe utilizar un nodo Distribución.

## Pestaña Gráfico de Histograma

Figura 5-48  
Configuración de la pestaña Gráfico de un nodo Histograma

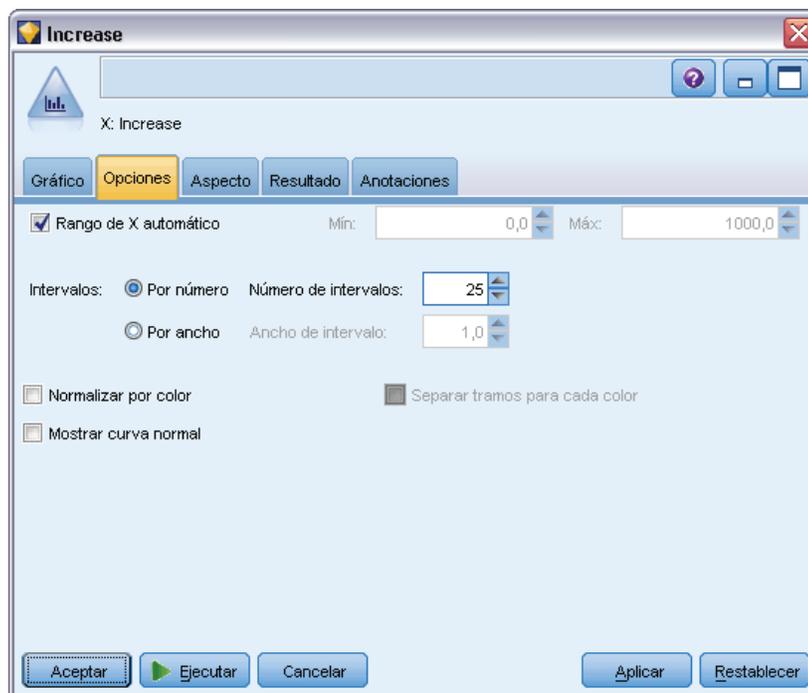


**Campo.** Permite seleccionar un campo numérico para el que se va a mostrar la distribución de los valores. Solamente aparecen en la lista los campos que no se han definido específicamente como simbólicos (categóricos).

**Superponer.** Permite seleccionar un campo simbólico con objeto de mostrar categorías de valores para el campo especificado. Al seleccionar un campo de superposición, el histograma se convierte en un gráfico apilado donde los colores representan distintas categorías del campo de superposición. Si utiliza el nodo Histograma, hay tres tipos de superposiciones: color, panel y animación. Si desea obtener más información, consulte el tema [Aspecto estético, superposiciones, paneles y animación](#) el p. 244.

## Pestaña Opciones de Histograma

Figura 5-49  
Configuración de la pestaña Opciones de un nodo Histograma



**Rango de X automático.** Seleccione esta opción para utilizar todo el rango de valores de los datos en el eje. Anule la selección para utilizar un subconjunto explícito de valores en función de los valores Mín. y Máx. que especificó. Introduzca los valores o utilice las flechas. Los rangos automáticos se seleccionan por defecto para activar la generación rápida de gráficos.

**Intervalos.** Seleccione Por número o Por ancho.

- Seleccione Por número para mostrar un número fijo de barras de colección cuya anchura depende del rango especificado anteriormente y el número de buckets especificado. Indique el número de intervalos que se utilizarán en el gráfico en la opción Número de intervalos. Utilice las flechas para ajustar el número.
- Seleccione Por ancho para crear un gráfico con barras de un ancho fijo. El número de intervalos depende del ancho especificado y el rango de valores. Indique la anchura de las barras en la opción Ancho de intervalo.

**Normalizar por color.** Seleccione esta opción para ajustar todas las barras a la misma altura, mostrando los valores superpuestos como un porcentaje del número total de casos de cada barra.

**Mostrar curva normal.** Seleccione esta opción para añadir una curva normal al gráfico, mostrando la media y la varianza de los datos.

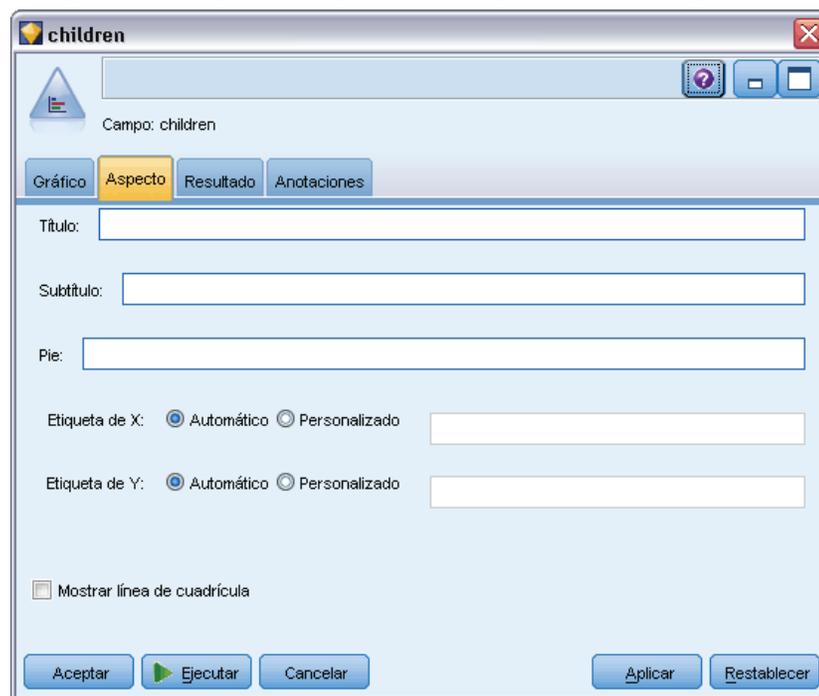
**Separar tramos para cada color** Seleccione esta opción para mostrar cada valor superpuesto como un tramo independiente en el gráfico.

## Pestaña Aspecto de Histograma

Puede especificar las opciones de aspecto antes de crear gráficos.

Figura 5-50

Configuración de la pestaña Aspecto de la mayoría de nodos de gráficos



**Título.** Introduzca el texto que desea para el título del gráfico.

**Subtítulo.** Introduzca el texto que desea para el subtítulo del gráfico.

**Pie.** Introduzca el texto que desea para el pie del gráfico.

**Etiqueta de X.** Permite aceptar la etiqueta de eje  $x$  (horizontal) generada automáticamente o seleccionar Personalizado para especificar una etiqueta.

**Etiqueta de Y.** Permite aceptar la etiqueta del eje  $y$  (vertical) generada automáticamente o seleccionar Personalizado para especificar una etiqueta.

**Mostrar línea de cuadrícula.** Seleccionada por defecto, esta opción muestra una línea de cuadrícula detrás del gráfico o el diagrama que permite determinar con mayor facilidad los puntos de corte de la región y el tramo. Las líneas de cuadrícula siempre se muestran en blanco a menos que el fondo del gráfico sea blanco; en cuyo caso se muestran en gris.

## Uso de histogramas

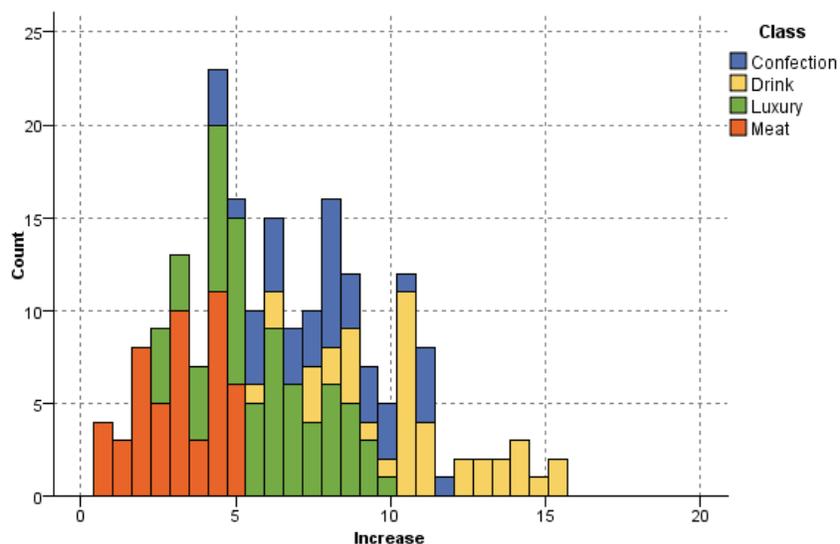
Los histogramas muestran la distribución de valores en un campo numérico cuyos valores se extienden por el eje  $x$ . Los histogramas funcionan de forma similar a los gráficos de colecciones. Las colecciones muestran la distribución de los valores de un campo numérico *relativo a los valores de otro*, en lugar de las instancias de los valores de un solo campo.

Una vez haya creado un gráfico, puede examinar los resultados y definir los tramos para dividir los valores a lo largo del eje  $x$  o definir regiones. También puede marcar elementos dentro del gráfico. Si desea obtener más información, consulte el tema [Exploración de gráficos](#) el p. 355.

Puede utilizar opciones del menú Generar para crear nodos Equilibrar, Seleccionar o Derivar mediante los datos de gráfico o más específicamente dentro de tramos, regiones o elementos marcados. Este tipo de gráfico se suele utilizar antes de los nodos de manipulación para explorar los datos y corregir cualquier desequilibrio al generar un nodo Equilibrar desde el gráfico que desea utilizar en la ruta. También puede generar un nodo de derivación como marca para añadir un campo que muestre el tramo donde se encuentra cada registro o un nodo Seleccionar para seleccionar todos los registros de un rango de valores o un conjunto determinado. Dichas operaciones ayudan a centrarse en un subconjunto de datos determinado para su mejor exploración. Si desea obtener más información, consulte el tema [Generación de nodos a partir de gráficos](#) el p. 364.

Figura 5-51

*Histograma con la distribución de compras incrementadas por categoría debido a la promoción*

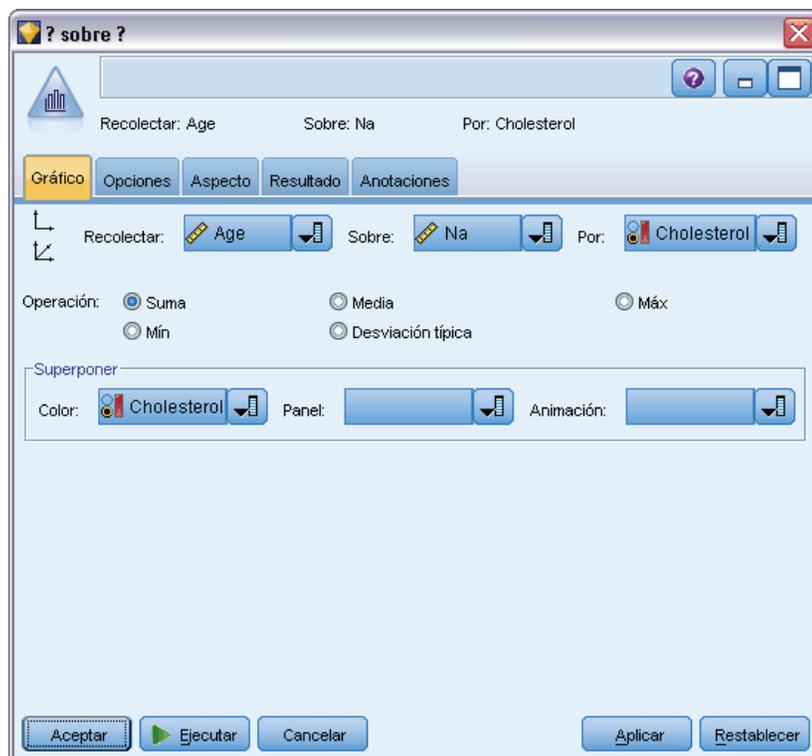


## Nodo Colección

Las colecciones son similares a los histogramas salvo por el hecho de que las colecciones muestran la distribución de los valores de un campo numérico relativo a los valores de otro, en lugar de las instancias de los valores de un solo campo. Las colecciones son útiles para ilustrar una variable o un campo cuyos valores cambian con el tiempo. Con los gráficos 3D también puede incluir un eje simbólico que muestra las distribuciones por categoría. Las colecciones bidimensionales se muestran como gráficos de barras apilados, con superposiciones donde se utilicen. Si desea obtener más información, consulte el tema [Aspecto estético, superposiciones, paneles y animación](#) el p. 244.

## Pestaña Gráfico de Colección

Figura 5-52  
Configuración de la pestaña Gráfico de un nodo Colección



**Recolectar.** Permite seleccionar un campo cuyos valores se recopilarán y mostrarán en el rango de valores para el campo especificado en Sobre. Solamente se enumeran los campos definidos como simbólicos.

**Sobre.** Permite seleccionar un campo cuyos valores se utilizan para mostrar el campo especificado en Recolectar.

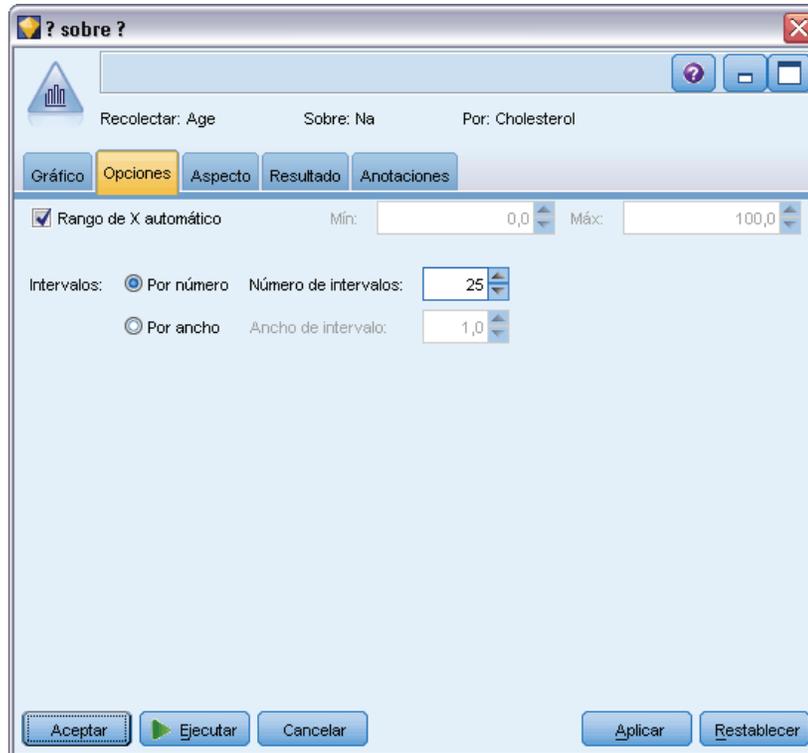
**Por.** Activada al crear un gráfico 3D, esta opción permite seleccionar un campo nominal o marca utilizado para mostrar el campo de colección por categorías.

**Operación.** Permite seleccionar lo que representa cada barra del gráfico de colección. Las opciones son Suma, Media, Máx, Mín y Desviación típica.

**Superponer.** Permite seleccionar un campo simbólico con objeto de mostrar categorías de valores para el campo especificado. Al seleccionar un campo de superposición, la colección se convierte y se crean múltiples barras de distintos colores para cada categoría. Este nodo tiene tres tipos de superposiciones: color, panel y animación. Si desea obtener más información, consulte el tema [Aspecto estético, superposiciones, paneles y animación](#) el p. 244.

## Pestaña Opciones de Colección

Figura 5-53  
Configuración de la pestaña Opciones de un nodo Colección



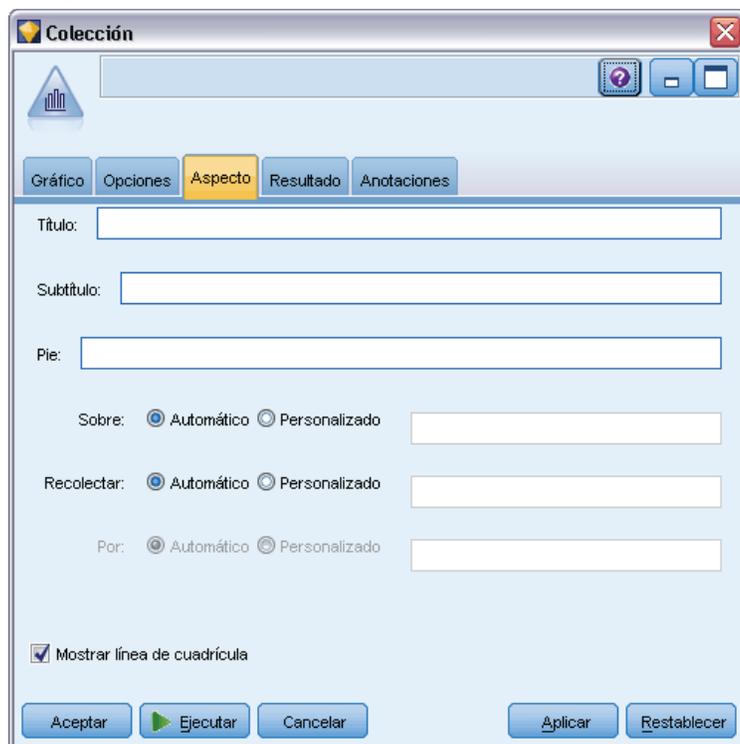
**Rango de X automático.** Seleccione esta opción para utilizar todo el rango de valores de los datos en el eje. Anule la selección para utilizar un subconjunto explícito de valores en función de los valores Mín. y Máx. que especificó. Introduzca los valores o utilice las flechas. Los rangos automáticos se seleccionan por defecto para activar la generación rápida de gráficos.

**Intervalos.** Seleccione Por número o Por ancho.

- Seleccione Por número para mostrar un número fijo de barras de colección cuya anchura depende del rango especificado anteriormente y el número de buckets especificado. Indique el número de intervalos que se utilizarán en el gráfico en la opción Número de intervalos. Utilice las flechas para ajustar el número.
- Seleccione Por ancho para crear un gráfico con barras de un ancho fijo. El número de intervalos depende del ancho especificado y el rango de valores. Indique la anchura de las barras en la opción Ancho de intervalo.

## Pestaña Aspecto de Colección

Figura 5-54  
Configuración de la pestaña Aspecto de un nodo Colección



Puede especificar las opciones de aspecto antes de crear gráficos.

**Título.** Introduzca el texto que desea para el título del gráfico.

**Subtítulo.** Introduzca el texto que desea para el subtítulo del gráfico.

**Pie.** Introduzca el texto que desea para el pie del gráfico.

**Etiqueta Sobre.** Permite aceptar la etiqueta generada automáticamente o seleccionar Personalizado para especificar una etiqueta.

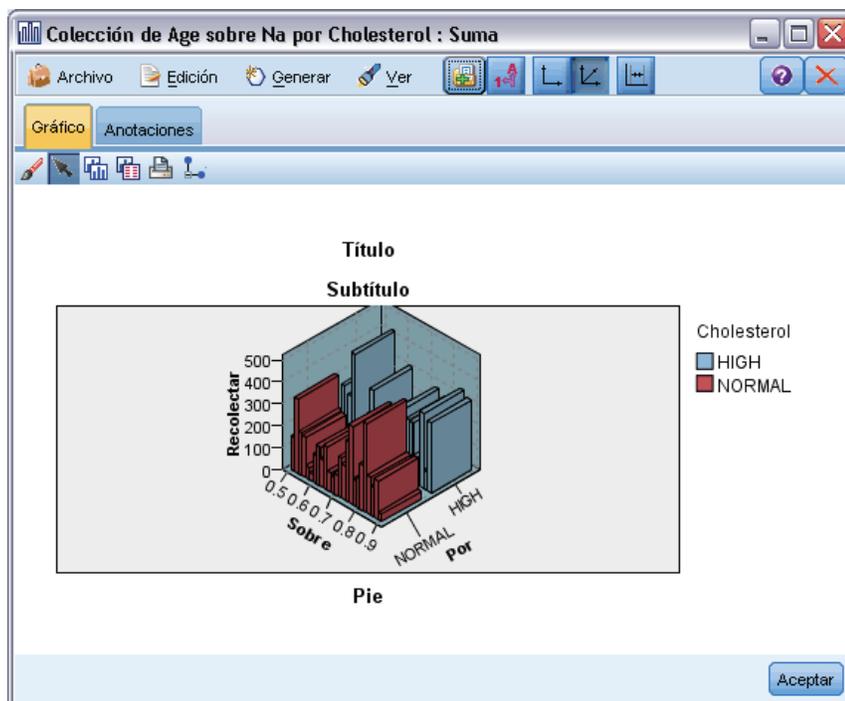
**Etiqueta Recolectar.** Permite aceptar la etiqueta generada automáticamente o seleccionar Personalizado para especificar una etiqueta.

**Etiqueta Por.** Permite aceptar la etiqueta generada automáticamente o seleccionar Personalizado para especificar una etiqueta.

**Mostrar línea de cuadrícula.** Seleccionada por defecto, esta opción muestra una línea de cuadrícula detrás del gráfico o el diagrama que permite determinar con mayor facilidad los puntos de corte de la región y el tramo. Las líneas de cuadrícula siempre se muestran en blanco a menos que el fondo del gráfico sea blanco; en cuyo caso se muestran en gris.

El siguiente ejemplo muestra dónde se colocan las opciones de apariencia en una versión 3D del gráfico.

Figura 5-55  
Posición de las opciones de apariencia de gráfico en gráfico de colección 3D



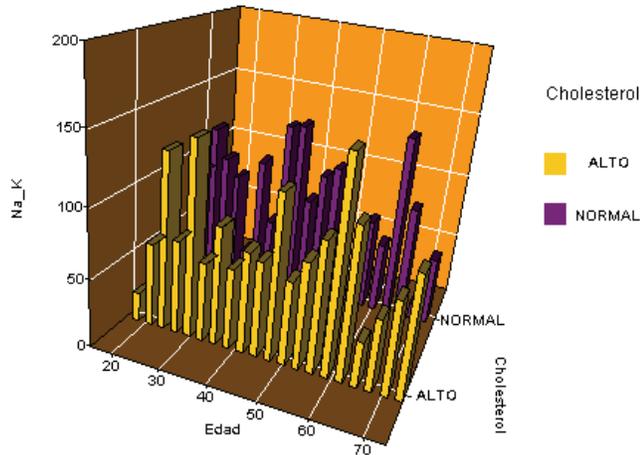
### Utilización de un gráfico de colección

Las colecciones muestran la distribución de los valores de un campo numérico *relativo a los valores de otro*, en lugar de las instancias de los valores de un solo campo. Los histogramas funcionan de forma similar a los gráficos de colecciones. Los histogramas muestran la distribución de valores en un campo numérico cuyos valores se extienden por el eje  $x$ .

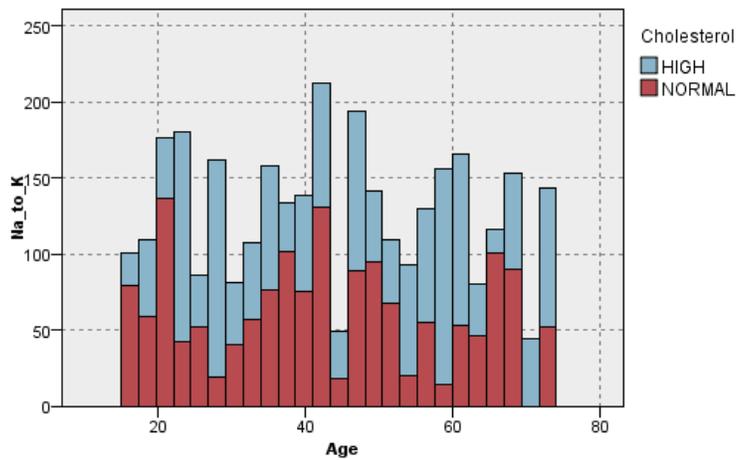
Una vez haya creado un gráfico, puede examinar los resultados y definir los tramos para dividir los valores a lo largo del eje  $x$  o definir regiones. También puede marcar elementos dentro del gráfico. Si desea obtener más información, consulte el tema [Exploración de gráficos](#) el p. 355.

Puede utilizar opciones del menú Generar para crear nodos Equilibrar, Seleccionar o Derivar mediante los datos de gráfico o más específicamente dentro de tramos, regiones o elementos marcados. Este tipo de gráfico se suele utilizar antes de los nodos de manipulación para explorar los datos y corregir cualquier desequilibrio al generar un nodo Equilibrar desde el gráfico que desea utilizar en la ruta. También puede generar un nodo de derivación como marca para añadir un campo que muestre el tramo donde se encuentra cada registro o un nodo Seleccionar para seleccionar todos los registros de un rango de valores o un conjunto determinado. Dichas operaciones ayudan a centrarse en un subconjunto de datos determinado para su mejor exploración. Si desea obtener más información, consulte el tema [Generación de nodos a partir de gráficos](#) el p. 364.

**Figura 5-56**  
Gráficos de colección 3D mostrando la suma de Na\_K sobre Age para niveles de colesterol altos y normales



**Figura 5-57**  
Gráfico de colección sin eje z pero con el colesterol como superposición de colores

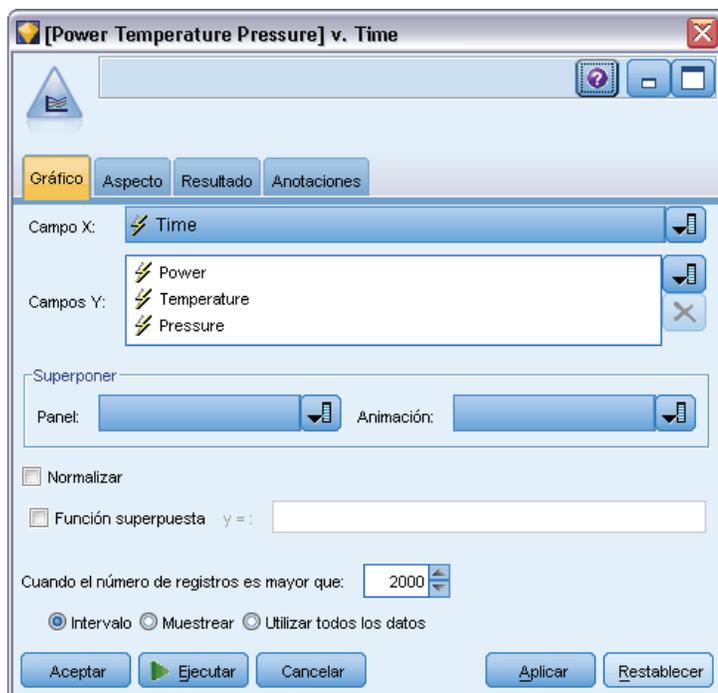


## Nodo G. múltiple

Un gráfico múltiple es un tipo especial de gráfico que muestra varios campos  $Y$  sobre un único campo  $X$ . Los campos  $Y$  están trazados como líneas coloreadas y cada uno equivale a un nodo Gráfico con el estilo establecido en Línea y Modo para  $X$  establecido en Ordenar. Los gráficos múltiples son útiles cuando se tienen datos de una secuencia temporal y se desea explorar la fluctuación de diversas variables durante un período de tiempo.

## Pestaña Gráfico de G. múltiple

Figura 5-58  
Configuración de la pestaña Gráfico de un nodo G. múltiple



**Campo X.** Desde la lista, seleccione el campo que desea visualizar en el eje horizontal x.

**Campos Y.** Seleccione uno o varios campos de la lista para mostrarlo en el rango de los valores del campo X. Utilice el botón del selector de campos para seleccionar múltiples campos. Pulse en el botón de eliminación para eliminar campos de la lista.

**Superponer.** Existen varias formas de ilustrar las categorías de los valores de datos. Por ejemplo, puede utilizar una superposición de animaciones para mostrar múltiples gráficos para cada valor de los datos. Es útil para los conjuntos que contienen más de 10 categorías. Cuando se utiliza para conjuntos de más de 15 categorías, podrá apreciar una disminución del rendimiento. Si desea obtener más información, consulte el tema [Aspecto estético, superposiciones, paneles y animación](#) el p. 244.

**Normalizar.** Seleccione esta opción para escalar todos los valores de Y para el rango de 0 a 1 al mostrarlos en el gráfico. La normalización ayuda a explorar la relación entre líneas que, de lo contrario, estaría oculta debido las diferencias existentes en el rango de valores de cada serie, y se recomienda cuando se representan varias líneas en el mismo gráfico, o cuando se contrastan representaciones en paneles a cada lado. (La normalización no es necesaria cuando todos los valores de datos se encuentran en un rango parecido.)

Figura 5-59

Gráfico múltiple estándar que muestra la fluctuación de la planta de energía durante un período de tiempo (tenga en cuenta que sin la normalización, resulta imposible ver el gráfico Pressure)

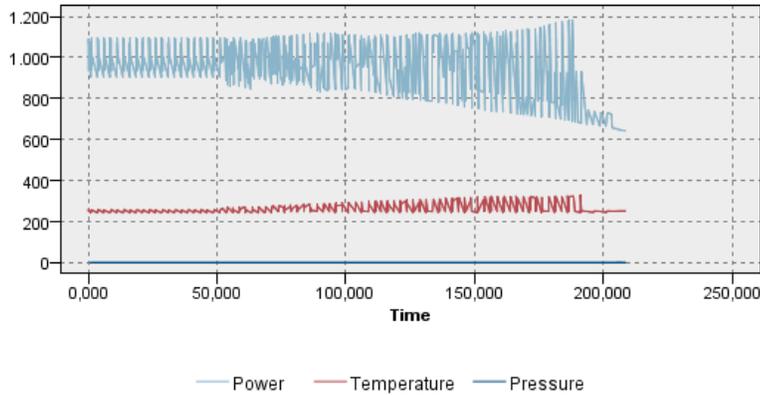
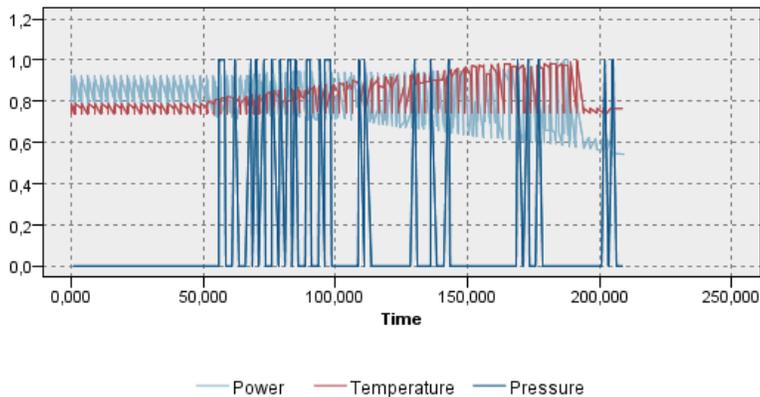


Figura 5-60

Gráficos múltiples normalizados que muestran un gráfico para Pressure



**Función superpuesta.** Seleccione esta opción para especificar una función conocida que comparar con los valores reales. Por ejemplo, para comparar valores reales con valores pronosticados, puede representar la función  $y = x$  como una superposición. Especifique una función para  $y =$  en el cuadro de texto. La función por defecto es  $y = x$ , pero puede especificar cualquiera, como una función cuadrática o una expresión arbitraria, en cuanto a  $x$ .

*Nota:* las funciones superpuestas no están disponibles para un panel o gráfico de animación.

**Cuando el número de registros es mayor que.** Permite especificar un método para trazar conjuntos de datos grandes. Puede especificar un tamaño de conjunto de datos máximo o utilizar el valor por defecto de 2.000 puntos. El rendimiento mejora en el caso de conjuntos de datos grandes al seleccionar las opciones Intervalo o Muestra. Si lo prefiere, puede trazar todos los puntos de datos seleccionando Utilizar todos los datos, pero debe tener en cuenta que esto puede hacer que el rendimiento del software descienda radicalmente.

*Nota:* cuando la opción de Modo para X es Superponer o Como se lee, estas opciones están desactivadas y solamente se utilizan los primeros  $n$  registros.

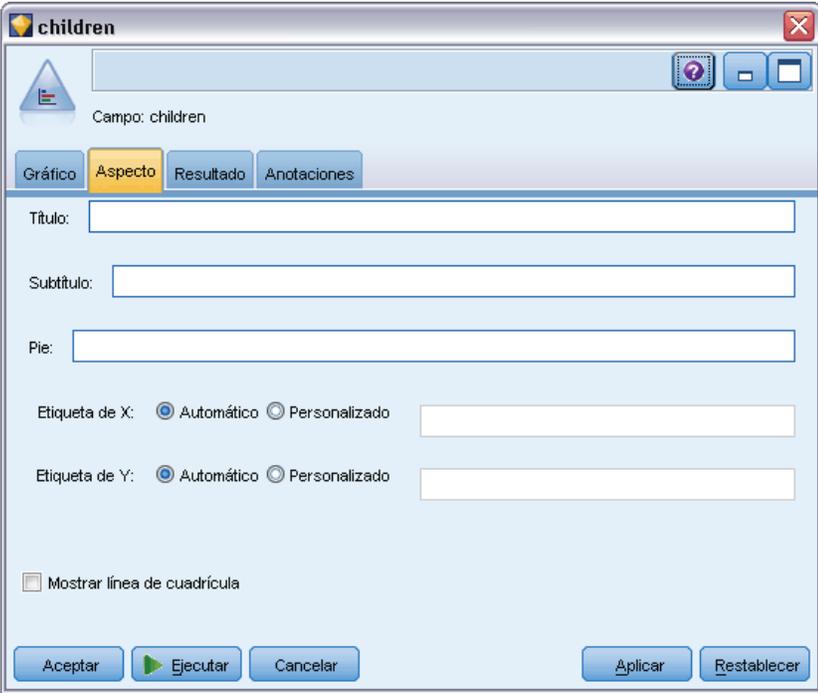
- **Intervalo.** Seleccione esta opción para activar los intervalos cuando el conjunto de datos contiene más registros de los especificados. Los intervalos dividen el gráfico en cuadrículas finas antes de realizar el trazado y realiza el recuento del número de conexiones que aparecerán en cada una de las casillas de la cuadrícula. En el gráfico final se utiliza una conexión por casilla en el centroide del intervalo (media de todos los puntos de conexión del intervalo).
- **Muestra.** Seleccione esta opción para muestrear los datos de forma aleatoria en el número de registros especificado.

### ***Pestaña Aspecto de G. múltiple***

Puede especificar las opciones de aspecto antes de crear gráficos.

Figura 5-61

Configuración de la pestaña Aspecto de la mayoría de nodos de gráficos



The screenshot shows a window titled 'children' with a standard Windows-style title bar. Below the title bar, there is a toolbar with a question mark icon and a close button. The main area contains a tabbed interface with four tabs: 'Gráfico', 'Aspecto' (which is selected and highlighted in yellow), 'Resultado', and 'Anotaciones'. Under the 'Aspecto' tab, there are several input fields and options:

- 'Título:' followed by a text input field.
- 'Subtítulo:' followed by a text input field.
- 'Pie:' followed by a text input field.
- 'Etiqueta de X:' with two radio buttons: 'Automático' (selected) and 'Personalizado', followed by a text input field.
- 'Etiqueta de Y:' with two radio buttons: 'Automático' (selected) and 'Personalizado', followed by a text input field.
- A checkbox labeled 'Mostrar línea de cuadrícula' which is currently unchecked.

At the bottom of the window, there are five buttons: 'Aceptar', 'Ejecutar' (with a play icon), 'Cancelar', 'Aplicar', and 'Restablecer'.

**Título.** Introduzca el texto que desea para el título del gráfico.

**Subtítulo.** Introduzca el texto que desea para el subtítulo del gráfico.

**Pie.** Introduzca el texto que desea para el pie del gráfico.

**Etiqueta de X.** Permite aceptar la etiqueta de eje x (horizontal) generada automáticamente o seleccionar Personalizado para especificar una etiqueta.

**Etiqueta de Y.** Permite aceptar la etiqueta del eje y (vertical) generada automáticamente o seleccionar Personalizado para especificar una etiqueta.

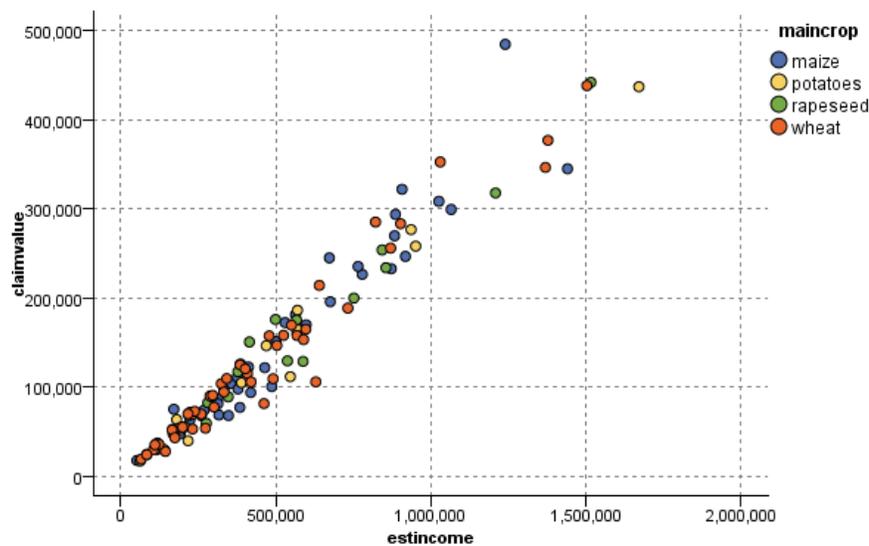
**Mostrar línea de cuadrícula.** Seleccionada por defecto, esta opción muestra una línea de cuadrícula detrás del gráfico o el diagrama que permite determinar con mayor facilidad los puntos de corte de la región y el tramo. Las líneas de cuadrícula siempre se muestran en blanco a menos que el fondo del gráfico sea blanco; en cuyo caso se muestran en gris.

### Utilización de un gráfico múltiple

Los gráficos y los gráficos múltiples son esencialmente gráficos de  $X$  frente a  $Y$ . Por ejemplo, si explora un fraude potencial en las solicitudes de subvenciones para agricultura (tal y como se ilustra en el archivo *fraud.str* de la carpeta *Demos*, en el directorio de instalación de IBM® SPSS® Modeler), puede que desee trazar el ingreso reclamado en la solicitud para compararlo con el ingreso estimado por una red neuronal. Al utilizar una superposición, como el tipo de cosecha, se ilustrará si existe una relación entre las reclamaciones (valor o número) y el tipo de cosecha.

Figura 5-62

Gráfico de la relación entre el ingreso estimado y el valor de reclamación con el tipo de cosecha principal como una superposición



Como los gráficos, los gráficos múltiples y los diagramas de evaluación son representaciones bidimensionales de  $Y$  frente a  $X$ , es fácil interactuar con ellos definiendo regiones, marcando elementos o incluso dibujando tramos. También puede generar nodos para los datos representados por dichos tramos, elemento o regiones. Si desea obtener más información, consulte el tema [Exploración de gráficos](#) el p. 355.

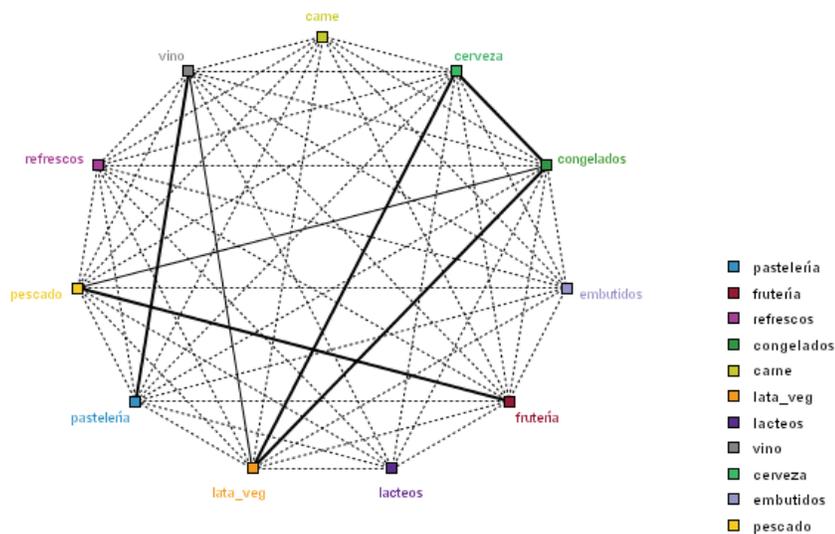
### Nodo Malla

Los nodos Malla muestran la fuerza de las relaciones entre los valores de dos o más campos simbólicos. El gráfico muestra las conexiones usando varios tipos de líneas para indicar la fuerza de conexión. Puede utilizar un nodo Malla, por ejemplo, para explorar las relaciones existentes

entre la compra de varios artículos en un sitio de comercio electrónico o un punto de venta al por menor tradicional.

Figura 5-63

Gráfico Malla que muestra las relaciones entre la compra de comestibles

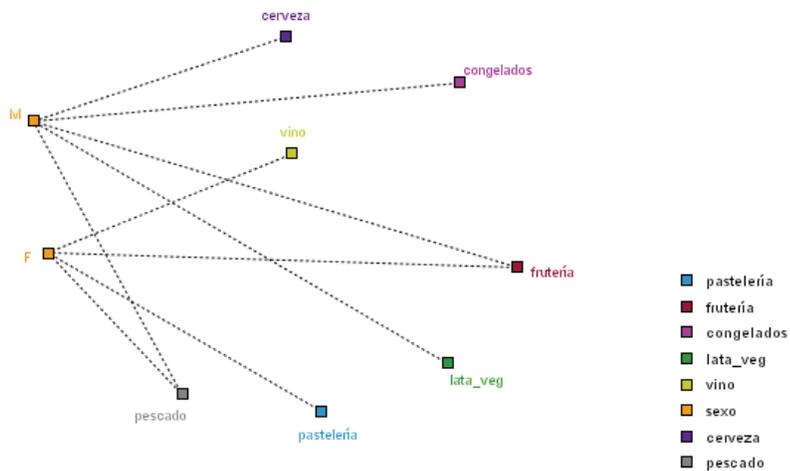


### Mallas direccionales

Los nodos de mallas direccionales son similares a los nodos Mallas en cuanto a que muestran la fuerza de las relaciones entre campos simbólicos. Sin embargo, los gráficos de mallas direccionales muestran solamente las conexiones de uno o más campos de fuente (Desde) con un único campo de destino (Hacia). Las conexiones son unidireccionales en el sentido de que son conexiones de una sola dirección.

Figura 5-64

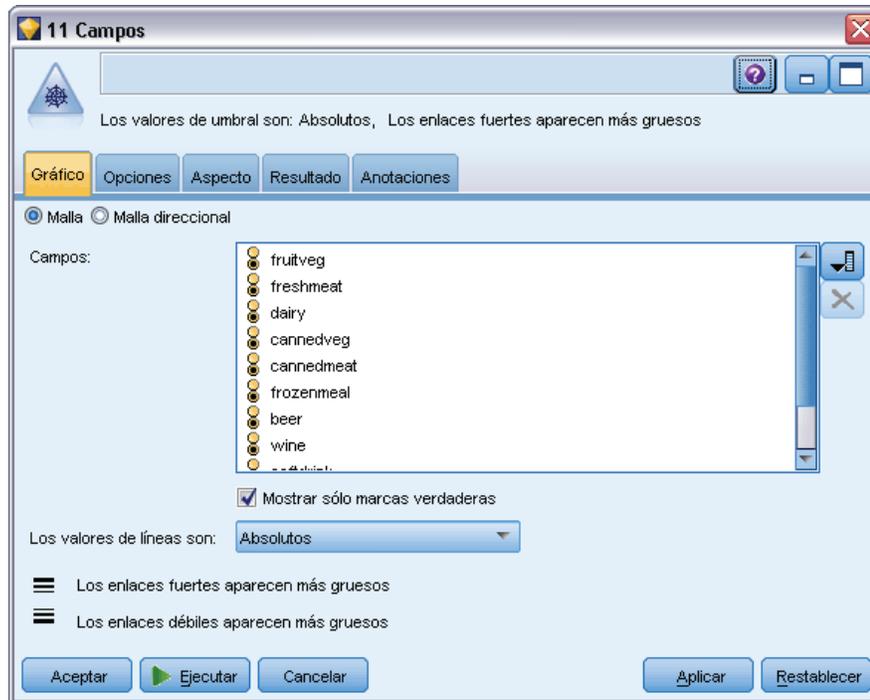
Gráfico Malla direccional que muestra la relación existente entre la compra de comestibles y el género



A semejanza de los nodos Malla, el gráfico muestra las conexiones usando varios tipos de líneas para indicar la fuerza de conexión. Puede utilizar un nodo de malla direccional, por ejemplo, para explorar las relaciones entre el género y una propensión a ciertos artículos de compra.

## Pestaña Gráfico de Malla

Figura 5-65  
Configuración de la pestaña Gráfico de un nodo Malla



**Malla.** Seleccione esta opción para crear un gráfico de malla que ilustre la fuerza de las relaciones entre todos los campos especificados.

**Malla direccional.** Seleccione esta opción para crear un gráfico de malla direccional que ilustre la fuerza de las relaciones entre múltiples campos y los valores de un campo, como el género o la religión. Cuando esta opción está seleccionada, se activa un Campo Hacia y el nombre del control de campos que aparece a continuación cambia a Campos Desde para ofrecer mayor claridad.

Figura 5-66  
Opciones de malla direccional



**Campo Hacia (solamente malla direccional).** Permite seleccionar un campo nominal o marca utilizados para crear una malla direccional. Solamente se enumeran los campos establecidos explícitamente como numéricos.

**Campos/Campos Desde.** Permite seleccionar campos para crear un gráfico de malla. Solamente se enumeran los campos establecidos explícitamente como numéricos. Utilice el selector de campos para elegir varios campos o campos por tipo.

*Nota:* en las mallas direccionales, este control se utiliza para seleccionar campos Desde.

**Mostrar solamente marcas verdaderas.** Seleccione esta opción para mostrar solamente marcas verdaderas para un campo de marcas. Esta opción simplifica la visualización de la malla y suele utilizarse para datos donde las instancias de valores positivos son de especial importancia.

**Los valores de líneas son.** Seleccione un tipo de umbral de la lista desplegable.

- Absolutos establece umbrales en función del número de registros que tengan cada par de valores.
- Porcentajes globales muestra el número de casos absolutos representados por el enlace como una proporción de todas las instancias de cada par de valores representados en el gráfico de malla.
- Porcentajes del campo/valor menor y Porcentajes del campo/valor mayor indican el campo o valor que se debe utilizar para evaluar los porcentajes. Por ejemplo, suponga que 100 registros tienen el valor *drugY* para el campo *Drug* y solamente 10 tienen el valor *LOW* para el campo *BP*. Si siete registros tienen ambos valores *drugY* y *LOW*, este porcentaje es 70% o 7%, en función del campo al que se está haciendo referencia, (*PS*) más pequeño o (*Drug*) más grande.

*Nota:* en el caso de gráficos de mallas direccionales, las opciones tercera y cuarta anteriores no están disponibles. En su lugar, puede seleccionar Porcentaje del campo/valor "Hasta" y Porcentaje del campo/valor "Desde".

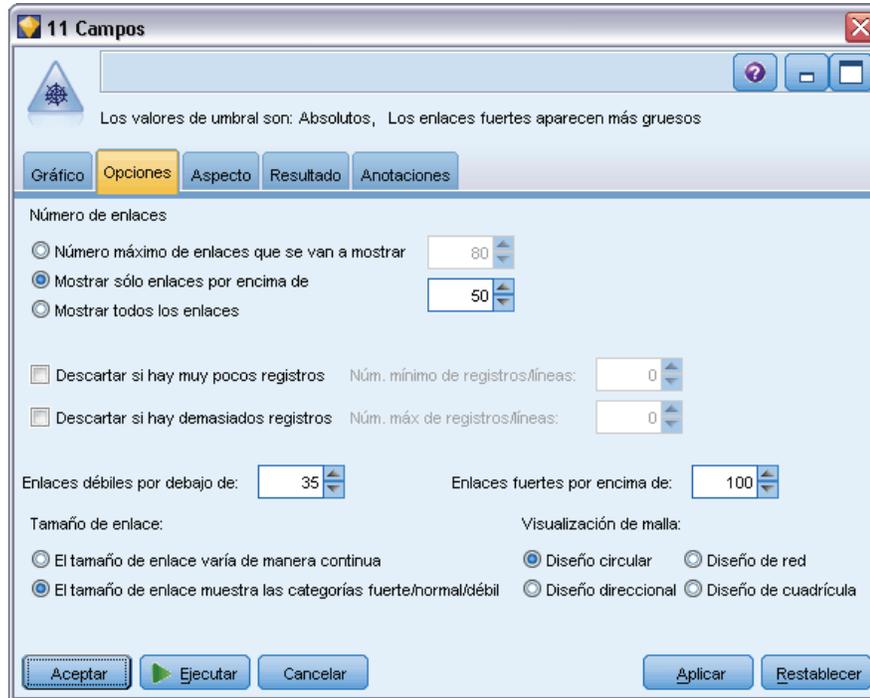
**Los enlaces fuertes aparecen más gruesos.** Esta opción, seleccionada por defecto, es la forma estándar de ver el tipo de enlace entre los campos.

**Los enlaces débiles aparecen más gruesos.** Seleccione esta opción para invertir el significado de los enlaces indicados por líneas gruesas Esta opción se suele utilizar para la detección de fraudes o el examen de valores atípicos.

## ***Pestaña Opciones de Malla***

La pestaña Opciones de los nodos Malla contiene un número de opciones adicionales para personalizar el gráfico de resultados.

Figura 5-67  
Configuración de la pestaña Opciones de un nodo Malla



**Número de enlaces.** Las siguientes opciones se utilizan para controlar el número de enlaces que se muestra en el gráfico de resultados. Algunas de estas opciones, como Enlaces débiles por encima de y Enlaces fuertes por encima de están también disponibles en la ventana del gráfico de resultado. También puede utilizar un control deslizante en el gráfico final para ajustar el número de enlaces mostrado.

- **Número máximo de enlaces que se van a mostrar.** Especifique el número máximo de enlaces que mostrar en el gráfico de resultados. Utilice las flechas para ajustar el valor.
- **Mostrar solamente enlaces por encima de.** Especifique un valor mínimo para el que mostrar una conexión en la malla. Utilice las flechas para ajustar el valor.
- **Mostrar todos los enlaces.** Especifique esta opción para mostrar todos los enlaces independientemente de los valores mínimo y máximo. Al seleccionar esta opción, el tiempo de procesamiento puede aumentar si existe un número de campos elevado.

**Descartar si hay muy pocos registros.** Seleccione esta opción para omitir las conexiones admitidas por muy pocos registros. Establezca el umbral para esta opción especificando un número en Núm. mínimo de registros/líneas.

**Descartar si hay demasiados registros.** Seleccione esta opción para omitir conexiones fuertemente admitidas. Introduzca un número en Núm. máx de registros/líneas.

**Enlaces débiles por debajo de.** Permite especificar un número para indicar el umbral para las conexiones débiles (líneas punteadas) y las conexiones habituales (líneas normales). Todas las conexiones que estén por debajo de este valor se consideran débiles.

**Enlaces fuertes por encima de.** Permite especificar un umbral para las conexiones fuertes (líneas fuertes) y las conexiones habituales (líneas normales). Todas las conexiones que estén por encima de este valor se consideran fuertes.

**Tamaño de enlace.** Permite especificar opciones para controlar el tamaño de los enlaces:

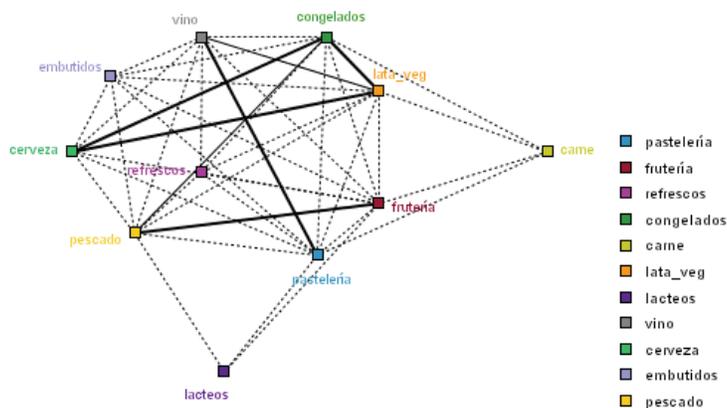
- **El tamaño de enlace varía de manera continua.** Seleccione esta opción para mostrar un rango de tamaños de enlace que refleje la variación de las fuerzas de conexión en función de los valores de datos reales.
- **El tamaño de enlace muestra las categorías fuerte/normal/débil.** Seleccione esta opción para mostrar tres fuerzas de conexión (fuerte, normal y débil). Los puntos de corte para estas categorías pueden especificarse con anterioridad y en el gráfico final.

**Visualización de malla.** Permite seleccionar un tipo de malla que mostrar:

- **Diseño circular.** Seleccione esta opción para utilizar la visualización de malla estándar.
- **Diseño de red.** Seleccione esta opción para utilizar un algoritmo para agrupar los vínculos más fuertes. Esta opción está pensada para resaltar enlaces fuertes usando diferenciaciones espaciales y líneas ponderadas.
- **Diseño direccional.** Seleccione esta opción para crear un diseño de malla direccional que utilice la selección Campo Hacia de la pestaña Gráfico como centro de la dirección.
- **Diseño de cuadrícula.** Seleccione esta opción para crear un diseño de malla en un patrón de cuadrícula con espacios regulares.

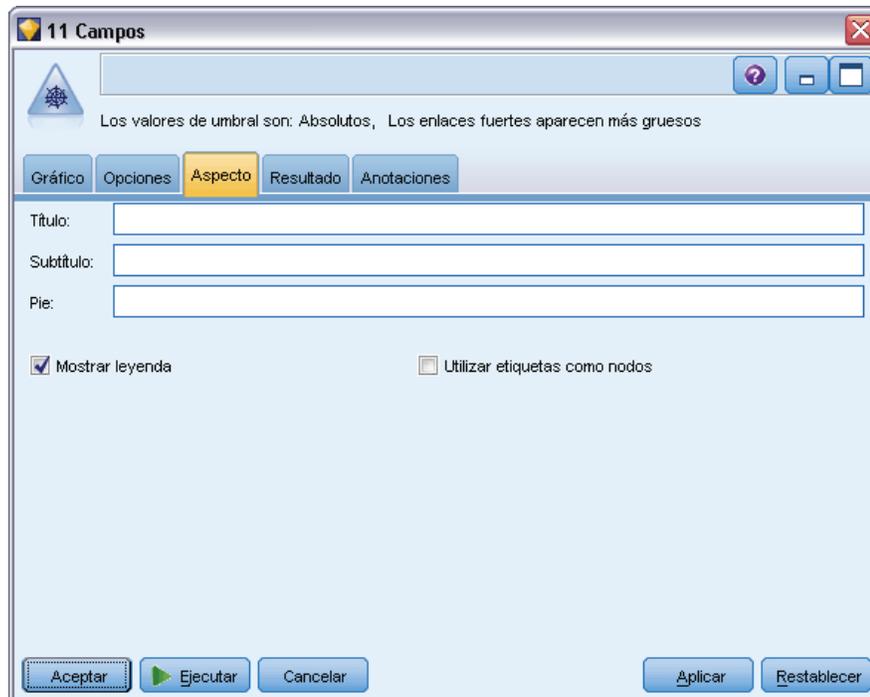
Figura 5-68

Gráfico Malla que muestra conexiones fuertes desde congelados y lata\_veg a otros comestibles



## Pestaña Aspecto de Malla

Figura 5-69  
Configuración de la pestaña Aspecto de un nodo Malla



Puede especificar las opciones de aspecto antes de crear gráficos.

**Título.** Introduzca el texto que desea para el título del gráfico.

**Subtítulo.** Introduzca el texto que desea para el subtítulo del gráfico.

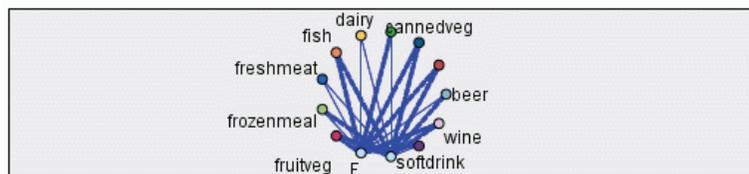
**Pie.** Introduzca el texto que desea para el pie del gráfico.

**Mostrar leyenda.** Puede especificar si se visualiza la leyenda. En el caso de los gráficos con un número elevado de campos, el aspecto del gráfico puede mejorar al ocultar la leyenda.

**Utilizar etiquetas como nodos.** Puede incluir el texto de la etiqueta en cada nodo en lugar de mostrar las etiquetas adyacentes. En el caso de los gráficos con un número pequeño de campos, el gráfico puede resultar más legible.

Figura 5-70  
Gráfico Malla que muestra etiquetas como nodos

**Relación entre el género y las compras de la tienda de comestibles**



## Utilización de un gráfico de malla

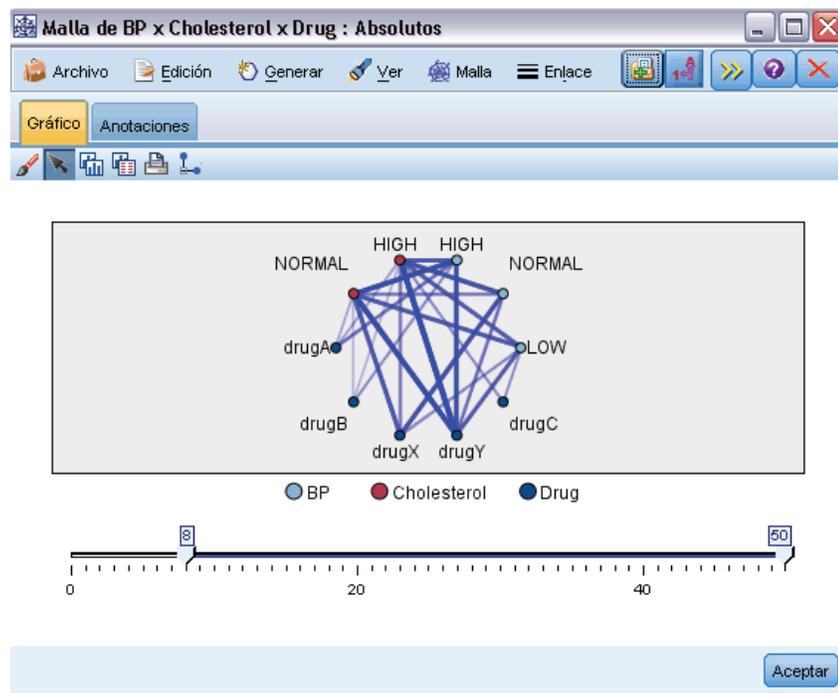
Los nodos Malla se utilizan para ilustrar la fuerza de las relaciones existentes entre los valores de dos o más campos simbólicos. Las conexiones se muestran en un gráfico con distintos tipos de líneas para indicar conexiones de creciente fuerza. Por ejemplo, puede utilizar un nodo Malla para explorar la relación existente entre los niveles de colesterol, la presión sanguínea y los medicamentos que resultaron eficaces en el tratamiento de las enfermedades de los pacientes.

- Las conexiones fuertes se indican con una línea gruesa. Así se indica que los dos valores están fuertemente relacionados y se deben explorar más.
- Las conexiones medias se muestran con una línea de grosor normal.
- Las conexiones débiles se muestran con una línea punteada.
- Si no se muestra ninguna línea entre dos valores, éstos nunca se dan en el mismo registro o esta combinación se produce en un número de registros inferior al umbral especificado en el cuadro de diálogo del nodo Malla.

Una vez creado un nodo Malla, existen varias opciones para ajustar la visualización del gráfico y generar nodos para su posterior análisis.

Figura 5-71

Gráfico de malla que indica un número de relaciones fuertes, como presión sanguínea normal con drugX y colesterol alto con drugY.



En el caso de los nodos Malla y los nodos de mallas direccionales, se puede:

- Cambiar el diseño de la visualización de la malla.
- Ocultar puntos para simplificar la visualización.
- Cambiar los umbrales que controlan los estilos de línea.

- Resaltar las líneas situadas entre los valores para indicar una relación “seleccionada”.
- Generar un nodo Seleccionar para uno o varios registros “seleccionados” o un nodo de derivación como marca asociado con una o varias relaciones en la malla.

#### ***Para ajustar puntos***

- **Mueva** puntos pulsando con el ratón en ellos y arrastrándolos a una nueva ubicación. La malla se volverá a generar para reflejar la nueva ubicación.
- **Oculte** puntos de la malla pulsando con el botón derecho del ratón en ellos y seleccionando Ocultar u Ocultar y volver a planear en el menú contextual. Ocultar solamente oculta el punto seleccionado y cualquier línea asociada a él. Ocultar y volver a planear vuelve a dibujar la malla, ajustándola a los cambios realizados. Los movimientos manuales se deshacen.
- **Muestre** todos los puntos ocultos seleccionando Mostrar todo o Mostrar todo y volver a planear en el menú de malla de la ventana del gráfico. Al seleccionar Mostrar todo y volver a planear, se vuelve a dibujar la malla y se ajusta para incluir todos los puntos ocultos previamente y sus conexiones.

#### ***Para seleccionar o “resaltar” líneas***

Las líneas seleccionadas están resaltadas en rojo.

- ▶ Para seleccionar una única línea, pulse con el botón izquierdo sobre la línea.
- ▶ Para seleccionar varias líneas, realice una de las siguientes acciones:
  - Utilizando el cursor, dibuje un círculo alrededor de los puntos de las líneas que desee seleccionar.
  - Mantenga pulsada la tecla Ctrl y pulse con el botón izquierdo las líneas que desee seleccionar.

Puede anular la selección de todas las líneas seleccionadas pulsando sobre el fondo del gráfico o seleccionando Borrar selección en el menú Malla de la ventana del gráfico.

#### ***Para ver la malla mediante un diseño diferente***

- ▶ En el menú Malla, seleccione Diseño circular, Diseño de red, Diseño direccional o Diseño de cuadrícula para cambiar el diseño del gráfico.

#### ***Para activar o desactivar el control deslizante de enlaces***

- ▶ En el menú Ver, seleccione Control deslizante de enlaces.

#### ***Para seleccionar o marcar registros de una relación simple***

- ▶ Pulse con el botón derecho del ratón en la línea que representa la relación de interés.
- ▶ En el menú contextual, elija Generar nodo Seleccionar para el enlace o Generar nodo Derivar para el enlace.

Automáticamente se añade un nodo Seleccionar o un nodo Derivar al lienzo de rutas con las opciones y condiciones especificadas adecuadas:

- El nodo Seleccionar selecciona todos los registros de una relación dada.
- El nodo Derivar genera una marca que indica si la relación seleccionada es verdadera para los registros de todo el conjunto de datos. El campo de marcas toma su nombre de la unión de dos valores de la relación unidos por guión bajo (por ejemplo *LOW\_drugC* o *drugC\_LOW*).

#### **Para seleccionar o marcar registros para un grupo de relaciones**

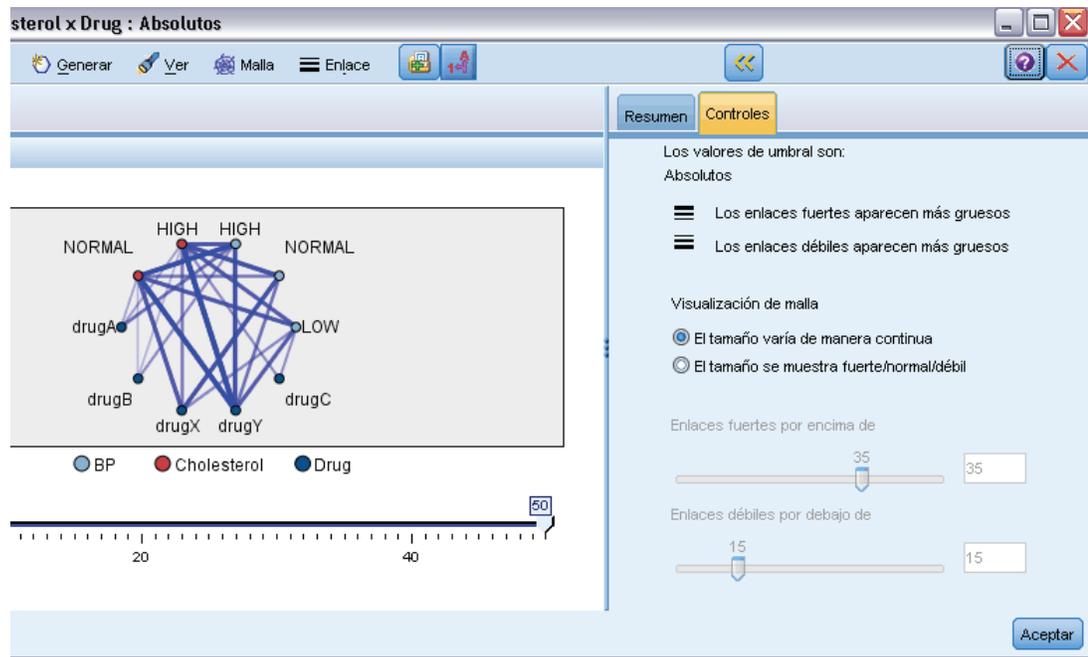
- ▶ Seleccione las líneas de la visualización de malla que representan relaciones de interés.
- ▶ En el menú Generar de la ventana del gráfico, elija Nodo Seleccionar (“Y”), Nodo Seleccionar (“O”), Nodo Derivar (“Y”) o Nodo Derivar (“O”).
  - El nodo “O” proporciona la disyunción de condiciones. Esto significa que el nodo se aplicará a los registros para los que cualquiera de las relaciones seleccionadas sea válida.
  - El nodo “Y” proporciona la conjunción de condiciones. Esto significa que el nodo se aplicará solamente a los registros para los que todas las relaciones seleccionadas sean válidas. Se producirá un error si cualquiera de las relaciones seleccionadas es mutuamente excluyente.

Una vez completada la selección, se añade un nodo Seleccionar o un nodo Derivar automáticamente al lienzo de rutas con las opciones y condiciones especificadas adecuadas.

#### **Ajuste de umbrales de malla**

Una vez creado el gráfico de malla, puede ajustar los umbrales que controlan los estilos de línea mediante la barra de herramientas para cambiar la línea mínima visible. También puede ver opciones de umbrales adicionales al pulsar en el botón de flecha doble amarilla de la barra de herramientas para expandir la ventana de gráfico de malla. A continuación, pulse en la pestaña Controles para ver opciones adicionales.

Figura 5-72  
Ventana expandida que muestra las opciones de umbral y presentación



**Los valores de umbral son.** Muestra el tipo de umbral seleccionado durante la creación en el cuadro de diálogo del nodo Malla.

**Los enlaces fuertes aparecen más gruesos.** Esta opción, seleccionada por defecto, es la forma estándar de ver el tipo de enlace entre los campos.

**Los enlaces débiles aparecen más gruesos.** Seleccione esta opción para invertir el significado de los enlaces indicados por líneas gruesas. Esta opción se suele utilizar para la detección de fraudes o el examen de valores atípicos.

**Visualización de malla.** Permite especificar las opciones para controlar el tamaño de los gráficos de resultado:

- **El tamaño varía de manera continua.** Seleccione esta opción para mostrar un rango de tamaños de enlace que refleje la variación de las fuerzas de conexión en función de los valores de datos reales.
- **El tamaño se muestra fuerte/normal/débil.** Seleccione esta opción para mostrar tres fuerzas de conexión (fuerte, normal y débil). Los puntos de corte para estas categorías pueden especificarse con anterioridad y en el gráfico final.

**Enlaces fuertes por encima de.** Permite especificar un umbral para las conexiones fuertes (líneas fuertes) y las conexiones habituales (líneas normales). Todas las conexiones que estén por encima de este valor se consideran fuertes. Utilice el control deslizante para ajustar el valor o introducir un número en el campo.

**Enlaces débiles por debajo de.** Permite especificar un número para indicar el umbral para las conexiones débiles (líneas punteadas) y las conexiones habituales (líneas normales). Todas las conexiones que estén por debajo de este valor se consideran débiles. Utilice el control deslizante para ajustar el valor o introducir un número en el campo.

Una vez ajustados los umbrales de la malla, puede volver a planear o a dibujar la visualización de la malla con los valores de umbral nuevos a través del menú de malla ubicado en la barra de herramientas de gráfico de malla. Una vez encontrada la configuración que muestra los patrones más significativos, puede actualizar la configuración original en el nodo Malla (también denominado nodo de malla parental) seleccionando Actualizar nodo parental en el menú Malla de la ventana del gráfico.

### Creación de un resumen de malla

Puede crear un documento de resumen de malla que enumere enlaces fuertes, medios y débiles pulsando en el botón de flecha doble amarilla de la barra de herramientas para expandir la ventana del gráfico de malla. A continuación, pulse en la pestaña Resumen para ver las tablas de cada tipo de enlace. Las tablas pueden expandirse y contraerse mediante los botones de conmutación de cada una.

Figura 5-73

Resumen de malla que enumera las conexiones existentes entre la presión sanguínea, el colesterol y el tipo de medicamento.

Resumen		
Controles		
Enlaces fuertes		
Enlaces	Campo 1	Campo 2
47	Cholesterol = "HIGH"	Drug = "drugY"
44	Cholesterol = "NORMAL"	Drug = "drugY"
42	BP = "HIGH"	Cholesterol = "NORMAL"
38	BP = "HIGH"	Drug = "drugY"
37	BP = "NORMAL"	Cholesterol = "HIGH"
36	BP = "NORMAL"	Drug = "drugX"
Enlaces medios		
Enlaces	Campo 1	Campo 2
35	BP = "HIGH"	Cholesterol = "HIGH"
34	Cholesterol = "NORMAL"	Drug = "drugX"
33	BP = "LOW"	Cholesterol = "NORMAL"
31	BP = "LOW"	Cholesterol = "HIGH"
30	BP = "LOW"	Drug = "drugY"
23	BP = "NORMAL"	Drug = "drugY"
23	BP = "HIGH"	Drug = "drugA"
22	BP = "NORMAL"	Cholesterol = "NORMAL"
20	Cholesterol = "HIGH"	Drug = "drugX"
18	BP = "LOW"	Drug = "drugX"
16	BP = "LOW"	Drug = "drugC"
16	Cholesterol = "HIGH"	Drug = "drugC"
16	BP = "HIGH"	Drug = "drugB"
Enlaces débiles		
Enlaces	Campo 1	Campo 2
12	Cholesterol = "HIGH"	Drug = "drugA"
11	Cholesterol = "NORMAL"	Drug = "drugA"
8	Cholesterol = "HIGH"	Drug = "drugB"
8	Cholesterol = "NORMAL"	Drug = "drugB"

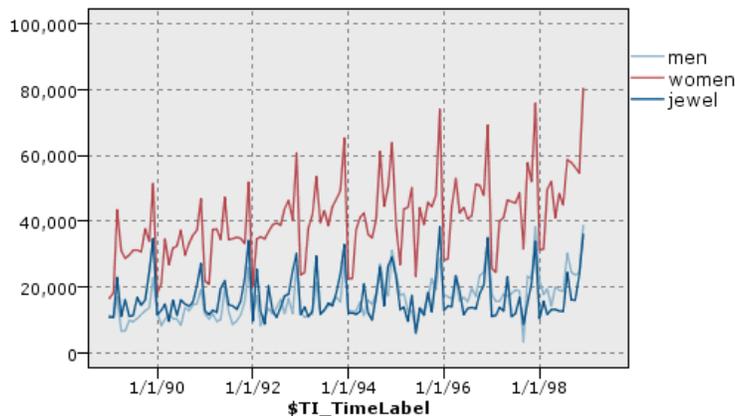
Para imprimir el resumen, seleccione lo siguiente desde el menú en la ventana del gráfico Web:  
File > Imprimir resumen

## Nodo Gráfico de tiempo

Los nodos Gráfico de tiempo permiten ver una o más series representadas durante un período de tiempo. La serie que represente debe contener valores numéricos que deben producirse durante un intervalo de tiempo en el que los períodos son uniformes. Se suele utilizar un nodo Intervalos de tiempo antes de un nodo Gráfico de tiempo para crear un campo *EtiquetaTiempo*, que se utiliza por defecto para etiquetar el eje x en los gráficos. Si desea obtener más información, consulte el tema [Nodo de intervalos de tiempo](#) en el capítulo 4 el p. 216.

Figura 5-74

*Representación de las ventas de ropa y joyería para hombre y mujer durante un período de tiempo*

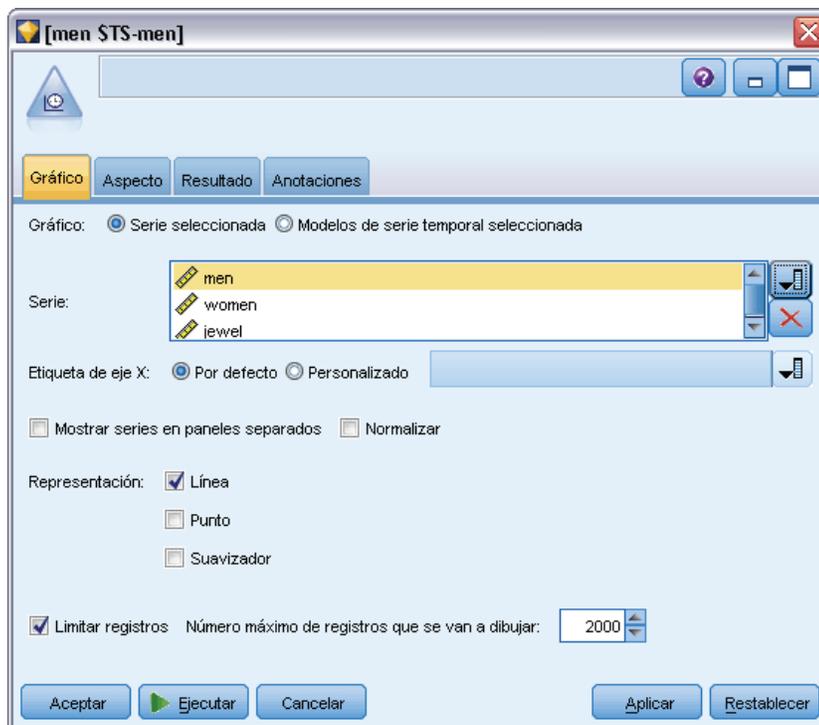


### Creación de eventos e intervenciones

Puede crear campos de eventos e intervenciones a partir de los resultados del gráfico de tiempo mediante la generación de un nodo Derivar (marca o nominal) desde los menús contextuales. Por ejemplo, podría crear un campo de evento en el caso de una huelga de trenes, donde el estado sea Verdadero si se produce el evento y Falso en el caso contrario. En un campo de intervención, por ejemplo para una subida de precios, podría utilizar una derivación como recuento para identificar la fecha de la subida, con 0 para el precio anterior y 1 para el nuevo. Si desea obtener más información, consulte el tema [Nodo Derivar](#) en el capítulo 4 el p. 165.

## Pestaña Gráfico de Tiempo

Figura 5-75  
Configuración de la pestaña Gráfico de un nodo Gráfico de Tiempo



**Gráfico.** Permite seleccionar el modo de representación de los datos de series.

- **Serie seleccionada.** Representa los valores para las series temporales seleccionadas. Si selecciona esta opción para representar los intervalos de confianza, anule la selección de la casilla de verificación Normalizar.
- **Modelos de serie temporal seleccionada.** Utilizada junto con un modelo Serie temporal, esta opción representa todos los campos relacionados (valores reales y pronosticados, e intervalos de confianza) para una o más series temporales seleccionadas. Esta opción deshabilita algunas de las demás opciones del cuadro de diálogo. Ésta es la opción más adecuada para la representación de intervalos de confianza.

**Serie.** Seleccione uno o más campos con los datos de series temporales que desea representar. Los datos deben ser numéricos.

**Etiqueta de eje X.** Elija la etiqueta por defecto o un campo único para utilizarlo como etiqueta del eje  $x$  en los gráficos. Si selecciona Por defecto, el sistema utilizará el campo EtiquetaTiempo creado a partir de números enteros anteriores de la ruta o secuenciales del nodo Intervalos de tiempo si no hay ningún nodo Intervalos de tiempo. Si desea obtener más información, consulte el tema [Nodo de intervalos de tiempo](#) en el capítulo 4 el p. 216.

**Mostrar series en paneles independientes.** Determina si cada serie se muestra en un panel independiente. Como alternativa, si no selecciona ningún panel, todas las series temporales se representarán en el mismo gráfico y los suavizamientos no estarán disponibles. Al representar todas las series temporales en el mismo gráfico, cada serie se representará en un color diferente.

**Normalizar.** Seleccione esta opción para escalar todos los valores de  $Y$  para el rango de 0 a 1 al mostrarlos en el gráfico. La normalización ayuda a explorar la relación entre líneas que, de lo contrario, estaría oculta debido las diferencias existentes en el rango de valores de cada serie, y se recomienda cuando se representan varias líneas en el mismo gráfico, o cuando se contrastan representaciones en paneles a cada lado. (La normalización no es necesaria cuando todos los valores de datos se encuentran en un rango parecido.)

**Mostrar.** Seleccione uno o más elementos para mostrarlos en el gráfico. Puede elegir entre líneas, puntos y suavizamientos (LOESS). Los suavizamientos están disponibles solamente si muestra las series en paneles independientes. Por defecto, se selecciona el elemento de línea. Asegúrese de que selecciona al menos un elemento de gráfico antes de ejecutar el nodo de gráficos; de lo contrario el sistema devolverá un error indicando que no ha seleccionado ningún elemento para representar.

**Limitar registros.** Seleccione esta opción si desea limitar el número de registros representados. Especifique el número de registros, leídos desde el inicio del archivo de datos, que se representarán en la opción Número máximo de registros que se van a dibujar. Por defecto el número se define como 2.000. Si desea representar los últimos  $n$  registros del archivo de datos, puede utilizar un nodo Ordenar antes de este nodo para organizar los registros en orden descendente por tiempo.

## Pestaña Aspecto de Gráfico de Tiempo

Figura 5-76  
Configuración de la pestaña Aspecto de un nodo Gráfico de tiempo

The image shows a software dialog box titled "Gráfico de tiempo". It has a standard Windows-style title bar with a close button. Below the title bar is a toolbar with a help icon, a question mark icon, and a close icon. The dialog is divided into four tabs: "Gráfico", "Aspecto" (which is selected and highlighted in yellow), "Resultado", and "Anotaciones". Under the "Aspecto" tab, there are several input fields and options:

- "Título:" followed by a text input field.
- "Subtítulo:" followed by a text input field.
- "Pie:" followed by a text input field.
- "Etiqueta de X:" with two radio buttons: "Automático" (selected) and "Personalizado". To the right of "Personalizado" is a text input field.
- "Etiqueta de Y:" with two radio buttons: "Automático" (selected) and "Personalizado". To the right of "Personalizado" is a text input field.
- A checked checkbox labeled "Mostrar línea de cuadrícula".
- "Diseño:" with two radio buttons: "Horizontal" (selected) and "Vertical".

At the bottom of the dialog, there are five buttons: "Aceptar", "Ejecutar" (with a play icon), "Cancelar", "Aplicar", and "Restablecer".

Puede especificar las opciones de aspecto antes de crear gráficos.

**Título.** Introduzca el texto que desea para el título del gráfico.

**Subtítulo.** Introduzca el texto que desea para el subtítulo del gráfico.

**Pie.** Introduzca el texto que desea para el pie del gráfico.

**Etiqueta de X.** Permite aceptar la etiqueta de eje x (horizontal) generada automáticamente o seleccionar Personalizado para especificar una etiqueta.

**Etiqueta de Y.** Permite aceptar la etiqueta del eje y (vertical) generada automáticamente o seleccionar Personalizado para especificar una etiqueta.

**Mostrar línea de cuadrícula.** Seleccionada por defecto, esta opción muestra una línea de cuadrícula detrás del gráfico o el diagrama que permite determinar con mayor facilidad los puntos de corte de la región y el tramo. Las líneas de cuadrícula siempre se muestran en blanco a menos que el fondo del gráfico sea blanco; en cuyo caso se muestran en gris.

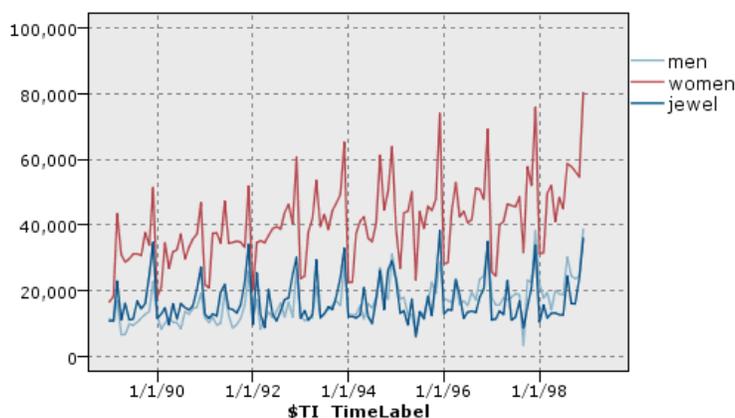
**Diseño.** Sólo en el caso de los gráficos de tiempo, puede especificar si los valores de tiempo se representan en un eje horizontal o vertical.

## Utilización de un gráfico de tiempo

Una vez creado un gráfico de tiempo, existen varias opciones para ajustar la visualización del gráfico y generar nodos para su posterior análisis. Si desea obtener más información, consulte el tema [Exploración de gráficos](#) el p. 355.

Figura 5-77

*Representación de las ventas de ropa y joyería para hombre y mujer durante un período de tiempo*



Una vez creado un gráfico de tiempo, definidos los tramos y examinados los resultados, puede utilizar las opciones del menú Generar y el menú contextual para crear nodos Seleccionar o Derivar. Si desea obtener más información, consulte el tema [Generación de nodos a partir de gráficos](#) el p. 364.

## Nodo Evaluación

El nodo Evaluación ofrece una forma sencilla de evaluar y comparar modelos predictivos para elegir el mejor modelo para su aplicación. Los diagramas de evaluación muestran el comportamiento de los modelos pronosticando determinados resultados. Funcionan ordenando los

registros en función del valor pronosticado y confianza del pronóstico, dividiendo los registros en grupos de igual tamaño (**cuantiles**) y, a continuación, dibujando el valor del criterio de negocios de cada cuantil, del más alto al más bajo. El gráfico muestra múltiples modelos como líneas independientes.

Los resultados se gestionan definiendo un valor o rango de valores específicos como un **acierto**. Los aciertos suelen indicar algún tipo de éxito (como una venta a un cliente) o un evento de interés (como un diagnóstico médico específico). Puede definir criterios de aciertos en la pestaña Opciones del cuadro de diálogo o utilizar los criterios de aciertos por defecto como se indica a continuación:

- Los campos de salida de **marcas** son directos; los aciertos corresponden a valores *verdaderos*.
- Para los campos de resultado **nominales**, el primer valor del conjunto define un acierto.
- Para los campos de resultado **Continuos**, los aciertos equivalen a valores mayores que el punto medio del rango del campo.

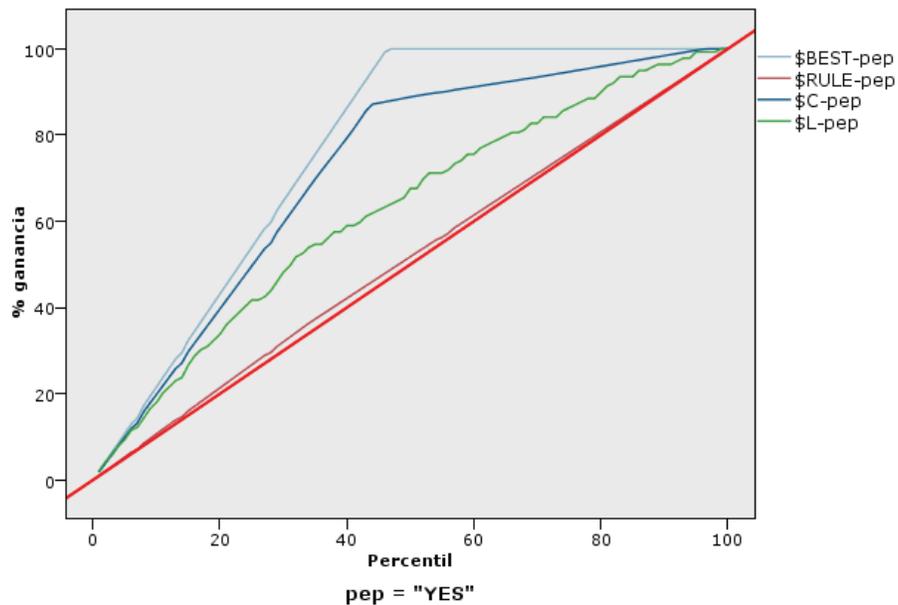
Existen cinco tipos de diagramas de evaluación, cada uno de ellos con el énfasis puesto en un criterio de evaluación diferente.

### Ganancias

Las ganancias se definen como la proporción de aciertos totales que se produce en cada cuantil. Las ganancias se calculan como el resultado de: (número de aciertos en cuantil / número total de aciertos)  $\times$  100%.

Figura 5-78

Gráfico de ganancias (acumulado) con línea base que incluye mejor línea y regla de negocio

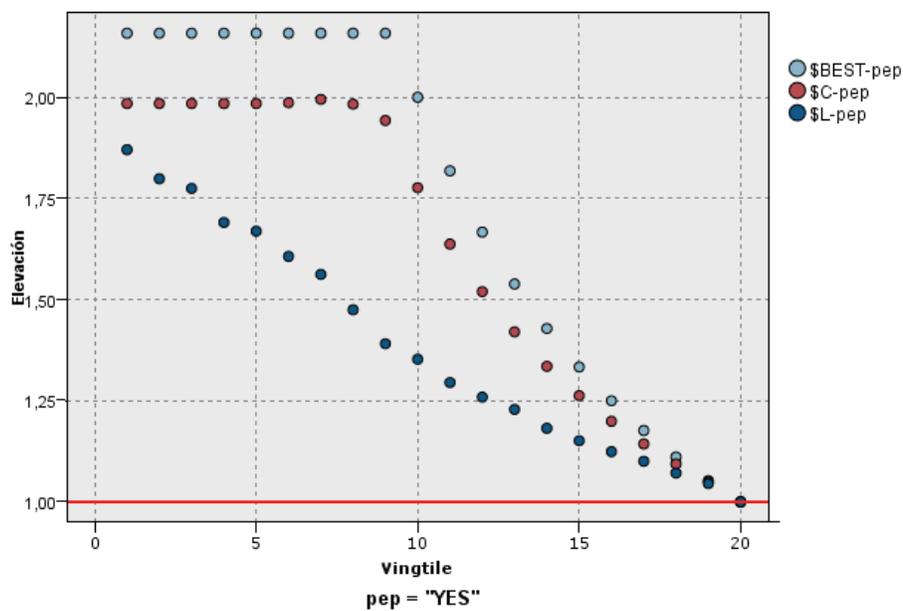


### Elevación

La elevación compara el porcentaje de registros de cada cuantil que supone aciertos con el porcentaje global de aciertos de los datos de entrenamiento. Se calcula como el resultado de: (aciertos del cuantil / registros del cuantil) / (aciertos totales / registros totales).

Figura 5-79

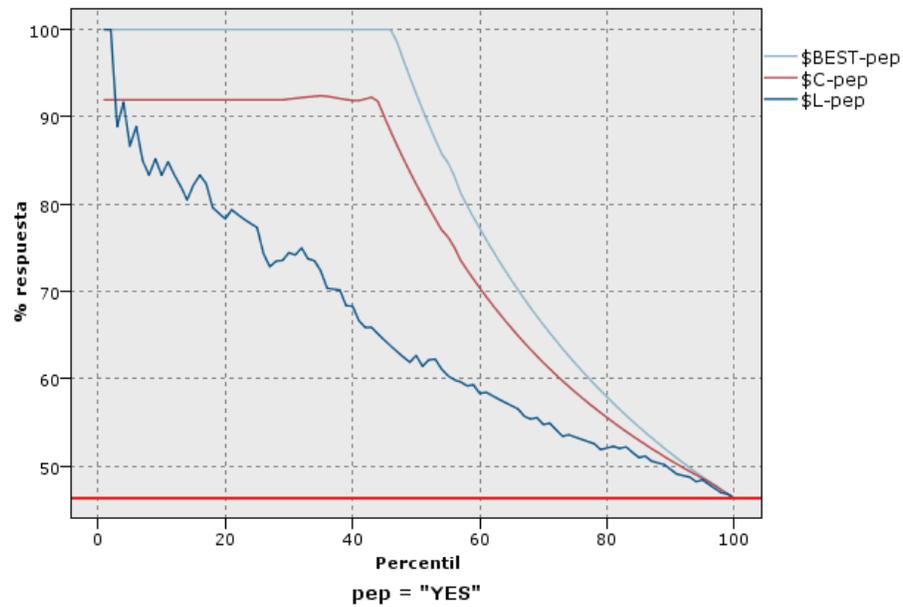
Gráfico de elevación (acumulado) mediante puntos y mejor línea.



### Respuesta

La respuesta es sencillamente el porcentaje de registros del cuantil que son aciertos. La respuesta se calcula como el resultado de: (aciertos del cuantil / registros del cuantil) × 100%.

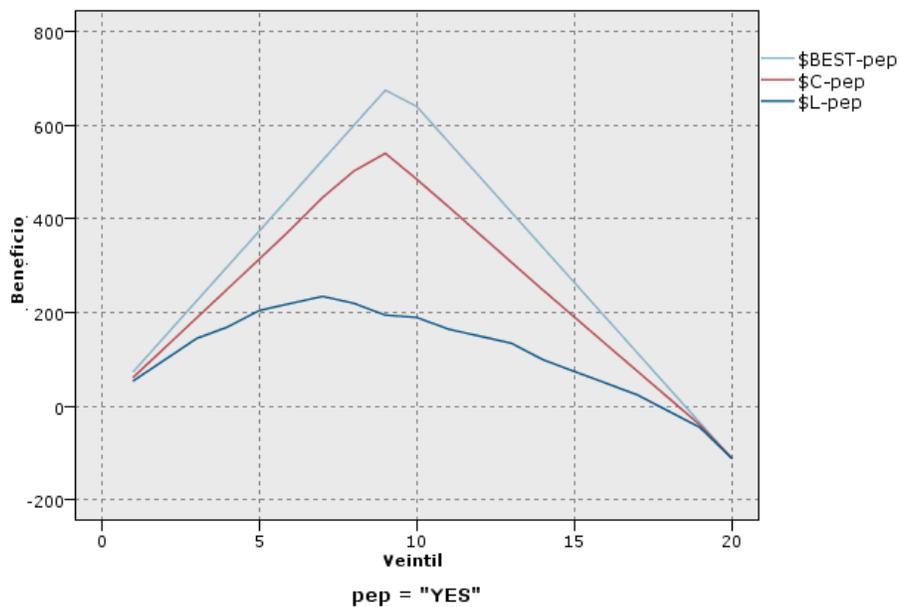
Figura 5-80  
Gráfico de respuestas (acumulado) con mejor línea



### **Beneficio**

El beneficio es igual a los **ingresos** de cada registro menos el **coste** del registro. Los beneficios de un cuantil son la suma de los beneficios de todos los registros del cuantil. Se asume que los ingresos se aplican solamente a los aciertos, pero los costes se aplican a todos los registros. Los beneficios y los costes se pueden fijar o estar definidos por campos en los datos. Los beneficios se calculan como el resultado de: (suma de los ingresos de los registros del cuantil – suma de los costes de los registros del cuantil).

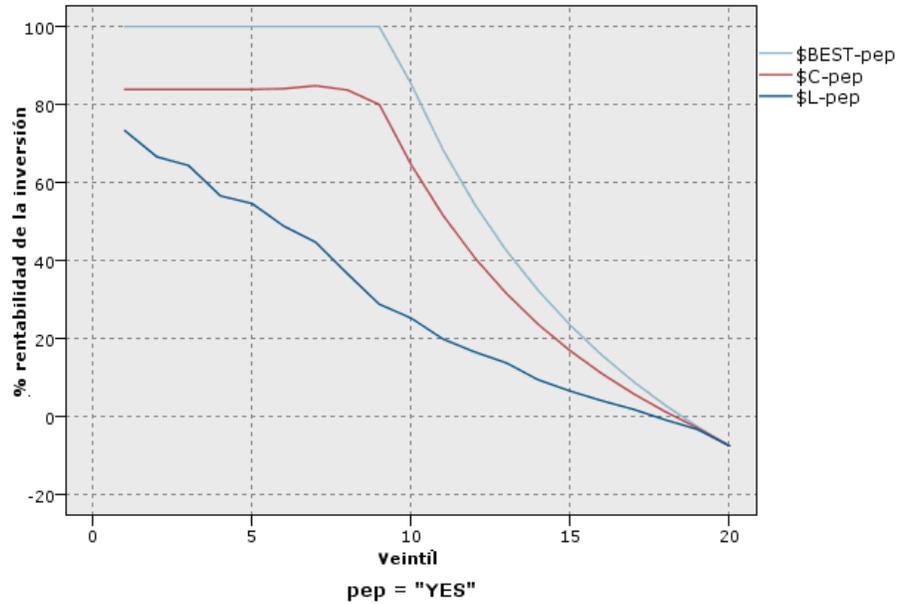
Figura 5-81  
Gráfico de beneficios (acumulado) con mejor línea



### ***Rentabilidad de la inversión***

La rentabilidad de la inversión (ROI, del inglés 'Return On Investment') es similar al beneficio en cuanto a que implica la definición de ingresos y costes. La rentabilidad de la inversión compara los beneficios con los costes del cuantil. La rentabilidad de la inversión se calcula como el resultado de:  $(\text{beneficios del cuantil} / \text{costes del cuantil}) \times 100\%$ .

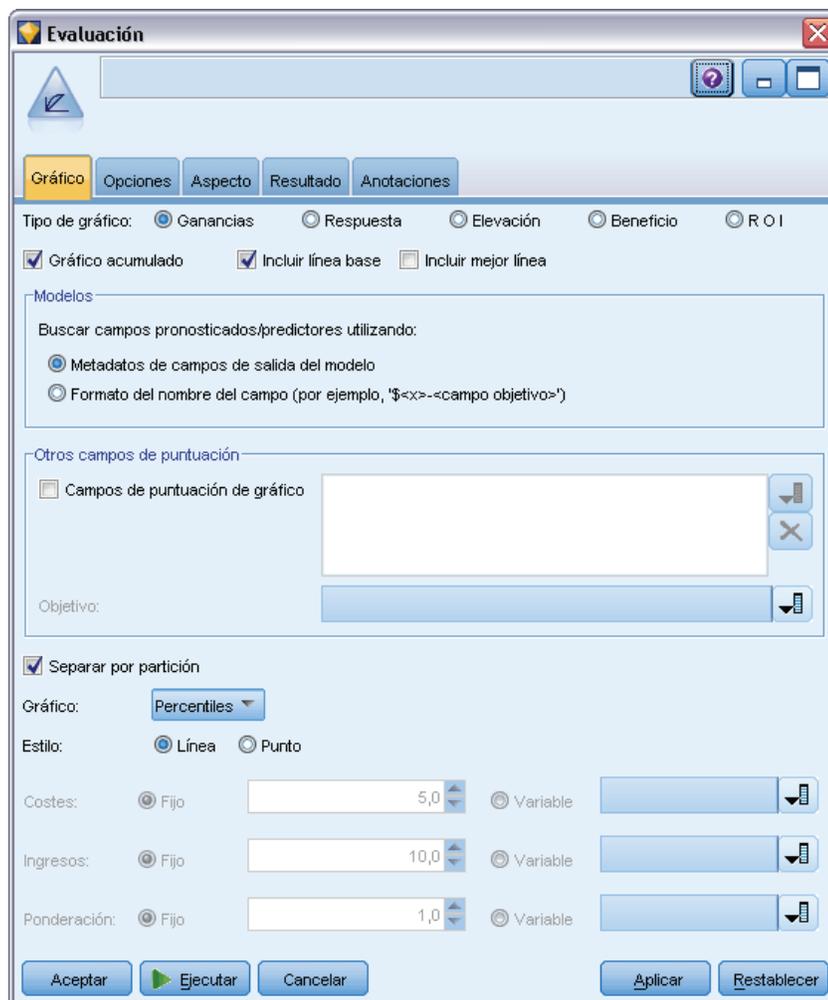
Figura 5-82  
Gráfico de rentabilidad de la inversión (acumulado) con mejor línea



Los diagramas de evaluación también pueden ser acumulados, de forma que cada punto equivalga al valor del cuantil correspondiente más todos los cuantiles mayores. Los gráficos acumulados suelen mostrar mejor el rendimiento global de modelos, mientras que los gráficos no acumulados suelen ser mejores para indicar determinadas áreas de problemas para los modelos.

## Pestaña Gráfico de Evaluación

Figura 5-83  
Configuración de la pestaña Gráfico de un nodo Evaluación



**Tipo de gráfico.** Seleccione uno de los siguientes tipos: Ganancias, Respuesta, Elevación, Beneficio o Rentabilidad de la inversión.

**Gráfico acumulado.** Seleccione esta opción para crear un gráfico acumulado. Los valores de los gráficos acumulados se dibujan para cada cuantil más todos los cuantiles mayores.

**Incluir línea base.** Seleccione esta opción para incluir una línea base en el gráfico, indicando una distribución aleatoria perfecta de los aciertos donde la confianza se vuelve irrelevante. (Incluir línea base no está disponible para los gráficos Beneficio y Rentabilidad de la inversión.)

**Incluir mejor línea.** Seleccione esta opción para incluir una mejor línea en el gráfico, que indique una confianza perfecta (donde los aciertos = 100% de los casos).

**Buscar campos pronosticados/predictores utilizando.** Seleccione Metadatos de campos de salida del modelo para buscar los campos pronosticados del gráfico mediante metadatos, o seleccione Formato del nombre del campo para buscarlos por nombre.

**Campos Puntuación de Gráfico.** Seleccione esta casilla de verificación para activar el selector de campos de puntuación. A continuación, seleccione uno o más rangos o campos de puntuación continuos, es decir, campos que no sean modelos estrictamente predictivos que puedan ser útiles para clasificar registros en términos de propensión para ser un acierto. El nodo Evaluación puede comparar cualquier combinación de uno o más campos de puntuación o modelos más predictivos. Un ejemplo típico puede ser comparar varios campos RFM con su mejor modelo predictivo.

**Objetivo.** Seleccione el campo objetivo mediante el selector de campo. Seleccione cualquier campo de nominal o marca instanciado o defina el campo con dos o más valores.

*Nota:* Este campo objetivo solamente se aplica a los campos de puntuación (modelos predictivos definidos por sus propios objetivos) y se ignora si se define un criterio de acierto personalizado en la pestaña Opciones.

**Dividir por partición.** Si utiliza un campo de partición para dividir los registros en muestreos de entrenamiento, comprobación y validación, seleccione esta opción para mostrar un diagrama de evaluación independiente para cada partición. Si desea obtener más información, consulte el tema [Nodo Partición](#) en el capítulo 4 el p. 205.

*Nota:* al dividir en particiones, los registros con valores nulos en el campo de partición se excluyen de la evaluación. Esto nunca debería ser un problema si se utiliza un nodo Partición ya que este tipo de nodos no genera valores nulos.

**Gráfico.** Permite seleccionar el tamaño de los cuantiles para dibujarlos en el gráfico desde la lista desplegable. Las opciones son Cuantiles, Quintiles, Deciles, Veintiles, Percentiles y 1000-tilas.

**Estilo.** Seleccione Línea o Punto.

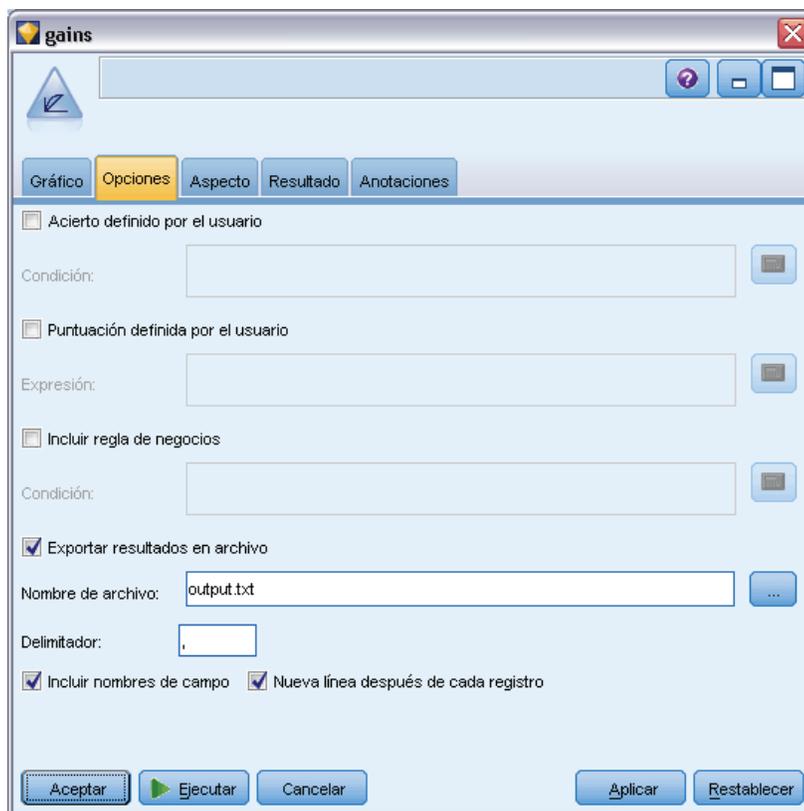
**Gráficos de beneficios y ROI** En el caso de los gráficos Beneficios y Rendimiento de la inversión, los controles adicionales permiten especificar costes, ingresos y ponderaciones.

- **Costes.** Permite especificar el coste asociado con cada registro. Puede seleccionar costes fijos o variables. En el caso de los costes fijos, especifique el valor del coste. En el caso de los costes variables, pulse en el selector de campos para elegir un campo de costes.
- **Ingresos.** Permite especificar los ingresos asociados con cada registro que representa un acierto. Puede seleccionar costes fijos o variables. En el caso de los ingresos fijos, especifique el valor del ingreso. En el caso de los ingresos variables, pulse en el selector de campos para elegir un campo de ingresos.
- **Ponderación.** Si los registros de los datos representan más de una unidad, puede utilizar ponderaciones de frecuencias para ajustar los resultados. Especifique la ponderación asociada con cada registro mediante valores fijos o variables. En el caso de las ponderaciones fijas, especifique el valor de ponderación (el número de unidades por registro). En el caso de ponderaciones variables, pulse en el selector de campos para elegir un campo de ponderaciones.

## ***Pestaña Opciones de Evaluación***

La pestaña Opciones de los diagramas de evaluación proporciona flexibilidad al definir aciertos, criterios de puntuación y reglas de negocios mostradas en el gráfico. También puede establecer opciones para exportar los resultados de la evaluación del modelo.

Figura 5-84  
Configuración de la pestaña Opciones de un nodo Evaluación



**Acierto definido por el usuario.** Seleccione esta opción para especificar una condición personalizada con objeto de indicar un acierto. Esta opción resulta útil para definir el resultado de interés en lugar de deducirlo del tipo de campo objetivo y el orden de valores.

- **Condición.** Cuando se selecciona Acierto definido por el usuario, se debe especificar una expresión CLEM para una condición de acierto. Por ejemplo, `@TARGET = "YES"` es una condición válida que indica que un valor *Sí* para el campo objetivo se tomará como un acierto en la evaluación. La condición especificada se utilizará para todos los campos objetivo. Para crear una condición, escriba en el campo o utilice el generador de expresiones para generar una expresión de condición. Si los datos están instanciados, puede insertar valores directamente desde el generador de expresiones.

**Puntuación definida por el usuario.** Seleccione esta opción para especificar una condición utilizada para puntuar casos antes de asignarlos a los cuantiles. La puntuación por defecto se calcula a partir del valor pronosticado y la confianza. Utilice el campo Expresión para crear una expresión de puntuación personalizada.

- **Expresión.** Permite especificar una expresión CLEM utilizada para la puntuación. Por ejemplo, si un resultado numérico en el rango de 0 a 1 se ordena de forma que los valores más bajos sean mejores que los más altos, puede definir un acierto como `@TARGET < 0.5` y la puntuación asociada como `1 • @PREDICTED`. La expresión de puntuación debe producir un valor numérico. Para crear una condición, escriba en el campo o utilice el generador de expresiones para generar una expresión de condición.

**Incluir regla de negocios.** Seleccione esta opción para especificar una condición de regla que refleje los criterios de interés. Por ejemplo, es posible que desee visualizar una regla para todos los casos en los que mortgage = "Y" and income >= 33000. Las reglas de negocios se dibujan en el gráfico y se etiquetan en la clave como *Regla*.

- **Condición.** Permite especificar una expresión CLEM utilizada para definir una regla de negocios en el gráfico de resultados. Simplemente escriba en el campo o utilice el generador de expresiones para generar una expresión de condición. Si los datos están instanciados, puede insertar valores directamente desde el generador de expresiones.

**Exportar resultados en archivo.** Seleccione esta opción para exportar los resultados de la evaluación del modelo a un archivo de texto delimitado. Puede leer este archivo para realizar análisis especializados en los valores calculados. Establezca las siguientes opciones de exportación:

- **Nombre de archivo.** Permite introducir el nombre del archivo de resultados. Utilice el botón de puntos suspensivos (...) para seleccionar la carpeta deseada.
- **Delimitador.** Permite introducir un carácter, como una coma o un espacio, que utilizar como delimitador de campos.

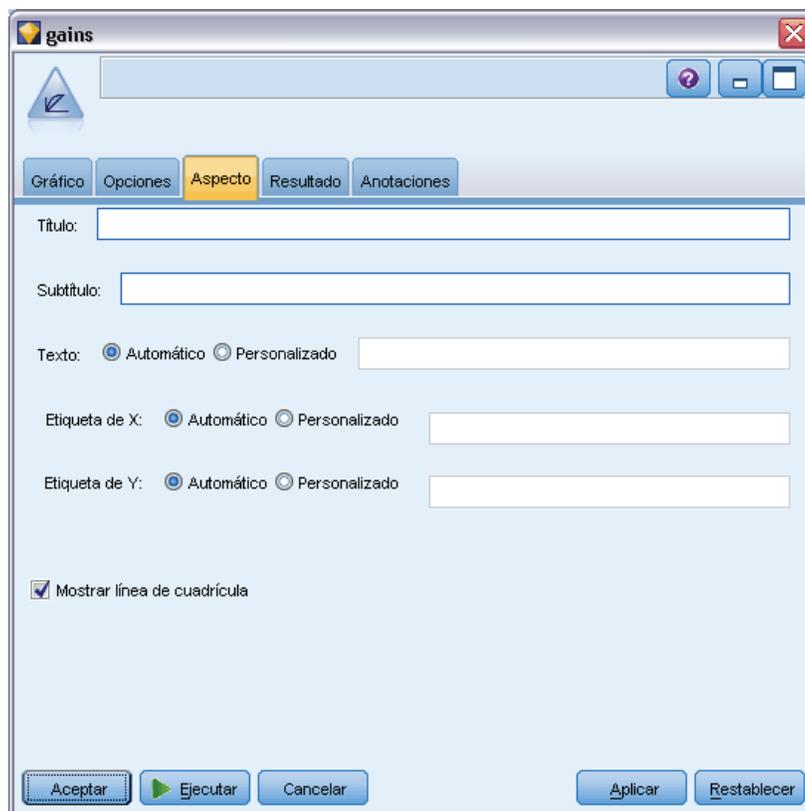
**Incluir nombres de campo.** Seleccione esta opción para incluir nombres de campos como la primera línea del archivo de resultados.

**Nueva línea después de cada registro.** Seleccione esta opción para comenzar cada registro en una nueva línea.

### ***Pestaña Aspecto de Evaluación***

Puede especificar las opciones de aspecto antes de crear gráficos.

Figura 5-85  
Configuración de la pestaña Aspecto de un nodo Evaluación



**Título.** Introduzca el texto que desea para el título del gráfico.

**Subtítulo.** Introduzca el texto que desea para el subtítulo del gráfico.

**Texto.** Permite aceptar la etiqueta generada automáticamente o seleccionar Personalizado para especificar una etiqueta.

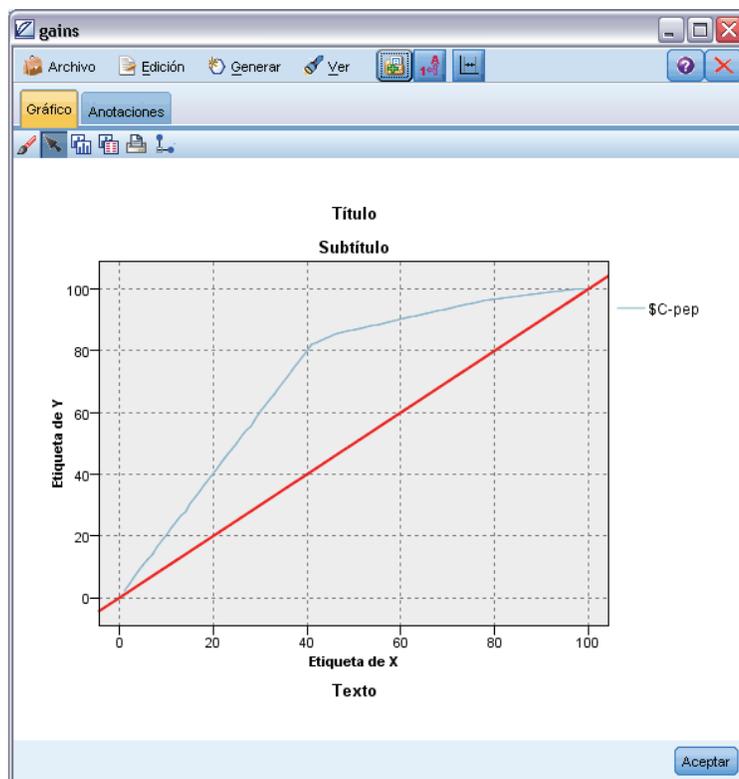
**Etiqueta de X.** Permite aceptar la etiqueta de eje  $x$  (horizontal) generada automáticamente o seleccionar Personalizado para especificar una etiqueta.

**Etiqueta de Y.** Permite aceptar la etiqueta del eje  $y$  (vertical) generada automáticamente o seleccionar Personalizado para especificar una etiqueta.

**Mostrar línea de cuadrícula.** Seleccionada por defecto, esta opción muestra una línea de cuadrícula detrás del gráfico o el diagrama que permite determinar con mayor facilidad los puntos de corte de la región y el tramo. Las líneas de cuadrícula siempre se muestran en blanco a menos que el fondo del gráfico sea blanco; en cuyo caso se muestran en gris.

El siguiente ejemplo muestra dónde se colocan las opciones de apariencia en el gráfico.

Figura 5-86  
Posición de las opciones de apariencia de gráfico en el gráfico Evaluación



### ***Lectura de los resultados de una evaluación del modelo***

La interpretación de un diagrama de evaluación depende hasta cierto punto del tipo de gráfico, pero hay algunas características comunes a todos los diagramas de evaluación. En el caso de los gráficos acumulados, las líneas más altas indican modelos mejores, especialmente en la parte izquierda del gráfico. En muchos casos, al comparar múltiples modelos las líneas se cruzan para que uno de ellos sea más alto en una parte del gráfico y otro sea más alto en otra parte distinta. En este caso, debe tener en cuenta la porción de la muestra que desee (que define un punto en el eje  $x$ ) al decidir el modelo que va a elegir.

La mayoría de los gráficos no acumulados será muy similar. Para obtener buenos modelos, los gráficos no acumulados deben ser altos en la parte izquierda del gráfico y bajos en la parte derecha. (Si un gráfico no acumulados muestra un patrón de sierra, puede suavizarlo reduciendo el número de cuantiles que se van a dibujar y volviendo a ejecutarlo.) Los descensos de la parte izquierda del gráfico o los picos de la parte derecha pueden indicar áreas donde el modelo tiene un pronóstico pobre. Una línea plana que cruza el gráfico indica un modelo que no proporciona ninguna información.

**Ganancias.** Los gráficos de ganancias acumulados siempre comienzan en el 0% y terminan en el 100% de izquierda a derecha. Si el modelo es bueno, el gráfico de ganancias irá subiendo vertiginosamente hacia el 100% y, a continuación, se estabilizará. Los modelos que no

proporcionan información mostrarán una diagonal desde la esquina inferior izquierda a la superior derecha (que se muestra en el gráfico si se selecciona Incluir línea base).

**Elevación.** Los gráficos de elevación acumulados tienden a comenzar por encima de 1,0 y descender gradualmente hasta alcanzar 1,0 de izquierda a derecha. El margen derecho del gráfico representa el conjunto de datos completo, de modo que el cociente de aciertos en cantidades acumuladas con respecto a los aciertos de datos sea 1,0. Para obtener un buen modelo válido, la elevación debe comenzar muy por encima de 1,0 en la izquierda, permanecer en una meseta alta a medida que se mueve hacia la derecha y, a continuación, descender rápidamente hacia el 1,0 en la parte derecha del gráfico. En el caso de un modelo que no proporciona información, la línea se mantendrá alrededor del 1,0 en todo el gráfico. (Si la opción Incluir línea base está seleccionada, aparece una línea horizontal en el 1,0 en el gráfico como referencia.)

**Respuestas.** Los gráficos de respuestas acumulados tienden a ser muy similares a los gráficos de elevación salvo por la escala. Los gráficos de respuestas suelen comenzar cerca del 100% y descender gradualmente hasta alcanzar la tasa de respuesta global (total de aciertos / total de registros) en el margen derecho del gráfico. Para obtener un buen modelo, la línea debe comenzar cerca de 100% en la izquierda, permanecer en una meseta alta a medida que se mueve a la derecha y, a continuación, descender rápidamente hacia la tasa de respuesta global en la parte derecha del gráfico. En el caso de un modelo que no proporciona información, la línea se mantendrá alrededor de la tasa de respuesta global en todo el gráfico. (Si la opción Incluir línea base está seleccionada, aparecerá una línea horizontal en la tasa de respuesta global en el gráfico como referencia.)

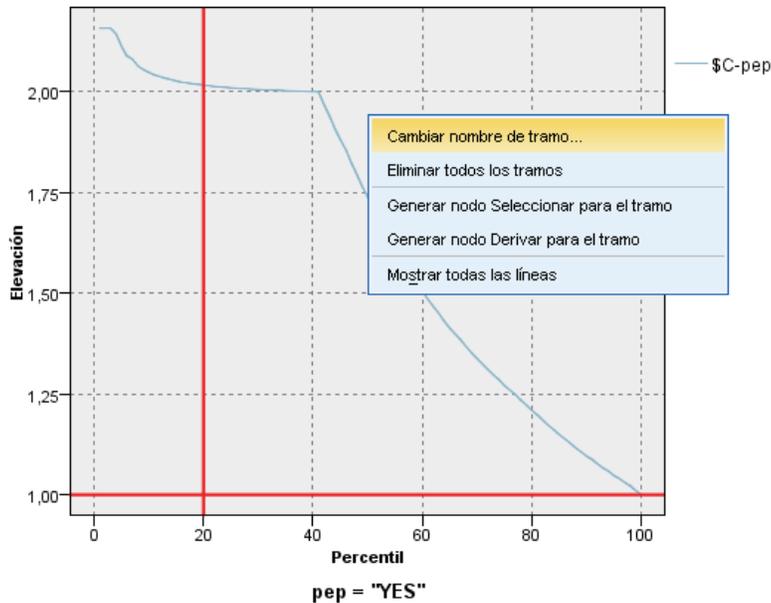
**Beneficios.** Los gráficos de beneficios acumulados muestran la suma de beneficios a medida que aumenta el tamaño de la muestra seleccionada, de izquierda a derecha. Los gráficos de beneficios suelen comenzar cerca del 0, aumentar rápidamente hacia la derecha hasta que alcanzan un pico o meseta en el centro y, a continuación, descender hacia el margen derecho del gráfico. Para obtener un modelo válido, los beneficios deben mostrar un pico bien definido en algún punto del centro del gráfico. En el caso de un modelo que no proporciona ninguna información, la línea será bastante recta y puede ascender, descender o permanecer en el mismo nivel en función de la estructura de costes/ingresos que se aplique.

**Rentabilidad de la inversión.** Los gráficos de rentabilidad de la inversión tienden a ser similares a los gráficos de respuestas y a los gráficos de elevación excepto por la escala. Los gráficos de rentabilidad de la inversión suelen comenzar por encima del 0% y descender gradualmente hasta alcanzar la rentabilidad de la inversión global de todo el conjunto de datos (que puede ser un valor negativo). Para obtener un buen modelo, la línea debe comenzar muy por encima de 0%, permanecer en una meseta alta a medida que se mueve a la derecha y, a continuación, descender rápidamente hacia la rentabilidad de la inversión global en la parte derecha del gráfico. En el caso de un modelo que no proporciona información, la línea debe mantenerse alrededor del valor de rentabilidad de la inversión global.

## ***Uso de un diagrama de evaluación***

El uso del ratón para explorar un diagrama de evaluación es similar al de un histograma o un gráfico de colección. El eje  $x$  representa puntuaciones de modelos en las cantidades especificadas, como los veintiles o los deciles.

Figura 5-87  
Cómo trabajar con un diagrama de evaluación

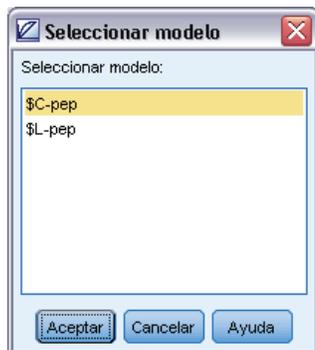


Puede dividir el eje  $x$  en tramos como lo haría con un histograma usando el icono de división, para mostrar opciones de división automática del eje en tramos iguales. Si desea obtener más información, consulte el tema [Exploración de gráficos](#) el p. 355. Puede editar los límites de los tramos manualmente seleccionando Tramos del gráfico en el menú Edición.

Una vez creado un diagrama de evaluación, definidos los tramos y examinados los resultados, puede utilizar las opciones del menú Generar y el menú contextual para crear nodos automáticamente en función de las selecciones del gráfico. Si desea obtener más información, consulte el tema [Generación de nodos a partir de gráficos](#) el p. 364.

Al generar nodos desde un diagrama de evaluación, se le solicitará que seleccione un modelo simple de entre todos los modelos disponibles en el gráfico.

Figura 5-88  
Selección de un modelo para la generación de nodos



Seleccione un modelo y pulse en Aceptar para generar el nuevo nodo en el lienzo de rutas.

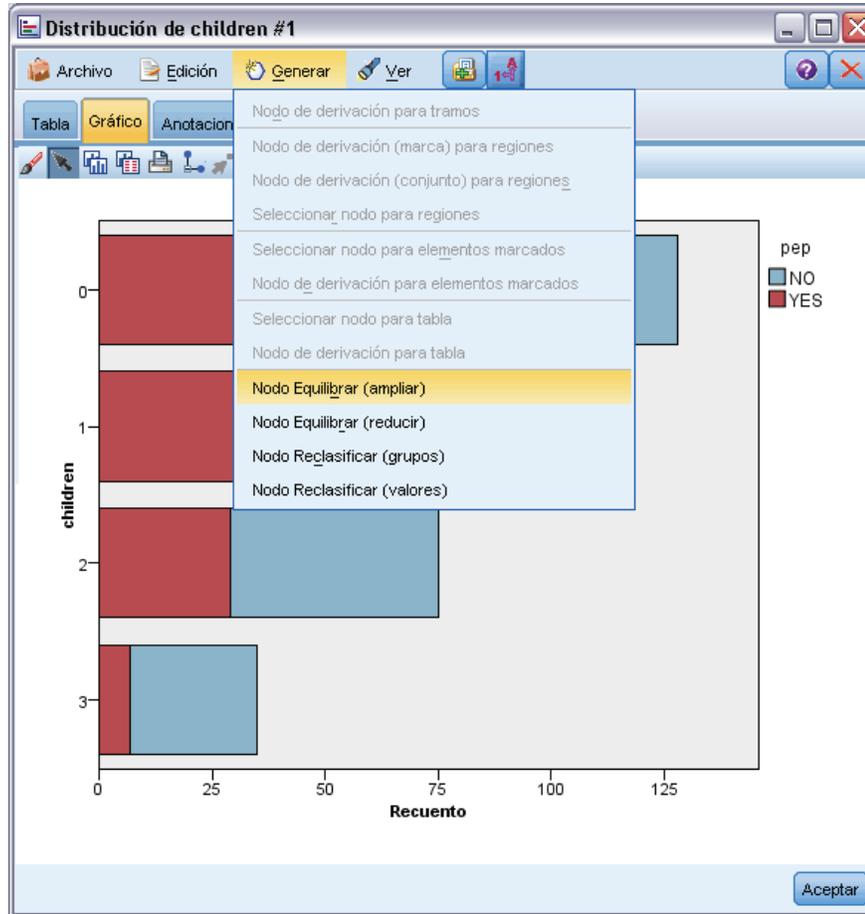
## ***Exploración de gráficos***

Mientras que el modo de edición le permite editar el diseño y apariencia del gráfico, Explorar modo permite explorar analíticamente los datos y valores que se representan en el gráfico. El objetivo principal de la exploración es analizar los datos e identificar los valores utilizando tramos, regiones y marcas para generar nodos Seleccionar, Derivar o Equilibrar. Para seleccionar este modo, seleccione Ver > Explorar modo de los menús (o pulse en el icono de la barra de herramientas).

Aunque algunos gráficos pueden utilizar todas las herramientas de exploración, otros sólo aceptan uno. Explorar modo incluye:

- Definición y edición de tramos, que se utilizan para dividir los valores en una escala del eje *x*. Si desea obtener más información, consulte el tema [Uso de tramos](#) el p. 356.
- Definición y edición de regiones, que se utilizan para identificar un grupo de valores en el área rectangular. Si desea obtener más información, consulte el tema [Usar regiones](#) el p. 360.
- Marcación y desmarcación de elementos para controlar la selección de los valores que se pueden utilizar para generar un nodo Seleccionar o Derivar. Si desea obtener más información, consulte el tema [Uso de elementos con marca](#) el p. 363.
- Generación de nodos utilizando los valores identificados por tramos, regiones, elementos con marca y enlaces Web que se utilizarán en la ruta. Si desea obtener más información, consulte el tema [Generación de nodos a partir de gráficos](#) el p. 364.

Figura 5-89  
Gráfico con el menú Generar



### Uso de tramos

En un gráfico con un campo de escala en el eje  $x$ , puede dibujar líneas de tramos verticales para dividir el rango de los valores en el eje  $x$ . Si un gráfico tiene múltiples paneles, una línea de tramo dibujada en un panel también se representa en el resto de paneles.

No todos los gráficos admiten tramos. Entre algunos de los gráficos que pueden tener tramos se incluyen: histogramas, gráficos de barras y distribuciones, gráficos (línea, dispersión, hora, etc.), colecciones y gráficos de evaluación. En gráficos con paneles, los tramos aparecen en todos los paneles. Y en algunos casos, en una SPLOM, verá una línea de tramo horizontal ya que el eje en el que el tramo de campo/variable se ha dibujado se ha desplazado.

Figura 5-90  
Gráfico con tres tramos

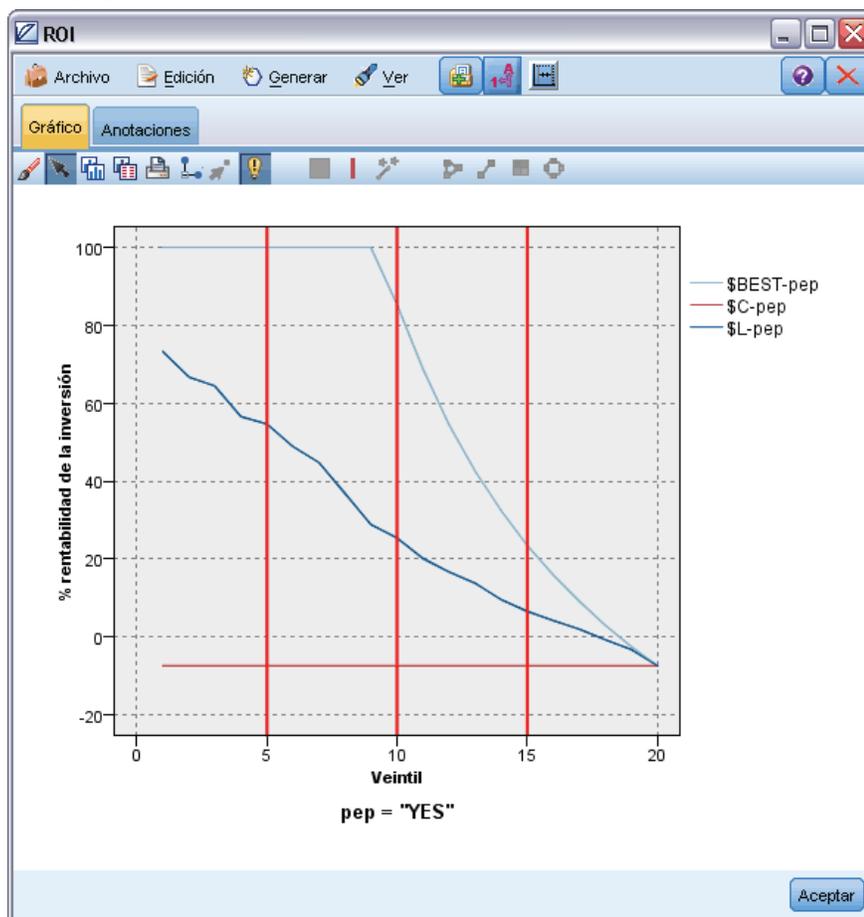
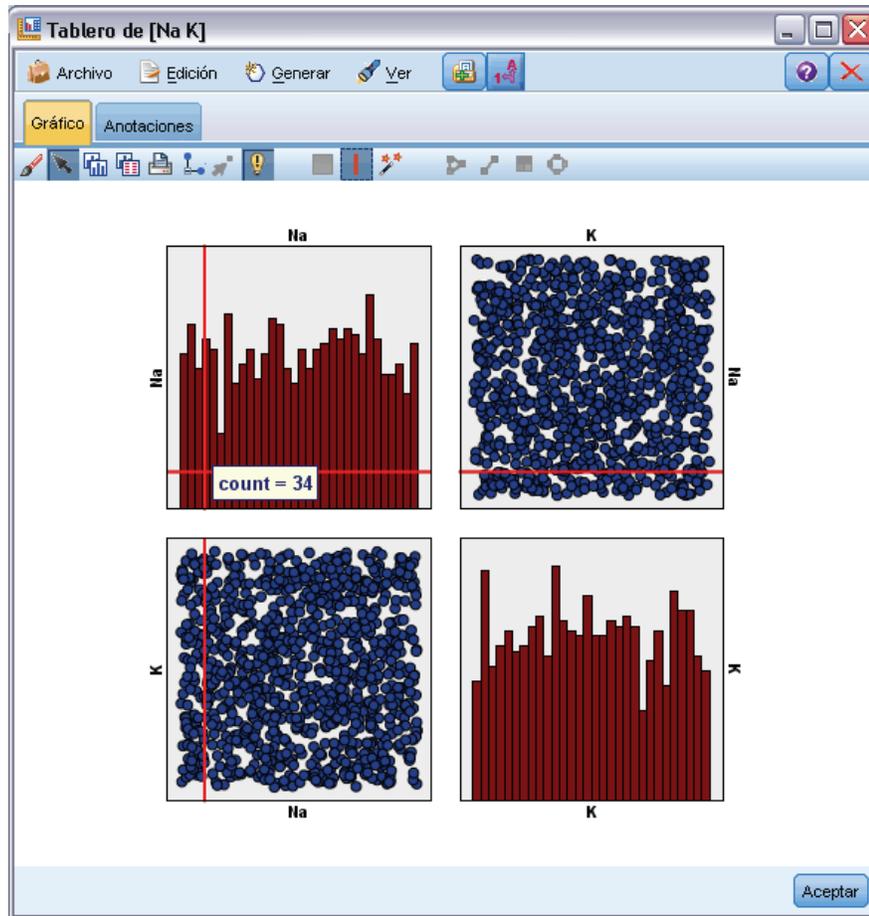


Figura 5-91  
SPLOM con tramos



### Definición de tramos

En un gráfico sin tramos, si añade una línea de tramo dividirá el gráfico en dos bandas. La línea de tramo representa el punto de inicio, que también se denomina límite inferior, del segundo tramo cuando lea el gráfico de izquierda a derecha. Del mismo modo, en un gráfico con dos tramos, si añade una línea de tramo dividirá uno de los tramos en dos, con lo que tendrá tres tramos. Por defecto, los tramos se denominan *tramoN*, donde *N* es el número de tramos de izquierda a derecha del eje *x*.

Una vez haya definido un tramo, puede arrastrar y soltar el tramo para volverlo a colocar en el eje *x*. Podrá ver más accesos directos si pulsa con el botón derecho en el tramo en tareas como cambiar el nombre, eliminar o generar nodos para ese tramo concreto.

### Para definir tramos:

- ▶ Compruebe que está en Explorar modo. Seleccione Ver > Explorar modo.
- ▶ En la barra de herramientas Explorar modo, pulse en el botón Dibujar tramo.

Figura 5-92  
Botón de la barra de herramientas Dibujar tramo



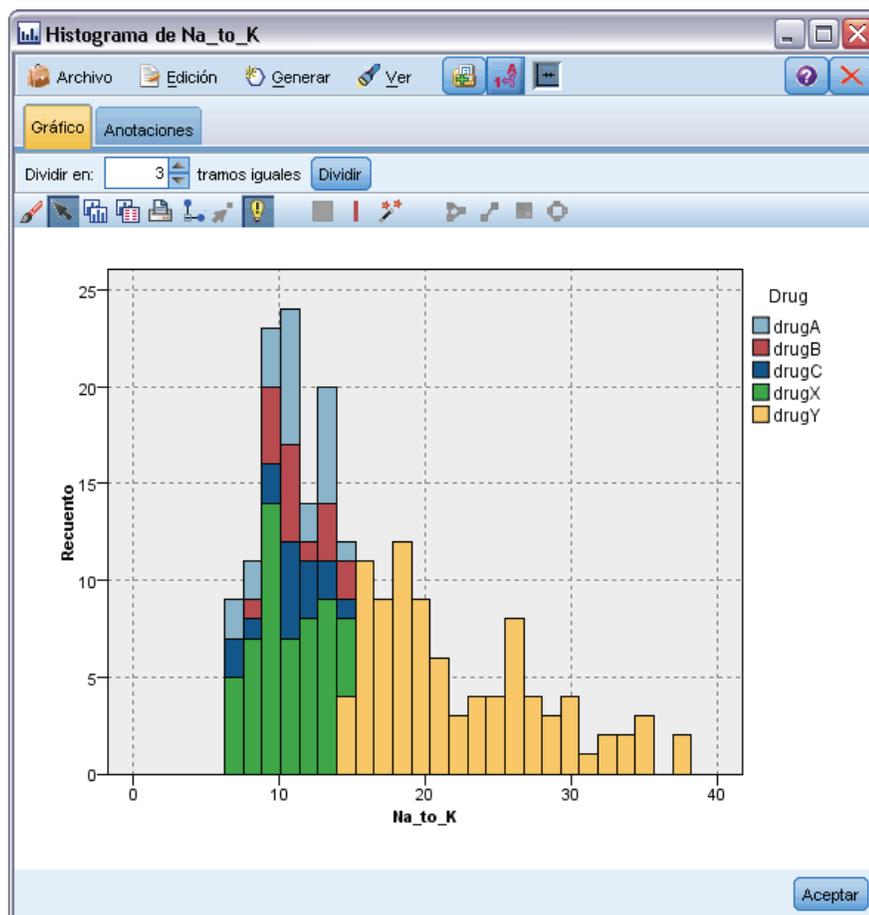
- En un gráfico que acepte tramos, pulse en el punto de valor del eje  $x$  en el que desea definir una línea de tramos.

*Nota:* También puede pulsar en el icono de la barra de herramientas Dividir el gráfico en tramos e introducir el número de tramos iguales que desea y pulsar en Dividir.

Figura 5-93  
Icono de división utilizado para expandir la barra de herramientas con opciones para dividir en tramos



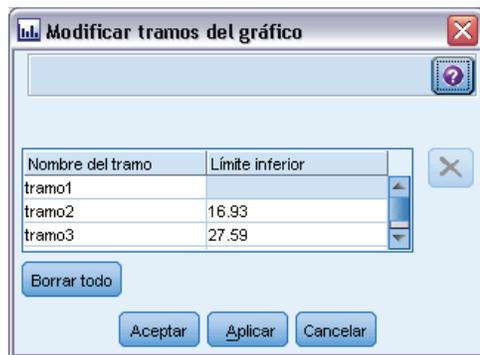
Figura 5-94  
Creación de la barra de herramientas de tramos iguales con tramos activados



### **Edición, cambio de nombre y eliminación de tramos**

Puede modificar las propiedades de tramos existentes en el cuadro de diálogo Modificar tramos del gráfico o mediante los menús contextuales del propio gráfico.

Figura 5-95  
Cuadro de diálogo Modificar tramos del gráfico



#### **Para modificar tramos:**

- ▶ Compruebe que está en Explorar modo. Seleccione Ver > Explorar modo.
- ▶ En la barra de herramientas Explorar modo, pulse en el botón Dibujar tramo.
- ▶ En los menús, seleccione Edición > Tramos del gráfico. Se abrirá el cuadro de diálogo Modificar tramos del gráfico.
- ▶ Si su gráfico contiene múltiples campos (como gráficos SPLOM), puede seleccionar el campo que desee en la lista desplegable.
- ▶ Añada un tramo nuevo introduciendo un nombre y límite inferior. Pulse la tecla Intro para iniciar una fila nueva.
- ▶ Edite el límite de un tramo ajustando el valor de Límite inferior.
- ▶ Cambie el nombre de un tramo introduciendo un nombre de tramo nuevo.
- ▶ Elimine un tramo seleccionando la línea en la tabla y pulsando en el botón de eliminación.
- ▶ Pulse en Aceptar para aplicar los cambios y cerrar el cuadro de diálogo.

*Nota:* También puede eliminar y cambiar el nombre de los tramos directamente en el gráfico pulsando con el botón derecho en la línea del tramo y seleccionando la opción de los menús contextuales.

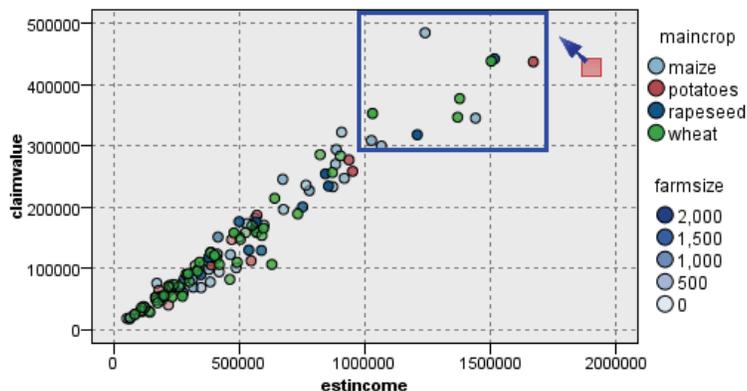
### **Usar regiones**

En cualquier gráfico con dos ejes de escala (o rango), puede dibujar regiones para agrupar valores en un área rectangular, denominada región. Una **región** es un área del gráfico que se describe por sus valores  $X$  e  $Y$  mínimo y máximo. Si un gráfico tiene múltiples paneles, una región dibujada en un panel también se representa en el resto de paneles.

No todos los gráficos admiten regiones. Entre los gráficos que aceptan regiones se incluyen: gráficos (línea, dispersión, burbujas, tiempo, etc.), SPLOM y colecciones. Estas regiones se dibujan en espacios de X,Y y no se pueden definir en gráficos 1-D, 3-D o animados. En gráficos con paneles, las regiones aparecen en todos los paneles. Con una matriz de diagrama de dispersión (SPLOM), aparecerá una región correspondiente en los gráficos superiores correspondientes pero no en los gráficos diagonales, ya que sólo muestran un campo de escala.

Figura 5-96

Definición de una región de valores de reclamación altos



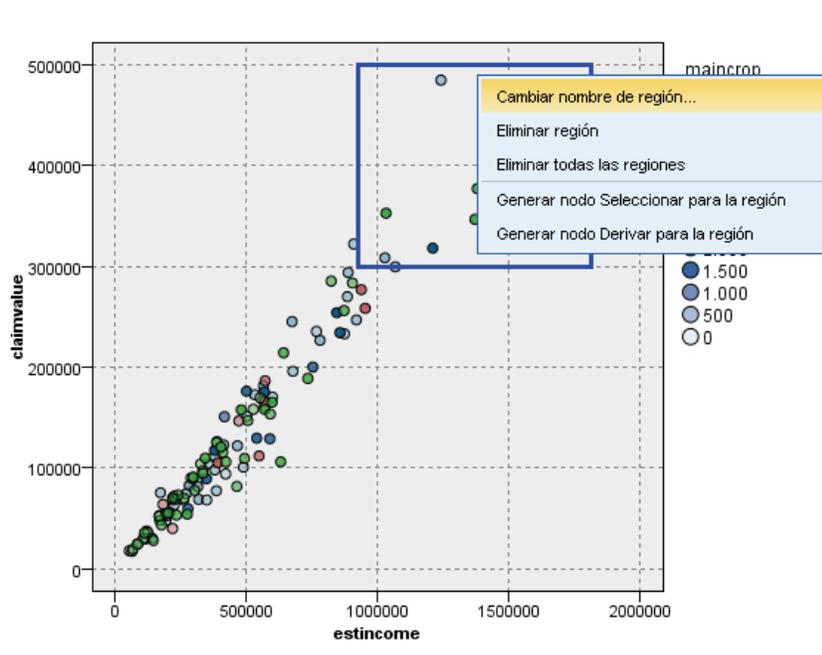
### Definición de regiones

Cuando crea una región, crea un grupo de valores. Por defecto, todas las regiones nuevas se denominan *Región<N>*, donde *N* corresponde al número de regiones ya creadas.

Una vez haya definido una región, pulse con el botón derecho en la línea de la región para acceder a algunos accesos directos básicos. Sin embargo, podrá ver más accesos directos si pulsa con el botón derecho en la región (no en la línea) en tareas como cambiar el nombre, eliminar o generar nodos Seleccionar y Derivar para esa región concreta.

Puede seleccionar subconjuntos de registros en función de su inclusión en una región determinada o en una de varias regiones. También puede incorporar información de región a un registro generando un nodo Derivar para marcar registros en función de su inclusión en una región. Si desea obtener más información, consulte el tema [Generación de nodos a partir de gráficos](#) el p. 364.

**Figura 5-97**  
*Exploración de la región de valores de reclamación altos*



**Para definir regiones:**

- ▶ Compruebe que está en Explorar modo. Seleccione Ver > Explorar modo.
- ▶ En la barra de herramientas Explorar modo, pulse en el botón Dibujar región.

**Figura 5-98**  
*Botón de la barra de herramientas Dibujar región*



- ▶ En un gráfico que acepte regiones, pulse el ratón y arrástrelo para dibujar la región rectangular.

***Edición, cambio de nombre y eliminación de regiones***

Puede modificar las propiedades de regiones existentes en el cuadro de diálogo Modificar regiones del gráfico o mediante los menús contextuales del propio gráfico.

Figura 5-99  
Especificación de propiedades de las regiones definidas



#### Para modificar regiones:

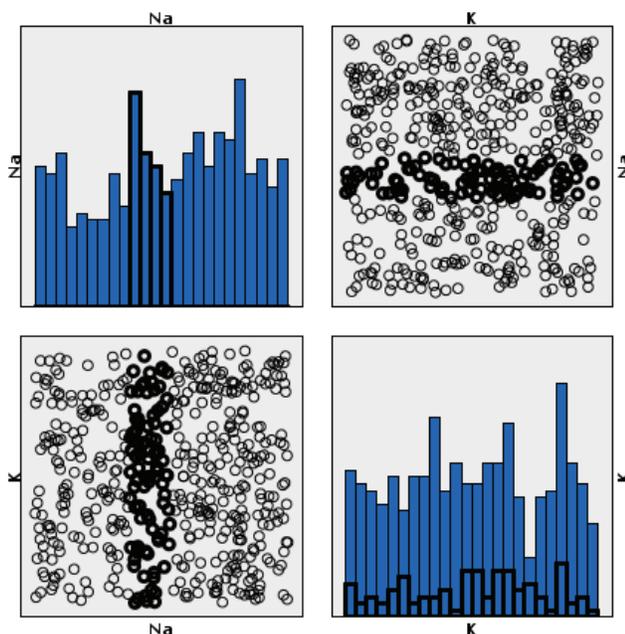
- ▶ Compruebe que está en Explorar modo. Seleccione Ver > Explorar modo.
- ▶ En la barra de herramientas Explorar modo, pulse en el botón Dibujar región.
- ▶ En los menús, seleccione Edición > Regiones del gráfico. Se abrirá el cuadro de diálogo Modificar regiones del gráfico.
- ▶ Si tiene múltiples campos en el gráfico (por ejemplo, gráficos SPLOM), debe definir el campo de la región en las columnas *Campo A* y *Campo B*.
- ▶ Añada una nueva región en una nueva línea introduciendo un nombre, seleccionando los campos de nombre (si procede) y definiendo los límites máximo y mínimo de cada campo. Pulse la tecla Intro para iniciar una fila nueva.
- ▶ Modifique los límites de la región existente ajustando los valores Min y Max de *A* y *B*.
- ▶ Cambie el nombre de una región cambiando el nombre de la región en la tabla.
- ▶ Elimine una región seleccionando la línea en la tabla y pulsando en el botón de eliminación.
- ▶ Pulse en Aceptar para aplicar los cambios y cerrar el cuadro de diálogo.

*Nota:* También puede eliminar y cambiar el nombre de las regiones directamente en el gráfico pulsando con el botón derecho en la línea de la región y seleccionando la opción de los menús contextuales.

### Uso de elementos con marca

Puede marcar elementos como barras, sectores y puntos en cualquier gráfico. Las líneas, áreas y superficies no se pueden marcar en gráficos, salvo en gráficos de tiempo, gráficos múltiples y gráficos de evaluación, ya que las líneas hacen referencia a los campos en esos casos. Siempre que marque un elemento, está resaltando todos los datos que representa ese elemento. En cualquier gráfico en que se represente el mismo caso en más de un lugar (como SPLOM), marcación es sinónimo de cepillado. Puede marcar elementos en gráficos y en tramos y regiones. Siempre que marque un elemento y vuelva a modo de edición, la marca seguirá siendo visible.

Figura 5-100  
 Marcación de elementos en una matriz SPLOM



Puede marcar y eliminar la marca pulsando en los elementos del gráfico. Cuando pulsa en un elemento para marcarlo por primera vez, el elemento aparecerá con un borde grueso para indicar que se ha marcado. Si vuelve a pulsar en el elemento, el borde desaparecerá y el elemento ya no estará marcado. Para marcar múltiples elementos, puede mantener pulsada la tecla Ctrl mientras pulsa los elementos o puede arrastrar el ratón por los elementos que desea marcar con la “varita mágica”. Recuerde que si pulsa en otra área o elemento sin mantener pulsada la tecla Ctrl, se eliminará la marca de todos los elementos marcados anteriormente.

Puede generar los nodos Seleccionar y Derivar de los elementos marcados en su gráfico. Si desea obtener más información, consulte el tema [Generación de nodos a partir de gráficos](#) el p. 364.

**Para marcar elementos:**

- ▶ Compruebe que está en Explorar modo. Seleccione Ver > Explorar modo.
- ▶ En la barra de herramientas Explorar modo, pulse en el botón Marcar elementos.
- ▶ Pulse en el elemento que necesita o pulse y arrastre el ratón para dibujar una línea alrededor de la región con varios elementos.

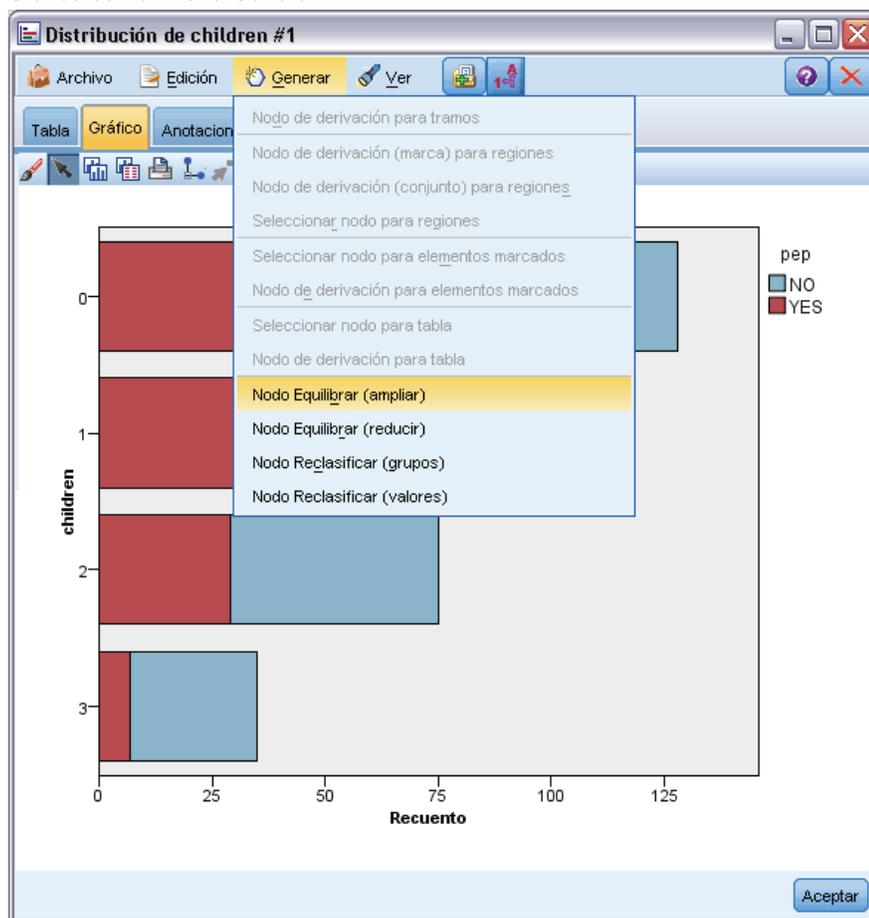
### **Generación de nodos a partir de gráficos**

Una de las funciones más útiles de los gráficos de IBM® SPSS® Modeler es la capacidad de generar nodos a partir de un gráfico o una selección en el gráfico. Por ejemplo, en un gráfico de tiempo, puede generar los nodos Derivar y Seleccionar basados en una selección o región de datos,

estableciendo los datos en subconjuntos de forma eficaz. Por ejemplo, puede utilizar esta potente función para identificar y excluir valores atípicos.

Cuando dibuje un tramo, también puede generar un nodo Derivar. En gráficos con dos ejes de escala, puede generar los nodos Seleccionar o Derivar de las regiones dibujadas en el gráfico. En gráficos con elementos marcados, puede generar nodos Derivar, Seleccionar y, en algunos casos, nodos Filtro a partir de estos elementos. Cualquier gráfico con una distribución de cuentas puede generar el nodo Equilibrar.

Figura 5-101  
Gráfico con el menú Generar



Siempre que genere un nodo, se coloca directamente en el lienzo de rutas, por lo que puede conectarlo a una ruta existente. Los siguientes nodos se pueden generar a partir de los gráficos: Seleccionar, Derivar, Equilibrar, Filtro y Reclasificar.

### **Nodos Seleccionar**

Los nodos Seleccionar se pueden generar para comprobar la inclusión de registros en una región y la exclusión de todos los registros ajenos a esa región o al revés para el procesamiento posterior.

- **Tramos.** Puede generar un nodo Seleccionar que incluya o excluya los registros de ese tramo. Nodo seleccionar para tramos únicamente sólo está disponible en los menús contextuales, ya que necesita seleccionar el tramo que va a utilizar en el nodo Seleccionar.
- **Regiones.** Puede generar un nodo Seleccionar que incluya o excluya los registros de una región.
- **Elementos marcados.** Puede generar nodos Seleccionar para capturar los registros correspondientes a los elementos marcados o enlaces de gráfico de mallas.

### ***Nodos Derivar***

Los nodos derivar se pueden generar a partir de regiones, tramos y elementos marcados. Todos los gráficos pueden producir nodos Derivar. En el caso de los gráficos de evaluación, se abre un cuadro de diálogo para seleccionar el modelo. En el caso de gráficos de malla, puede seleccionar los nodos Derivar (“Y”) y Derivar (“O”).

- **Tramos.** Puede generar un nodo Derivar que produzca una categoría para cada intervalo marcado en el eje, utilizando los nombres de tramos que aparecen en el cuadro de diálogo Editar tramos como nombres de categoría.
- **Regiones.** Puede generar un nodo Derivar (Derivar como marca) que crea un campo de marca denominado *en\_región* con las marcas definidas como *T* para los registros dentro de cualquier región y *F* para registros fuera de todas las regiones. También puede generar un nodo Derivar [Derivar como conjunto] que produce un conjunto con un valor para cada región con un nuevo campo denominado *región* para cada registro, que toma como su valor el nombre de la región en la que se encuentran los registros. Los registros que no están en ninguna región reciben el nombre de la región por defecto. Los nombres de los valores son los nombres de la región que aparecen en el cuadro de diálogo Editar regiones.
- **Elementos marcados.** Puede generar un nodo Derivar que calcula una marca que es *True* para todos los elementos marcados y *False* para el resto de registros.

### ***Nodos Equilibrar***

Los nodos Equilibrar se pueden generar para corregir desequilibrios en los datos, como reducir la frecuencia de valores comunes [utilice la opción de menú Nodo Equilibrar (reducir)] o para aumentar las instancias de los valores menos frecuentes [utilice la opción de menú Nodo Equilibrar (ampliar)]. La generación del nodo Equilibrar está activada para cualquier gráfico que muestra una distribución de cuentas, como Histograma, Puntos, Colección, Barras de recuentos, Sectores de recuentos y G. múltiple.

### ***Nodos Filtro***

Los nodos Filtro se pueden generar para cambiar el nombre o filtrar los campos en función de las líneas o nodos marcados en el gráfico. En el caso de los gráficos de evaluación, la mejor línea de ajuste no genera un nodo de filtro.

### ***Nodos Reclasificar***

Los nodos Reclasificar se pueden generar para volver a codificar los valores. Esta opción se utiliza para la distribución de gráficos. Puede generar un nodo Reclasificar para **grupos** para volver a codificar valores concretos de un campo representado en función de su inclusión en

un grupo (seleccione los grupos mediante Ctrl+pulsar en la pestaña Tablas). También puede generar un nodo Reclasificar para **valores** para volver a codificar datos en un conjunto existente de numerosos valores, como reclasificar datos en un conjunto estándar de valores para fusionar datos financieros de varias empresas para su análisis.

*Nota:* Si los valores se definen previamente, puede leerlos en SPSS Modeler como un archivo plano y utilizar una distribución para mostrar todos los valores. A continuación, puede generar un nodo Reclasificar (valores) para este campo directamente desde el gráfico. Si lo hace, colocará todos los valores objetivo en la columna *Valores nuevos* del nodo Reclasificar (lista desplegable).

### **Generación de nodos a partir de gráficos**

Puede utilizar el menú Generar en la ventana del resultado de gráfico para generar nodos. El nodo generado se colocará en el lienzo de rutas. Para utilizar el nodo, conéctelo a una ruta existente.

#### **Para generar un nodo a partir de un gráfico:**

- ▶ Compruebe que está en Explorar modo. Seleccione Ver > Explorar modo.
- ▶ En la barra de herramientas Explorar modo, pulse en el botón Región.
- ▶ Defina los tramos, regiones o cualquier elemento marcado que necesite para generar el nodo.
- ▶ En el menú Generar, seleccione el tipo de nodo que desea producir. Sólo están activados los que son posibles.

*Nota:* También puede generar nodos directamente del gráfico pulsando con el botón derecho en el gráfico y seleccionando la opción de generación que desee de los menús contextuales.

## **Edición de visualizaciones**

Mientras el modo de exploración le permite explorar analíticamente los datos y valores representados por la visualización, el modo de edición le permite cambiar el diseño y aspecto de la visualización. Por ejemplo, puede cambiar las fuentes y colores para que coincidan con el manual de estilo de su organización. Para seleccionar este modo, seleccione Ver > Modo edición de los menús (o pulse en el icono de la barra de herramientas).

En el modo de edición hay varias barras de herramientas que afectan a distintos aspectos del diseño de la visualización. Si no utiliza algunas de ellas, puede ocultarlas para ampliar el espacio del cuadro de diálogo en el que aparece el gráfico. Para seleccionar o anular la selección de las barras de herramientas, pulse en el nombre de la barra de herramientas correspondiente en el menú Ver.

*Nota:* Para añadir más detalle a sus visualizaciones, puede aplicar un título, una nota al pie y etiquetas de ejes. Si desea obtener más información, consulte el tema [Adición de títulos y notas al pie](#) el p. 384.

Tiene varias opciones para editar una visualización en el **Modo de edición**. Tiene la posibilidad de:

- Editar texto y darle formato.
- Cambiar el color de relleno, la transparencia y el patrón de marcos y elementos gráficos.
- Cambiar el color y el estilo de guión de los bordes y las líneas.

- Girar y cambiar la forma y la relación de aspecto de los elementos de punto.
- Cambiar el tamaño de los elementos gráficos (como barras y puntos).
- Ajustar el espacio que rodea a los elementos mediante márgenes y relleno.
- Especificar un formato para los números.
- Cambiar la configuración del eje y la escala.
- Ordenar, excluir y contraer categorías en un eje categórico.
- Definir la orientación de los paneles.
- Aplicar transformaciones a un sistema de coordenadas.
- Cambiar estadísticos, tipos de elementos gráficos y modificadores de colisión.
- Cambiar la posición de la leyenda.
- Aplicar hojas de estilo de visualización.

Los siguientes temas describen cómo se pueden realizar estas distintas tareas. También se recomienda leer las reglas generales de edición de gráficos.

#### ***Para cambiar al modo de edición***

- ▶ Elija en los menús:  
Ver > Modo de edición

## ***Reglas generales de edición de visualizaciones***

### ***Modo de edición***

Todas las ediciones se realizan en modo de edición. Para activarlo, elija en los menús:  
Ver > Modo de edición

### ***Selección***

Las opciones disponibles para edición dependen de la selección. Se activan diferentes opciones de la barra de herramientas y la paleta de propiedades según lo que se seleccione. Sólo se aplican a la selección actual los elementos activados. Por ejemplo, si se selecciona un eje, las pestañas Escalas, Marca mayor y Marcas menores estarán disponibles en la paleta de propiedades.

A continuación mostramos algunos consejos para seleccionar elementos de la visualización:

- Pulse en un elemento para seleccionarlo.
- Seleccione un elemento gráfico (como los puntos de un diagrama de dispersión o las barras de un gráfico de barras) pulsando una vez. Tras la selección inicial, vuelva a pulsar para reducir la selección a grupos de elementos gráficos o a un único elemento gráfico.
- Pulse en Esc para anular la selección de todo.

### **Paletas**

Cuando se selecciona un elemento de la visualización, las diferentes paletas se actualizan para reflejar la selección. Las paletas contienen controles para realizar las modificaciones en la selección. Las paletas pueden ser barras de herramientas o un panel con múltiples controles y pestañas. Las paletas pueden estar ocultas, por lo que hay que asegurarse de que se muestra la paleta adecuada para realizar las ediciones. Compruebe en el menú Ver las paletas que se muestran actualmente.

Puede cambiar la posición de las paletas pulsando y arrastrando el espacio vacío de una paleta de barra de herramientas o en la parte izquierda de otras paletas. La información visual le permite saber dónde anclar la paleta. En el caso de paletas que no son de la barra de herramientas, también puede hacer en el botón cerrar para ocultar la paleta y en el botón desanclar para mostrar la paleta en una ventana diferente. Pulse el botón ayuda para mostrar la ayuda de la paleta específica.

### **Configuración automática**

Algunos parámetros de configuración incluyen la opción *-automático-*. Esto indica que se aplican valores automáticos. Los ajustes automáticos que se utilizarán dependen de la visualización específica y los valores de los datos. Puede introducir un valor que sustituya a la configuración automática. Si desea restaurar la configuración automática, elimine el valor actual y pulse en Entrar. El parámetro volverá a mostrar la opción *-automático-*.

### **Eliminación y ocultación de elementos**

Puede eliminar u ocultar varios elementos en la visualización. Por ejemplo, puede ocultar la leyenda o la etiqueta del eje. Seleccione un elemento y pulse la tecla Supr para eliminarlo. Si no se permite eliminar ese elemento, no se realizará ninguna acción. Si elimina un elemento por accidente, pulse Ctrl+Z para deshacer la eliminación.

### **State**

Algunas barras de herramientas muestran el estado de la selección actual, mientras que otras no. La paleta de propiedades siempre muestra el estado. Si una barra de herramientas *no* muestra el estado, se indicará en el tema que describe dicha barra de herramientas.

## **Edición y formato de texto**

Puede editar texto directamente y cambiar el formato de un bloque de texto entero. Tenga en cuenta que no podrá editar texto directamente enlazado a valores de datos. Por ejemplo, no podrá editar una etiqueta de marca, ya que su contenido se deriva de los datos subyacentes. Sin embargo, puede formatear cualquier texto de la visualización.

### **Cómo editar texto directamente**

- Pulse dos veces en el bloque de texto. De esta forma se selecciona todo el texto. Todas las barras de texto se desactivarán en este momento, ya que no podrá cambiar ninguna otra parte de la visualización mientras edita el texto.

- ▶ Escriba sobre el texto que desea sustituir. También puede volver a pulsar en el texto para que aparezca un cursor. Coloque el cursor en la posición deseada y escriba el texto adicional.

### ***Cómo dar formato al texto***

- ▶ Seleccione el marco que contiene el texto. No pulse dos veces en el texto.
- ▶ Use la barra de herramientas de fuentes para dar formato al texto. Si la barra de herramientas no está activada, asegúrese de seleccionar sólo el *marco* que contiene el texto. Si el texto está seleccionado, se desactivará la barra de herramientas.

Figura 5-102

*Barra de herramientas de fuentes*



Puede cambiar distintas características de la fuente:

- Color
- Familia (por ejemplo, Arial o Verdana)
- Tamaño (la unidad es pt, a menos que se indique otra distinta, como pc)
- Ponderación
- Alineación relativa al marco del texto

El formato se aplica a todo el texto de un marco. No puede cambiar el formato de letras o palabras individuales de ningún bloque de texto concreto.

## ***Cambio de colores, tramas, estilos de guión y transparencia***

Muchos elementos diferentes de una visualización tienen un relleno y un borde. El ejemplo más obvio es la barra de un gráfico de barras. El color de las barras es el color de relleno. También pueden llevar un borde negro y continuo alrededor.

Hay otros elementos menos obvios en la visualización que tienen colores de relleno. Si el color de relleno es transparente, es posible que no se vea que existe dicho relleno. Tomemos como ejemplo el texto de una etiqueta de eje. Da la sensación de que el texto es “flotante”, pero en realidad aparece en un marco con color de relleno transparente. Puede ver el marco si selecciona la etiqueta del eje.

Cualquier marco de la visualización puede tener un estilo de relleno y borde, incluyendo el marco alrededor de toda la visualización. Además, cualquier relleno tiene un nivel de opacidad/transparencia asociado que puede ajustarse.

### ***Para cambiar de colores, tramas, estilos de guión y transparencia***

- ▶ Seleccione el elemento cuyo formato desea cambiar. Por ejemplo, seleccione las barras de un gráfico de barras o un marco que contenga texto. Si la visualización está dividida por un campo o variable categórica, puede seleccionar también el grupo que corresponda a una categoría individual. De esta forma podrá cambiar el aspecto estético asignado por defecto a dicho grupo. Por ejemplo, puede cambiar el color de uno de los grupos de apilado de un gráfico de barras apiladas.

- Use la barra de herramientas de colores para cambiar el color de relleno o del borde, o bien la trama de relleno.

Figura 5-103  
Barra de herramientas de colores

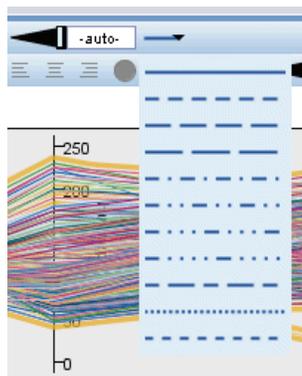


*Nota:* esta barra de herramientas no muestra el estado de la selección actual.

Para cambiar un color o relleno, puede pulsar el botón para seleccionar la opción mostrada o en la flecha desplegable para seleccionar otra opción. En el caso de los colores, observe que hay uno representado como blanco y atravesado por una línea diagonal roja. Se trata del color transparente. Puede usarlo, por ejemplo, para ocultar los bordes de las barras de un histograma.

- El primer botón controla el color de relleno.
  - El segundo botón controla el color del borde.
  - El tercer botón controla la trama de relleno. La trama de relleno utiliza el color del borde. Por tanto, la trama de relleno sólo se ve si el color del borde es visible.
  - El cuarto control es una barra deslizante y un cuadro de texto que controlan la opacidad del color y el patrón de relleno. Si el porcentaje es menor, habrá menor opacidad y más transparencia. 100% significa que es totalmente opaco (sin transparencia).
- Para cambiar el estilo de guión de un borde o una línea, use la barra de herramientas de líneas.

Figura 5-104  
Barra de herramientas de líneas



*Nota:* esta barra de herramientas no muestra el estado de la selección actual.

Al igual que con la otra barra de herramientas, puede pulsar en el botón para seleccionar la opción mostrada o en la flecha desplegable para seleccionar otra opción.

## ***Rotación y cambio de la forma y la relación de aspecto de los elementos de punto***

Puede girar elementos de punto, asignar una forma predefinida diferente o cambiar la relación de aspecto (la relación entre anchura y altura).

### **Cómo modificar elementos de punto**

- ▶ Seleccione los elementos de punto. No puede girar ni cambiar la forma y relación de aspecto de elementos de punto individuales.
- ▶ Use la barra de herramientas de símbolos para modificar los puntos.

Figura 5-105  
Barra de herramientas de símbolos



- El primer botón permite cambiar la forma de los puntos. Pulse en la flecha desplegable y seleccione una forma predefinida.
- Con el segundo botón puede girar los puntos hasta una posición específica como si lo hiciera con un compás. Pulse en la flecha desplegable y, a continuación, arrastre la punta hasta la posición deseada.
- El tercero permite cambiar la relación de aspecto. Pulse en la flecha desplegable y, a continuación, arrastre el rectángulo que aparece. La forma del rectángulo representa la relación de aspecto.

## **Cambio de tamaño de los elementos gráficos**

Puede cambiar el tamaño de los elementos gráficos de la visualización. Estos incluyen barras, líneas y puntos entre otros. Si el tamaño del elemento gráfico depende de una variable o un campo, el tamaño especificado es el *mínimo*.

### **Cómo cambiar el tamaño de los elementos gráficos**

- ▶ Seleccione los elementos gráficos cuyo tamaño desea cambiar.
- ▶ Use el control deslizante o introduzca un tamaño específico para la opción disponible en la barra de herramientas de símbolos. La unidad es píxeles, a menos que se indique otra distinta (consulte a continuación la lista completa de abreviaturas de unidad). También puede especificar un porcentaje (como 30 %), lo que significa que un elemento gráfico usa el porcentaje especificado del espacio disponible. El espacio disponible dependerá del tipo de elemento gráfico y la visualización específica.

Tabla 5-3  
Abreviaturas de unidad válidas

Abreviatura	Unidad
cm	centímetro
en	pulgada
mm	milímetro
pc	pica
pt	punto
px	pixel

Figura 5-106  
Control del tamaño en la barra de herramientas de símbolos



## Especificación de márgenes y relleno

Si hay demasiado espacio o un espacio insuficiente alrededor o dentro de un marco de la visualización, puede cambiar este margen y los ajustes de relleno. El **margen** es la cantidad de espacio existente entre el marco y otros elementos que lo rodean. El **relleno** es la cantidad de espacio existente entre el borde del marco y el *contenido* de dicho marco.

### Cómo especificar los márgenes y el relleno

- ▶ Seleccione el marco cuyos márgenes y relleno desea especificar. Puede tratarse de un marco de texto, el marco que rodea a la leyenda o incluso el marco de datos que muestra los elementos gráficos (como barras y puntos).
- ▶ Use la pestaña Márgenes de la paleta de propiedades para especificar la configuración. Todos los tamaños se expresan en píxeles, a menos que se indique una unidad distinta (como cm o pulg).

Figura 5-107  
Pestaña Márgenes



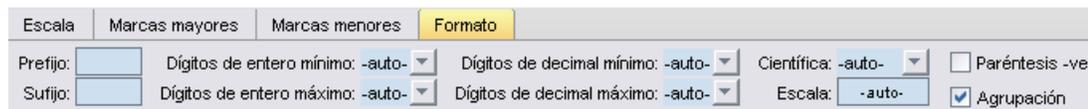
## Formato de numeración

Puede especificar el formato de los números de las etiquetas señalizadoras en el eje continuo o las etiquetas de valor de datos que muestran un número. Por ejemplo, puede especificar que los números que aparecen en las etiquetas de marcas se muestren en miles.

### Para especificar formatos de numeración

- ▶ Seleccione las etiquetas señalizadoras del eje continuo o las etiquetas de valor de datos si contienen números.
- ▶ Pulse en la pestaña Formato de la paleta Propiedades.

Figura 5-108  
Pestaña Formato



- Seleccione las opciones de formato de numeración que desee:

**Prefijo.** Carácter que se mostrará al comienzo del número. Introduzca, por ejemplo, un signo de dólar (\$) si los números se refieren a salarios en dólares estadounidenses.

**Sufijo.** Carácter que se mostrará al final del número. Introduzca, por ejemplo, un signo de porcentaje (%) si los números equivalen a porcentajes.

**Dígitos de entero mínimo.** Número mínimo de dígitos que mostrar en la parte entera de una representación decimal. Si el valor real no contiene el número mínimo de dígitos, la parte entera del valor se rellenará con ceros.

**Dígitos de entero máximo.** Número máximo de dígitos que mostrar en la parte entera de una representación decimal. Si el valor real supera el número máximo de dígitos, la parte entera del valor se rellenará con asteriscos.

**Dígitos de decimal mínimo.** Número mínimo de dígitos que mostrar en la parte decimal de una representación decimal o científica. Si el valor real no contiene el número mínimo de dígitos, la parte decimal del valor se rellenará con ceros.

**Dígitos de decimal máximo.** Número máximo de dígitos que mostrar en la parte decimal de una representación decimal o científica. Si el valor real supera el número mínimo de dígitos, la parte decimal se redondeará al número adecuado de dígitos.

**Científica.** Determina si se muestran o no los números en notación científica. La notación científica resulta útil para números muy grandes o muy pequeños. -auto- permite a la aplicación determinar cuándo es adecuado utilizar la notación científica.

**Escala.** Un factor de escala, que es un número por el que se divide el valor original. Utilice un factor de escala si los números son grandes, pero no desea que la etiqueta se extienda demasiado para adaptarse al número. Si cambia el formato de numeración de las etiquetas señalizadoras, asegúrese de editar el título del eje para indicar cómo se debe interpretar el número. Supongamos, por ejemplo, que el eje de escala muestra salarios y las etiquetas son 30.000, 50.000 y 70.000. Puede introducir un factor de escala de 1000 para mostrar 30, 50 y 70. A continuación, debe editar el título del eje para incluir el texto en miles.

**Paréntesis -ve.** Indica si deben incluirse paréntesis alrededor de los valores negativos.

**Agrupación.** Indica si se debe mostrar un carácter entre grupos de dígitos. La configuración regional actual de su equipo determinará qué carácter se utiliza para la agrupación de dígitos.

## ***Cambio de la configuración del eje y la escala***

Existen varias opciones para modificar ejes y escalas.

### ***Cómo cambiar la configuración del eje y la escala***

- Seleccione cualquier parte del eje (por ejemplo, la etiqueta del eje o las etiquetas de marca).
- Use las pestañas Escalas, Marca mayor y Marcas menores de la paleta de propiedades para cambiar la configuración del eje y la escala.

Figura 5-109  
Paleta de propiedades



### Pestaña Escalas

Nota: La pestaña Escalas no aparece para gráficos en los que los datos estén agregados previamente (por ejemplo, histogramas).

**Tipo.** Especifica si la escala es lineal o transformada. Las transformaciones de escala permiten comprender los datos o realizar los supuestos necesarios para la inferencia estadística. En los diagramas de dispersión, puede utilizar una escala transformada si la relación entre las variables o los campos independientes y dependientes es no lineal. Las transformaciones de escala también se pueden utilizar para lograr que un histograma asimétrico sea más simétrico y se parezca a una distribución normal. Tenga en cuenta que sólo se transforma la escala en que se muestran los datos; los datos reales no se transforman.

- **Lineal.** Especifica una escala lineal sin transformar.
- **Log.** Especifica una escala transformada de logaritmo en base 10. Para adaptarse a los valores cero y negativos, esta transformación usa una versión modificada de la función logarítmica. Esta función “logaritmo seguro” se define como  $\text{sign}(x) * \log(1 + \text{abs}(x))$ . De modo que  $\text{safeLog}(-99)$  es igual a:

$$\text{sign}(-99) * \log(1 + \text{abs}(-99)) = -1 * \log(1 + 99) = -1 * 2 = -2$$

- **potencia** Especifica una escala transformada de potencia con un exponente de 0,5. Para adaptarse a los valores negativos, esta transformación usa una versión modificada de la función exponencial. Esta función “potencia segura” se define como  $\text{sign}(x) * \text{pow}(\text{abs}(x), 0.5)$ . De modo que  $\text{safePower}(-100)$  es igual a:

$$\text{sign}(-100) * \text{pow}(\text{abs}(-100), 0.5) = -1 * \text{pow}(100, 0.5) = -1 * 10 = -10$$

**Mín/Máx/Bajo redondeado/Alto redondeado.** Especifica el rango de la escala. Si selecciona Bajo redondeado y Alto redondeado, la aplicación podrá seleccionar una escala apropiada de acuerdo con los datos. Los valores mínimo y máximo están redondeados porque suelen ser valores enteros mayores o menores que los valores de datos máximos y mínimos. Por ejemplo, si los datos oscilan entre 4 y 92, los valores alto y bajo redondeados para la escala pueden estar comprendidos entre 0 y 100 en lugar de los valores mínimo y máximo reales de los datos. Preste atención para no definir un rango demasiado pequeño que oculte elementos importantes. Tenga en cuenta también que no puede definir un valor mínimo y máximo explícito si está seleccionada la opción Incluir cero.

**Margen bajo/Margen alto.** Crea márgenes en los extremos inferior y/o superior del eje. El margen aparece en posición perpendicular con respecto al eje seleccionado. La unidad es píxeles, a menos que se indique otra distinta (como cm o pulg). Por ejemplo, si define Margen alto en 5 para el eje vertical, se extenderá un margen horizontal de 5 px a lo largo de la parte superior del marco de datos.

**Inversa.** Especifica si la escala está invertida.

**Incluir cero.** Indica que la escala debe incluir el valor 0. Esta opción suele utilizarse en gráficos de barras para asegurarse de que las barras empiezan en el 0, y no en un valor cercano a la altura de la barra más pequeña. Si se selecciona esta opción, Mín y Máx se desactivan ya que no se pueden definir los valores mínimo y máximo del rango de la escala.

### ***Pestañas Marca mayor/Marcas menores***

Las **marcas** o **marcas de comprobación** son líneas que aparecen sobre un eje. Indican valores en intervalos o categorías específicos. Las **marcas mayores** son marcas de comprobación con etiquetas. Son también más largas que las otras marcas de comprobación. Las **marcas menores** son marcas de comprobación que aparecen entre las marcas de comprobación mayores. Algunas opciones son específicas del tipo de marca, pero la mayoría de las opciones están disponibles para marcas mayores y menores.

**Mostrar marcas.** Especifica si aparecen marcas mayores o menores en el gráfico.

**Mostrar cuadrículas.** Especifica si las cuadrículas aparecen en las marcas mayores o menores. Las **cuadrículas** son líneas que atraviesan todo el gráfico de eje a eje.

**Posición.** Especifica la posición de las marcas de comprobación en relación con el eje.

**Longitud.** Especifica la longitud de las marcas de comprobación. La unidad es píxeles, a menos que se indique otra distinta (como cm o pulg).

**Base.** *Sólo se aplica a las marcas mayores.* Especifica el valor en el que aparece la primera marca mayor.

**Delta.** *Sólo se aplica a las marcas mayores.* Especifica la diferencia entre marcas mayores. Esto es, las marcas mayores aparecerán cada  $n$  valores, donde  $n$  es el valor delta.

**Divisiones.** *Sólo se aplica a las marcas menores.* Especifica el número de divisiones de marcas menores existentes entre las marcas mayores. El número de marcas menores es igual al número de divisiones menos uno. Por ejemplo, supongamos que hay marcas mayores en 0 y 100. Si introduce 2 como número de divisiones de marcas menores, habrá *una* marca menor en 50, dividiendo el intervalo de 0 a 100, lo que creará *dos* divisiones.

## ***Edición de categorías***

Puede editar las categorías en un eje categórico de varias formas:

- Cambie el orden de clasificación para mostrar las categorías.
- Excluya categorías específicas.
- Añada una categoría que no aparezca en el conjunto de datos.
- Contraiga/combine categorías pequeñas en una categoría.

### ***Cómo cambiar el orden de clasificación de categorías***

- ▶ Seleccione un eje categórico. La paleta Categorías muestra las categorías en el eje.

*Nota:* Si la paleta no está visible, asegúrese de haberla activado. En el menú Ver de IBM® SPSS® Modeler, seleccione Categorías.

- ▶ En la paleta Categorías, seleccione una opción de ordenación de la lista desplegable.

**Personalizado.** Ordena las categorías basándose en el orden en el que aparecen en la paleta. Utilice los botones de flecha para mover categorías a la parte superior de la lista, arriba, abajo y a la parte inferior de la lista.

**Datos.** Ordena las categorías basándose en el orden en el que aparecen en el conjunto de datos.

**Nombre.** Ordena las categorías alfabéticamente, utilizando los nombres como aparecen en la paleta. Esto puede ser tanto el valor como la etiqueta, dependiendo de si está seleccionado el botón de la barra de herramientas para mostrar valores y etiquetas.

**Valor.** Ordena las categorías según el valor de datos subyacente, utilizando los valores que aparecen entre paréntesis en la paleta. Sólo los orígenes de datos con metadatos (como los archivos de datos de IBM® SPSS® Statistics) admiten esta opción.

**Estadístico.** Ordena las categorías basándose en el estadístico calculado para cada categoría. Algunos ejemplos de estadísticos son recuentos, porcentajes y medias. Esta opción sólo está disponible si se utiliza un estadístico en el gráfico.

#### **Para añadir una categoría**

Por defecto, únicamente las categorías que aparecen en el conjunto de datos están disponibles. Puede añadir una categoría a la vista si fuera necesario.

- ▶ Seleccione un eje categórico. La paleta Categorías muestra las categorías en el eje.

*Nota:* Si la paleta no está visible, asegúrese de haberla activado. En el menú Ver de SPSS Modeler, seleccione Categorías.

- ▶ En la paleta de categorías, pulse en el botón Añadir categoría:

Figura 5-110  
Botón Añadir categoría



- ▶ En el cuadro de diálogo Añadir una nueva categoría, introduzca un nombre para la categoría.
- ▶ Pulse en Aceptar.

#### **Cómo excluir categorías específicas**

- ▶ Seleccione un eje categórico. La paleta Categorías muestra las categorías en el eje.

*Nota:* Si la paleta no está visible, asegúrese de haberla activado. En el menú Ver de SPSS Modeler, seleccione Categorías.

- ▶ En la paleta Categorías, seleccione un nombre de categoría en la lista Incluir y, a continuación, pulse en el botón X. Para mover la categoría hacia atrás, seleccione su nombre en la lista Excluido y, a continuación, pulse en la flecha situada a la derecha de la lista.

### ***Cómo contraer/combinar categorías pequeñas***

Puede combinar categorías que sean tan pequeñas que no necesite mostrarlas por separado. Por ejemplo, si tiene un gráfico de sectores con muchas categorías, considere la opción de contraer categorías con un porcentaje inferior a 10. Sólo se pueden contraer las estadísticas aditivas. Por ejemplo, no puede añadir medias porque las medias no son aditivas. Por lo tanto, no se pueden combinar/contraer categorías utilizando una media.

- ▶ Seleccione un eje categórico. La paleta Categorías muestra las categorías en el eje.

*Nota:* Si la paleta no está visible, asegúrese de haberla activado. En el menú Ver de SPSS Modeler, seleccione Categorías.

- ▶ En la paleta Categorías, seleccione Contraer y especifique un porcentaje. Cualquier categoría cuyo porcentaje del total sea inferior al número especificado se combinará en una categoría. El porcentaje se basa en el estadístico mostrado en el gráfico. La contracción sólo está disponible para estadísticos basados en recuentos y de suma.

### ***Cambio de la orientación de paneles***

Si utiliza paneles en su visualización, puede cambiar su orientación.

#### ***Cómo cambiar la orientación de los paneles***

- ▶ Seleccione cualquier parte de la visualización.
- ▶ Pulse en la pestaña Paneles de la paleta Propiedades.

Figura 5-111  
Pestañas Paneles



- ▶ Seleccione una opción en Diseño:

**Tabla.** Organiza los paneles como una tabla, de manera que hay una fila o columna asignada a cada valor individual.

**Transpuesto.** Organiza los paneles como una tabla, aunque también intercambia las filas y columnas originales. Esta opción no equivale a transponer el gráfico en sí. Observe que los ejes *x* e *y* no cambian al seleccionar esta opción.

**Lista.** Organiza los paneles como una lista, de manera que cada casilla representa una combinación de valores. Las columnas y filas ya no estarán asignadas a valores individuales. Esta opción permite que los paneles se ajusten si es necesario.

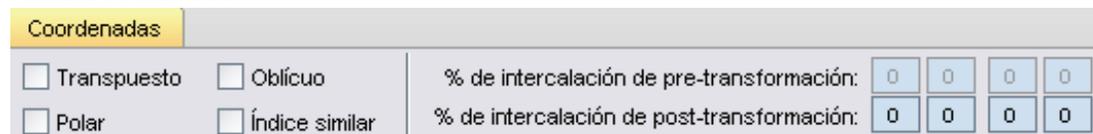
## Transformación del sistema de coordenadas

Muchas visualizaciones se muestran en un sistema de coordenadas plano y rectangular. Puede transformar el sistema de coordenadas según sea necesario. Por ejemplo puede aplicar una transformación polar al sistema de coordenadas, añadir efectos de sombra en sentido oblicuo y transponer los ejes. También puede deshacer cualquiera de estas transformaciones si ya se aplicaron a la visualización actual. Por ejemplo, se dibuja un gráfico de sectores en un sistema de coordenadas polares. Si lo desea, puede deshacer la transformación polar y mostrar el gráfico de sectores como una única barra apilada en un sistema de coordenadas rectangular.

### Para transformar el sistema de coordenadas

- ▶ Seleccione el sistema de coordenadas que desee transformar. El sistema de coordenadas se selecciona eligiendo el marco que rodea el gráfico individual.
- ▶ Pulse en la pestaña Coordenadas de la paleta Propiedades.

Figura 5-112  
Pestaña Coordenadas



- ▶ Seleccione las transformaciones que desee aplicar al sistema de coordenadas. También puede cancelar la selección de una transformación para deshacerla.

**Transpuesto.** El cambio de orientación de los ejes se denomina **transposición**. Es como intercambiar los ejes vertical y horizontal en una visualización 2-D.

**Polar.** Una transformación polar dibuja los elementos gráficos con un ángulo y una distancia específicos desde el centro del gráfico. Un gráfico de sectores es una visualización de una dimensión con una transformación polar que dibuja las barras individuales con ángulos específicos. Un gráfico radial es una visualización 2-D con una transformación polar que dibuja elementos gráficos con ángulos y distancias específicos desde el centro del gráfico. Una visualización 3-D también incluye una dimensión de profundidad adicional.

**Oblicuo.** Una transformación oblicua añade un efecto 3-D a los elementos gráficos. Esta transformación añade profundidad a los elementos gráficos, pero su profundidad es meramente decorativa. No se ve influenciado por valores de datos concretos.

**Misma proporción.** Si se aplica la misma proporción se especifica que la misma distancia en cada escala represente la misma diferencia de valores de datos. Por ejemplo, 2 cm en ambas escalas representan una diferencia de 1000.

**% del recuadro antes de transformación.** Si tras la transformación se recortan los ejes, puede que desee añadir recuadros al gráfico antes de aplicar la transformación. Los recuadros reducen las dimensiones en un porcentaje antes de aplicar transformaciones al sistema de coordenadas. Tiene control sobre las dimensiones  $x$  inferior,  $x$  superior,  $y$  inferior e  $y$  superior, en ese orden.

**% del recuadro tras la transformación.** Si desea cambiar la relación de aspecto de un gráfico, puede añadirle recuadros tras aplicar transformación. Los recuadros reducen las dimensiones en un porcentaje después de aplicar transformaciones al sistema de coordenadas. Estos recuadros también pueden aplicarse aunque no se haya realizado ninguna transformación en el gráfico. Tiene control sobre las dimensiones  $x$  inferior,  $x$  superior,  $y$  inferior e  $y$  superior, en ese orden.

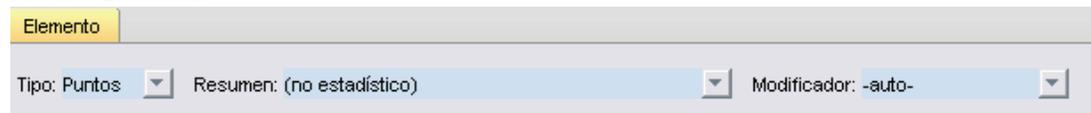
## Cambio de estadísticos y elementos gráficos

Puede convertir un en otro tipo, cambiar el estadístico utilizado para dibujar el elemento gráfico o especificar el modificador de colisión que determina lo que ocurre cuando los elementos gráficos se solapan.

### Para convertir un elemento gráfico

- ▶ Seleccione el elemento gráfico que desee convertir.
- ▶ Pulse en la pestaña Elemento de la paleta Propiedades.

Figura 5-113  
Pestaña Elemento.



- ▶ Seleccione un nuevo tipo de elemento gráfico en la lista Tipo.

Tipo de elemento gráfico	Descripción
Punto	Un marcador que identifica un punto de datos específico. Un elemento de punto se utiliza en diagramas de dispersión y otras visualizaciones relacionadas.
Intervalo	Una forma rectangular dibujada con un valor de datos específico y que rellena el espacio entre un origen y otro valor de datos. Se utiliza un elemento de intervalo en gráficos de barras e histogramas.
Línea	Una línea que conecta los valores de datos.
Ruta	Una línea que conecta los valores de datos en el orden en que aparecen en el conjunto de datos.
Área	Una línea que conecta los elementos de datos con el área entre la línea y un origen relleno.
Polígono	Una forma con varias caras que encierra una región de datos. Se puede utilizar un elemento poligonal en un diagrama de dispersión en intervalos o en un mapa.
Esquema	Un elemento que consta de un cuadro con bigotes y marcadores indicando valores atípicos. Se utilizan elementos de esquema para gráficos de caja.

### Para cambiar el estadístico

- ▶ Seleccione el elemento gráfico cuyo estadístico desee cambiar.
- ▶ Pulse en la pestaña Elemento de la paleta Propiedades.

- En la lista desplegable Resumen, seleccione un nuevo estadístico. Tenga en cuenta que al seleccionar un estadístico se agregan los datos. Si en su lugar desea que la visualización muestre datos sin agregar, seleccione (sin estadístico) en la lista Resumen.

### ***Estadísticos de resumen calculados a partir de un campo continuo***

- **Media.** Una medida de tendencia central. El promedio aritmético, la suma dividida por el número de casos.
- **Mediana.** Es el valor por encima y por debajo del cual se encuentran la mitad de los casos, el percentil 50. Si hay un número par de casos, la mediana es la media de los dos valores centrales, cuando los casos se ordenan en orden ascendente o descendente. La mediana es una medida de tendencia central que no es sensible a los valores atípicos (a diferencia de la media, que puede resultar afectada por unos pocos valores extremadamente altos o bajos).
- **Moda.** El valor que ocurre con mayor frecuencia. Si varios valores comparten la mayor frecuencia de aparición, cada uno de ellos es un modo.
- **Mínimo.** Valor más pequeño de una variable numérica.
- **Máximo.** El mayor valor de una variable numérica.
- **Rango.** La diferencia entre los valores mínimos y máximos.
- **Medio intervalo.** La mitad del rango, es decir, el valor cuya diferencia con el mínimo es igual a la diferencia con el máximo.
- **Sum.** Suma o total de todos los valores, a lo largo de todos los casos que no tengan valores perdidos.
- **Suma acumulada.** La suma acumulada de los valores. Cada elemento del gráfico muestra la suma de un subgrupo más la suma total de los grupos anteriores.
- **Porcentaje de la suma.** El porcentaje en cada subgrupo basado en un campo sumado en comparación con la suma en todos los grupos.
- **Suma del porcentaje acumulado.** El porcentaje acumulativo en cada subgrupo basado en un campo sumado en comparación con la suma en todos los grupos. Cada elemento del gráfico muestra el porcentaje de un subgrupo más el porcentaje total de los grupos anteriores.
- **Variance.** Es una medida de dispersión en torno a la media, igual a la suma de las desviaciones al cuadrado respecto a la media, dividida por el número de casos menos 1. La varianza se mide en unidades que son el cuadrado de las de la variable en cuestión.
- **Desviación típica.** Es una medida de la dispersión en torno a la media. En una distribución normal, el 68% de los casos se encuentra dentro de una desviación típica de la media y el 95% queda entre dos desviaciones típicas. Por ejemplo, si la edad media es de 45 años, con una desviación típica de 10, el 95% de los casos estaría entre los 25 y 65 en una distribución normal.
- **Error típico.** Medida de cuánto puede variar el valor de un estadístico de contraste de muestra en muestra. Es la desviación típica de la distribución muestral de un estadístico. Por ejemplo, el error típico de la media es la desviación típica de las medias muestrales.
- **Kurtosis.** Medida del grado en que las observaciones están agrupadas en torno al punto central. Para una distribución normal, el valor del estadístico de kurtosis es 0. Una kurtosis positiva indica que, con respecto a una distribución normal, las observaciones se concentran más en el centro de la distribución y presentan colas más estrechas hasta los valores extremos de la

distribución, en cuyo punto las colas de la distribución leptocúrtica son más gruesas con respecto a una distribución normal. Una curtosis negativa indica que, con respecto a una distribución normal, las observaciones se concentran menos y presentan colas más gruesas hasta los valores extremos de la distribución, en cuyo punto las colas de la distribución platicúrtica son más estrechas con respecto a una distribución normal.

- **Asimetría.** Medida de la asimetría de una distribución. La distribución normal es simétrica y tiene un valor de asimetría igual a 0. Una distribución que tenga una asimetría positiva significativa tiene una cola derecha larga. Una distribución que tenga una asimetría negativa significativa tiene una cola izquierda larga. Como regla aproximada, un valor de la asimetría mayor que el doble de su error típico se asume que indica una desviación de la simetría.

Los siguientes estadísticos de región pueden resultar en más de un elemento gráfico por subgrupo. Si utiliza los elementos de intervalo, área, o gráficos de borde, un estadístico de región da como resultado un elemento gráfico que muestra el rango. El resto de elementos gráficos dan como resultado dos elementos diferentes, uno que muestra el inicio del rango y el otro que muestra el final del rango.

- **Región: Rango.** El rango entre los valores mínimo y máximo.
- **Región: 95% de intervalo de confianza para la media.** Un rango de valores con un 95% de opciones de incluir la media poblacional.
- **Región: 95% de intervalo de confianza individual.** Un rango de valores con un 95% de opciones de incluir el valor predictivo especificando el caso individual.
- **Región: 1 Desviación típica por encima/debajo de la media.** Un rango de valores entre 1 desviación típica por encima y por debajo de la media.
- **Región: 1 Error típico por encima/debajo de la media.** Un rango de valores entre 1 error típico por encima y por debajo de la media.

#### ***Estadísticos de resumen basados en recuentos***

- **Recuento.** Número de filas/casos.
- **Recuento acumulado.** El número acumulado de filas/casos. Cada elemento del gráfico muestra el recuento de un subgrupo más el recuento total de los grupos anteriores.
- **Porcentaje de recuento.** El porcentaje de filas/casos de cada subgrupo en comparación con el número total de filas/casos.
- **Porcentaje acumulado de recuento.** El porcentaje acumulado de filas/casos de cada subgrupo en comparación con el número total de filas/casos. Cada elemento del gráfico muestra el porcentaje de un subgrupo más el porcentaje total de los grupos anteriores.

#### ***Para especificar el modificador de colisión***

El modificador de colisión determina lo que ocurre cuando los elementos gráficos se superponen.

- ▶ Seleccione el elemento gráfico cuyo modificador de colisión desee especificar.
- ▶ Pulse en la pestaña Elemento de la paleta Propiedades.

- ▶ En la lista desplegable Modificador, seleccione un modificador de colisión. -auto- permite a la aplicación determinar qué modificador de colisión es adecuado para el tipo de elemento gráfico y estadístico.

**Superpuesto.** Dibuje elementos gráficos encima de otros si tienen el mismo valor.

**Pila.** Elementos gráficos de pila que normalmente se superpondrían al tener los mismos valores de datos.

**Esquivar.** Mueve los elementos gráficos junto a otros elementos gráficos que aparecen con el mismo valor, en vez de superponerlos. Los elementos gráficos se ordenan simétricamente. Es decir, los elementos gráficos se trasladan a lados opuestos de una posición central. Esquivar es similar a conglomerar.

**Pila.** Mueve los elementos gráficos junto a otros elementos gráficos que aparecen con el mismo valor, en vez de superponerlos. Los elementos gráficos se ordenan asimétricamente. Es decir, los elementos gráficos se apilan unos sobre otros, con el elemento gráfico más bajo colocado en un valor específico de la escala.

**Perturbar (normal).** Recoloca los elementos gráficos en el mismo valor de datos aleatoriamente utilizando la distribución normal.

**Perturbar (uniforme).** Recoloca los elementos gráficos en el mismo valor de datos aleatoriamente utilizando la distribución uniforme.

## ***Cambio de la posición de la leyenda***

Si un gráfico incluye una leyenda, ésta suele mostrarse a la derecha de un gráfico. Si lo desea, puede cambiar esta posición.

### ***Cómo cambiar la posición de la leyenda***

- ▶ Seleccione la leyenda.
- ▶ Pulse en la pestaña Leyenda de la paleta Propiedades.

Figura 5-114  
Pestaña Leyenda



- ▶ Seleccione una posición.

## ***Copia de una visualización y de datos de visualización***

La paleta General incluye botones para copiar la visualización y sus datos.

Figura 5-116  
Botón Copiar visualización



**Copia de visualización.** Esta acción copia la visualización en el portapapeles como una imagen. Hay múltiples formatos de imagen disponibles. Cuando pegue la imagen en otra aplicación, puede seleccionar una opción de pegado especial para seleccionar uno de los formatos de imagen disponibles para pegar.

Figura 5-117  
Botón Copiar datos de visualización



**Copia de los datos de visualización.** Esta acción copia los datos subyacentes utilizados para dibujar la visualización. Los datos se copian en el portapapeles como texto normal o como texto con formato HTML. Cuando pegue los datos en otra aplicación, puede seleccionar una opción de pegado especial para seleccionar uno de los formatos disponibles para pegar.

## Métodos abreviados de teclado

Tabla 5-4  
Métodos abreviados de teclado

Tecla de método abreviado	Función
Ctrl + Espacio	Cambiar entre modo de exploración y el modo de edición
Suprimir	Eliminar un elemento de visualización
Ctrl + Z	Deshacer
Ctrl + Y	Rehacer
F2	Mostrar esquema para seleccionar los elementos del gráfico

## Adición de títulos y notas al pie

En todos los tipos de gráfico, es posible añadir títulos, notas al pie o etiquetas de eje únicos que ayuden a identificar lo que aparece en el gráfico.

Figura 5-118  
Adición de una nota al pie del gráfico



### **Adición de títulos a gráficos**

- ▶ Desde los menús, seleccione Edición > Añadir título de gráfico. Se mostrará un cuadro de texto que contiene <TITLE> sobre el gráfico.
- ▶ Compruebe que está en el modo de edición. Desde los menús, seleccione Ver > Modo edición.
- ▶ Pulse dos veces en el texto <TITLE>.
- ▶ Escriba el título que desee y pulse Volver.

### **Adición de notas al pie del gráfico**

- ▶ Desde los menús, seleccione Edición > Añadir nota al pie del gráfico. Se mostrará un cuadro de texto que contiene <FOOTNOTE> debajo del gráfico.
- ▶ Compruebe que está en el modo de edición. Desde los menús, seleccione Ver > Modo edición.
- ▶ Pulse dos veces en el texto <FOOTNOTE>.

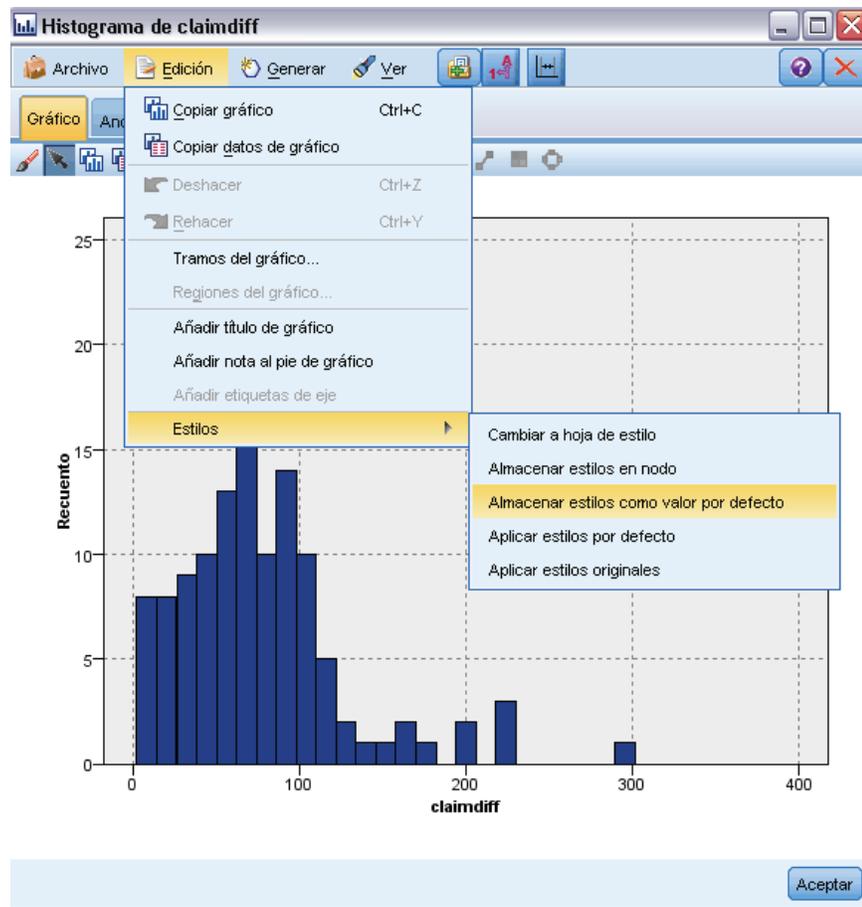
- Escriba el título que desee y pulse Volver.

## Uso de hojas de estilo de gráficos

La información de visualización de gráfico básica, como colores, fuentes, símbolos y grosor de líneas, se controla en una hoja de estilo. Se suministra una hoja de estilo por defecto con IBM® SPSS® Modeler, sin embargo, puede modificarla si lo necesita. Por ejemplo, es posible que desee utilizar en los gráficos el esquema de color corporativo que suele utilizar en las presentaciones. Si desea obtener más información, consulte el tema [Edición de visualizaciones](#) el p. 367.

En los nodos de gráfico, puede utilizar el modo de edición para realizar cambios de estilo al aspecto de un gráfico. Puede utilizar el menú Edición > Estilos para guardar los cambios como una hoja de estilo para aplicarlos a todos los gráficos que genere posteriormente desde el nodo de gráfico actual, o utilizarlo como nueva hoja de estilo por defecto para todos los gráficos que genere con SPSS Modeler.

Figura 5-119  
Selección de estilos de gráfico



Existen cinco opciones de hoja de estilo disponibles en la opción Estilos del menú Edición.

- **Cambiar a hoja de estilo.** Muestra una lista de diferentes estilos de hojas almacenados que puede seleccionar para cambiar la apariencia de los gráficos. Si desea obtener más información, consulte el tema [Aplicación de hojas de estilo](#) el p. 387.
- **Almacenar estilos en nodo.** Almacena las modificaciones realizadas a los estilos del gráfico seleccionado para poder aplicarlas a los futuros gráficos que se generen desde el mismo nodo de gráfico de la ruta actual.
- **Almacenar estilos como valor por defecto.** Almacena las modificaciones realizadas a los estilos del gráfico seleccionado para poder aplicarlas a los futuros gráficos que se generen desde cualquier nodo de gráfico de cualquier ruta. Al seleccionar esta opción, puede utilizar Aplicar estilos por defecto para que todos los gráficos existentes utilicen los mismos estilos.
- **Aplicar estilos por defecto.** Esta opción cambia los estilos del gráfico seleccionado por aquellos que se han guardado como estilos por defecto.
- **Aplicar estilos originales.** Esta opción vuelve a cambiar los estilos del gráfico seleccionado por aquellos que se suministran por defecto.

## ***Aplicación de hojas de estilo***

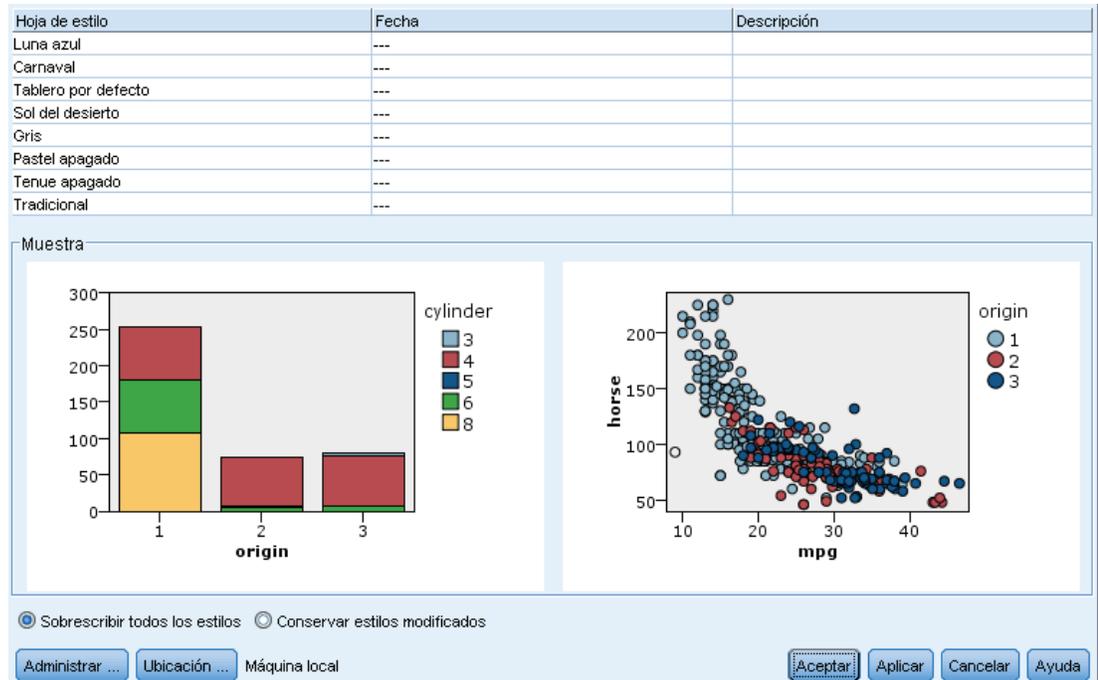
Puede aplicar una hoja de estilo de visualización que especifica propiedades de estilo de la visualización. Por ejemplo, la hoja de estilo puede definir fuentes, estilos de guión y colores, entre otras opciones. Hasta cierto punto, las hojas de estilo proporcionan un acceso directo a las modificaciones que tenga que realizar manualmente. Tenga en cuenta, sin embargo, que una hoja de estilo se limita a cambios de *estilo*. Otros cambios, como la posición de la leyenda o el rango de la escala no se guardan en la hoja de estilo.

### ***Cómo se aplica una hoja de estilo***

- ▶ Seleccione en los menús:  
Edición > Estilos > Cambiar a hoja de estilo
- ▶ Utilice el cuadro de diálogo Cambiar a hoja de estilo para seleccionar una hoja de estilo.
- ▶ Pulse Aplicar para aplicar la hoja de estilo a la visualización sin cerrar el cuadro de diálogo. Pulse Aceptar para la hoja de estilo y cerrar el cuadro de diálogo.

**Cuadro de diálogo Cambiar a/Seleccionar hoja de estilo**

Figura 5-120  
Cuadro de diálogo Cambiar a hoja de estilo



La tabla de la parte superior del cuadro de diálogo muestra todas las hojas de estilo de visualización que están actualmente disponibles. Algunas hojas de estilo están preinstaladas, mientras que otras se pueden haber creado en IBM® SPSS® Visualization Designer (como producto independiente).

La parte inferior del cuadro de diálogo muestra visualizaciones de ejemplo con datos de muestra. Seleccione una de las hojas de estilo que aplicarán sus estilos a las visualizaciones de ejemplo. Estos ejemplos le ayudan a determinar cómo afectarán las hojas de estilo a su visualización real.

El cuadro de diálogo ofrece las siguientes opciones.

**Estilos existentes.** Por defecto, una hoja de estilo puede sobrescribir todos los estilos de la visualización. Es posible modificar este comportamiento.

- **Sobrescribir todos los estilos.** Cuando se aplica la hoja de estilo, se sobrescriben todos los estilos de la visualización, incluyendo aquellos estilos que se modifican durante la sesión de modificación actual.
- **Conservar estilos modificados.** Cuando se aplica la hoja de estilo, se sobrescriben sólo aquellos estilos que *no* se han modificado en la visualización durante la sesión de modificación actual. Los estilos que se han modificado durante la sesión de modificación actual se conservan.

**Administrar.** Administra plantillas, hojas de estilo y mapas de visualización en su ordenador. Puede importar, exportar, cambiar el nombre y eliminar plantillas, hojas de estilo y mapas de visualización en su máquina local. Si desea obtener más información, consulte el tema [Administración de plantillas, hojas de estilo y archivos de mapas](#) el p. 289.

**Ubicación.** Cambia la ubicación donde están almacenadas las plantillas, hojas de estilo y mapas de visualización. La ubicación actual se enumera a la derecha del botón. Si desea obtener más información, consulte el tema [Configuración de la ubicación de las plantillas, hojas de estilo y mapas](#) el p. 287.

## Impresión, almacenamiento, copia y exportación de gráficos

Cada gráfico tiene un número de opciones que permiten guardar o imprimir el gráfico o exportarlo a otro formato. La mayoría de estas opciones están disponibles en el menú Archivo. Adicionalmente, en el menú Edición, puede seleccionar copiar el gráfico o los datos que contienen para utilizarlos en otras aplicaciones.

Figura 5-121

Menú Archivo y barra de herramientas de las ventanas Gráfico



### Impresión

- Para imprimir el gráfico, utilice el elemento de menú o el botón Imprimir. Antes de imprimir, puede utilizar Configurar página y Presentación preliminar para configurar las opciones de impresión y previsualizar el resultado.

**Almacenamiento de gráficos**

- ▶ Para guardar el gráfico en un archivo de resultados de IBM® SPSS® Modeler (\*.cou), seleccione Archivo > Guardar o Archivo > Guardar como en los menús.

o

Para guardar el gráfico en el repositorio, elija Archivo > Almacenar resultado en los menús.

**Copia de gráficos**

- ▶ Para copiar el gráfico para utilizarlo en otra aplicación, como MS Word o MS PowerPoint, elija Edición > Copiar gráfico en los menús.

**Copia de datos**

- ▶ Para copiar los datos para utilizarlos en otra aplicación, como MS Excel o MS Word, elija Edición > Copiar datos en los menús. Por defecto, los datos aparecen con formato HTML. Utilice Pegado especial en la otra aplicación para ver otras opciones de formato al pegar.

**Exportación de gráficos**

La opción Exportar gráfico permite exportar el gráfico en uno de los formatos siguientes: Mapa de bits (.bmp), JPEG (.jpg), PNG (.png), HTML (.html) o documento ViZml (.xml) para utilizarlo en otras aplicaciones de IBM® SPSS® Statistics

- ▶ Para exportar gráficos, seleccione Archivo > Exportar gráfico en los menús y seleccione el formato.

**Exportación de tablas**

La opción Exportar tabla permite exportar la tabla en uno de los formatos siguientes: delimitado por tabulaciones (.tab), delimitado por comas (.csv) o HTML (.html)

- ▶ Para exportar tablas, seleccione Archivo > Exportar tabla en los menús y seleccione el formato.

# Nodos de resultado

## Conceptos básicos sobre nodos de resultados

Los nodos de resultados ofrecen los medios para obtener información acerca de los datos y los modelos. También proporcionan un mecanismo para exportar datos en varios formatos y poder interactuar con otras herramientas de software.

Los siguientes nodos de resultado están disponibles:



El nodo Tabla muestra los datos en formato de tabla, que también se puede escribir en un archivo. Esto es útil en cualquier momento en que necesite inspeccionar sus valores de datos o exportarlos en un formato fácilmente legible. Si desea obtener más información, consulte el tema [Nodo Tabla](#) el p. 398.



El nodo Matriz crea una tabla que muestra las relaciones entre campos. Se suele utilizar normalmente para mostrar las relaciones entre dos campos simbólicos, pero también puede mostrar relaciones entre campos de marcas o entre campos numéricos. Si desea obtener más información, consulte el tema [Nodo Matriz](#) el p. 403.



El nodo Análisis evalúa la capacidad de los modelos predictivos para generar pronósticos precisos. Los nodos Análisis realizan varias comparaciones entre los valores pronosticados y los valores reales para uno o más nugget de modelo. También pueden comparar modelos predictivos entre sí. Si desea obtener más información, consulte el tema [Nodo Análisis](#) el p. 407.



El nodo Auditar datos permite echar un primer vistazo exhaustivo a los datos, incluyendo estadísticos de resumen, histogramas y distribución para cada campo, así como información sobre valores atípicos, valores perdidos y extremos. Los resultados se muestran en una matriz fácil de leer que se puede ordenar y utilizar para generar nodos de preparación de datos y gráficos de tamaño completo. Si desea obtener más información, consulte el tema [Nodo Auditar datos](#) el p. 412.



El nodo Transformación permite seleccionar y previsualizar los resultados de las transformaciones antes de aplicarlas a los campos seleccionados. Si desea obtener más información, consulte el tema [Nodo Transformación](#) el p. 427.



El nodo Estadísticos ofrece información básica de resumen acerca de los campos numéricos. Calcula estadísticos de resumen para campos individuales y correlaciones entre campos. Si desea obtener más información, consulte el tema [Nodo Estadísticos](#) el p. 432.



El nodo Medias compara las medias de grupos independientes o de pares de campos relacionados para probar si existen diferencias significativas. Por ejemplo, puede comparar los ingresos medios antes y después de poner en marcha una promoción o comparar los ingresos de los clientes que no recibieron esa promoción con los que sí lo hicieron. Si desea obtener más información, consulte el tema [Nodo Medias](#) el p. 437.



El nodo Informe crea informes con formato que contienen texto fijo, así como datos y otras expresiones derivadas de los datos. Puede especificar el formato del informe utilizando plantillas de texto para definir el texto fijo y las construcciones de resultados de datos. Puede proporcionar formato de texto personalizado utilizando etiquetas HTML de la plantilla y configurando opciones en la ficha Resultado. Puede incluir valores de datos y otros resultados condicionales mediante el uso de expresiones CLEM en la plantilla. Si desea obtener más información, consulte el tema [Nodo Informe](#) el p. 442.



El nodo Val. globales explora los datos y calcula los valores de resumen que se pueden utilizar en expresiones CLEM. Por ejemplo, puede utilizar este nodo para calcular estadísticos para un campo denominado *edad* y, a continuación, utilizar la media global de *edad* en expresiones CLEM insertando la función `@GLOBAL_MEAN(edad)`. Si desea obtener más información, consulte el tema [Nodo Val. globales](#) el p. 446.

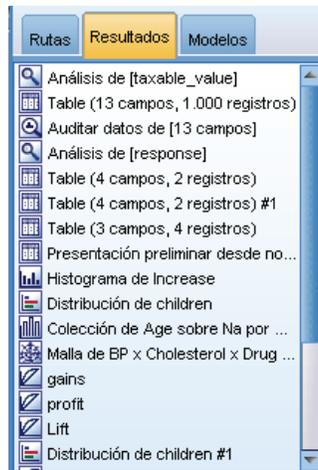
## Administración de resultados

El administrador de resultados muestra los diagramas, gráficos y tablas generados durante una sesión de IBM® SPSS® Modeler. Siempre puede volver a abrir un resultado pulsando dos veces en él en el administrador, sin tener que volver a ejecutar la ruta o el nodo correspondiente.

### Para ver el administrador de resultados

- Abra el menú Ver y elija Administradores. Pulse en la pestaña Resultados.

Figura 6-1  
Administrador de resultados



Desde el administrador de resultados, puede:

- Visualizar objetos de resultados existentes, como histogramas, diagramas de evaluación y tablas.
- Cambiar el nombre de los objetos de resultados.
- Guardar los objetos de resultados en disco o en IBM® SPSS® Collaboration and Deployment Services Repository (si está disponible).
- Añadir archivos de resultados al proyecto actual.

- Eliminar objetos de resultados no almacenados desde la sesión actual.
- Abrir los objetos de resultados guardados o recuperarlos de IBM SPSS Collaboration and Deployment Services Repository (si está disponible).

Para acceder a estas opciones, pulse con el botón derecho del ratón en la pestaña Resultados.

## Visualización de resultados

El resultado en pantalla se muestra en una ventana del explorador de resultados. La ventana del explorador de resultados tiene su propio conjunto de menús que permiten imprimir y guardar el resultado, así como exportarlo a otro formato. Tenga en cuenta que las opciones específicas pueden variar dependiendo del tipo de resultado.

**Impresión, almacenamiento y exportación de datos.** Hay más información disponible a continuación:

- Para imprimir el resultado, utilice la opción del menú o el botón Imprimir. Antes de imprimir, puede utilizar Configurar página y Presentación preliminar para configurar las opciones de impresión y previsualizar el resultado.
- Para guardar el resultado en un archivo de salida de IBM® SPSS® Modeler (.cou), seleccione Guardar o Guardar como en el menú Archivo.
- Para guardar los resultados en otro formato, como texto o HTML, elija Exportar en el menú Archivo. Si desea obtener más información, consulte el tema [Exportación de resultados](#) el p. 396.
- Para guardar el resultado en un repositorio compartido para que otros usuarios puedan visualizarlo utilizando e IBM® SPSS® Collaboration and Deployment Services Deployment Portal, seleccione Publicar en Web en el menú Archivo. Tenga en cuenta que esta opción requiere una licencia distinta de IBM® SPSS® Collaboration and Deployment Services.

**Selección de casillas y columnas.** El menú Edición contiene varias opciones para seleccionar, anular la selección y copiar casillas y columnas, según considere adecuado para el tipo de resultado en cuestión. Si desea obtener más información, consulte el tema [Selección de casillas y columnas](#) el p. 397.

**Generación de nuevos nodos.** El menú Generar permite generar nuevos nodos basados en el contenido del explorador de resultados. Las opciones varían dependiendo del tipo de resultado y los elementos del resultado actualmente seleccionados. Para obtener detalles acerca de las opciones de generación de nodos para un tipo particular de resultado, consulte la documentación de ese resultado.

## Publicar en Web

La función Publicar en Web le permite publicar ciertos tipos de resultados de rutas en un IBM® SPSS® Collaboration and Deployment Services Repository central compartido que conforma la base de IBM® SPSS® Collaboration and Deployment Services. Si utiliza esta opción, otros usuarios que deseen ver sus resultados podrán hacerlo utilizando un acceso a Internet y una cuenta de IBM SPSS Collaboration and Deployment Services, no necesitan tener instalado IBM® SPSS® Modeler.

*Nota:* Es necesario disponer de una licencia independiente para acceder a un repositorio IBM SPSS Collaboration and Deployment Services. Si desea obtener más información, consulte <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

La siguiente tabla enumera los nodos de SPSS Modeler que son compatibles con la función Publicar en Web. Los resultados de estos nodos se almacenan en IBM SPSS Collaboration and Deployment Services Repository en formato de objeto de resultado (.cou), y pueden verse directamente en IBM® SPSS® Collaboration and Deployment Services Deployment Portal.

El resto de los tipos de resultados pueden verse únicamente si la aplicación apropiada (por ej. SPSS Modeler para objetos de ruta) está instalada en la máquina del usuario.

Tabla 6-1  
*Nodos que son compatibles con Publicar en Web*

Tipo de nodo	Nodo
Gráficos	todos
Resultado	Tabla
	Matriz
	Auditar datos
	Transformar
	Medias
	Análisis
	Estadísticos
	Informe (HTML)
IBM® SPSS® Statistics	Resultado Statistics

### **Publicación de resultados en la Web**

Para publicar resultados en la Web:

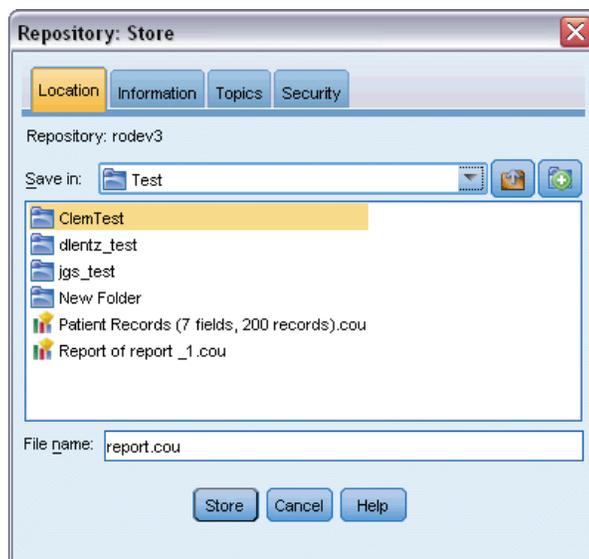
- ▶ En una ruta de IBM® SPSS® Modeler, ejecute uno de los nodos enumerados en la tabla. Al hacerlo se creará un objeto de resultado (por ejemplo, una tabla, una matriz o un objeto de informe) en una nueva ventana.
- ▶ Desde la ventana de objeto de resultado, seleccione:  
File > Publicar en Web

*Nota:* Si desea exportar archivos HTML simples para su uso en un navegador Web estándar, seleccione Exportar en el menú Archivo y seleccione HTML.

- ▶ Conecte con IBM® SPSS® Collaboration and Deployment Services Repository.

Cuando se ha conectado con éxito, se muestra el cuadro de diálogo Repositorio: Almacenar se muestra, ofreciendo una serie de opciones de almacenamiento.

Figura 6-2  
Cuadro de diálogo Repositorio: Almacenar



- ▶ Cuando haya seleccionado las opciones de almacenamiento que desea, pulse en Almacenar.

### **Visualización de resultados publicados en la Web**

Debe disponer de una cuenta de IBM SPSS Collaboration and Deployment Services para utilizar esta función. Si tiene la aplicación apropiada instalada para el tipo de objeto que desea ver (por ejemplo, IBM® SPSS® Modeler o IBM® SPSS® Statistics), el resultado se muestra dentro de la aplicación en vez de en el navegador.

*Nota:* Es necesario disponer de una licencia independiente para acceder a IBM® SPSS® Collaboration and Deployment Services. Si desea obtener más información, consulte <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>.

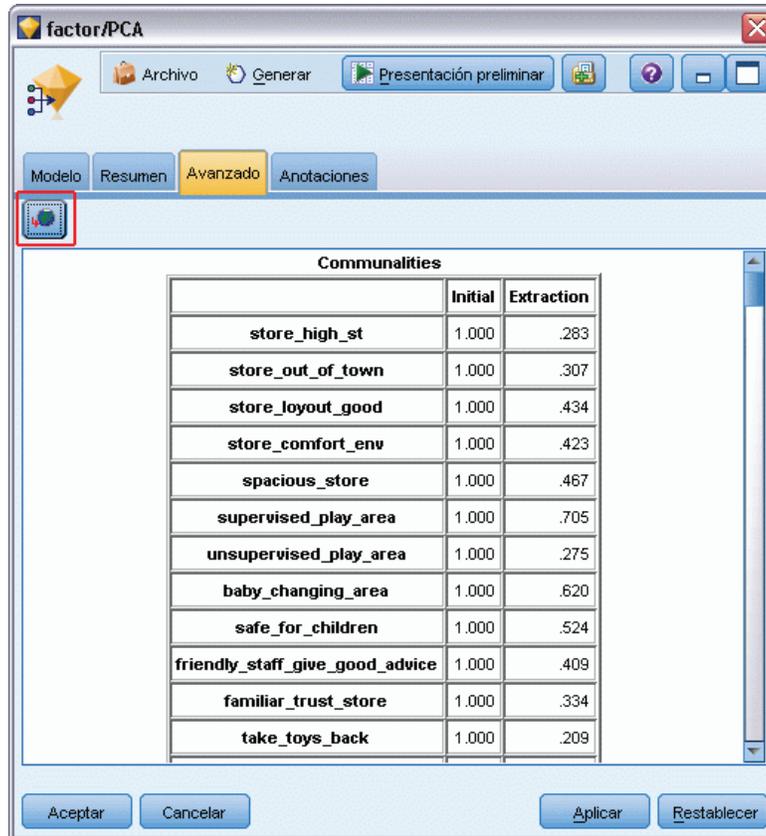
Para ver resultados publicados en la Web:

- ▶ Introduzca la siguiente dirección en su navegador: `http://<host_repos>:<puerto_repos>/peb` donde *host\_repos* y *puerto\_repos* representan el nombre de host y el número de puerto del host de IBM SPSS Collaboration and Deployment Services.
- ▶ Introduzca los detalles de inicio de sesión de su cuenta de IBM SPSS Collaboration and Deployment Services.
- ▶ Pulse en Repositorio de contenidos.
- ▶ Busque el objeto que desee visualizar.
- ▶ pulse en el nombre de objeto. En algunos tipos de objetos, como gráficos, puede haber una demora mientras se representa el objeto en el navegador.

## Visualización de resultados en un explorador HTML

En la pestaña Avanzado de los nuggets de modelo Regresión lineal, Regresión logística y PCA/Factorial, puede ver la información en un explorador diferente, como Internet Explorer. La información aparece en formato HTML, lo que le permite guardarla y reutilizarla en otros sitios, como una intranet de empresa, o un sitio de Internet.

Figura 6-3  
Botón de inicio en la pestaña Avanzado del nugget del modelo



Para mostrar la información en un explorador, pulse en el botón situado debajo del icono de modelo, en la esquina superior izquierda de la pestaña Avanzado del nugget del modelo.

## Exportación de resultados

En la ventana del explorador de resultados, puede elegir exportar el resultado a otro formato, como texto o HTML. Los formatos de exportación varían dependiendo del tipo de resultado, pero en general son parecidos a las opciones de tipo de archivo disponibles si selecciona Guardar en un archivo en el nodo utilizado para generar el resultado.

### Para exportar resultados

- ▶ En el explorador de resultados, abra el menú Archivo y elija Exportar. A continuación, elija el tipo de archivo que desea crear:
  - **Delimitado por tabulaciones (\*.tab).** Esta opción genera un archivo de texto con formato que contiene los valores de datos. Este estilo suele ser útil para generar una representación de texto sin formato de la información que se puede importar a otras aplicaciones. Esta opción está disponible para los nodos Tabla, Matriz y Medias.
  - **Delimitado por comas (\*.dat).** Esta opción genera un archivo de texto delimitado por comas que contiene los valores de datos. Este estilo suele ser útil como una manera rápida de generar un archivo de datos que se puede importar a hojas de cálculo u otras aplicaciones de análisis de datos. Esta opción está disponible para los nodos Tabla, Matriz y Medias.
  - **Delimitado por tabuladores transpuesto (\*.tab).** Esta opción es idéntica a la opción Delimitado por tabuladores, pero los datos se transponen de manera que las filas representan los campos y las columnas representan los registros.
  - **Delimitado por comas transpuesto (\*.dat).** Esta opción es idéntica a la opción Delimitado por comas, pero los datos se transponen de manera que las filas representan los campos y las columnas representan los registros.
  - **HTML (\*.html)** Esta opción escribe resultado con formato HTML en un archivo o archivos.

### Selección de casillas y columnas.

Figura 6-4  
Ventana del explorador de tablas

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	claimt
1	id602	name602	north	1780	42	9	734118.000	maize	arable
2	id606	name606	southeast	1580	42	7	445785.000	maize	arable
3	id607	name607	southeast	1820	29	6	211605.000	maize	arable
4	id608	name608	southeast	1640	108	7	1167040.0...	maize	arable
5	id610	name610	southeast	600	80	6	267928.000	wheat	arable
6	id611	name611	southeast	980	38	6	222703.000	maize	arable
7	id613	name613	southeast	440	86	3	115544.000	potatoes	arabi
8	id614	name614	southeast	1260	90	8	900243.000	maize	arable
9	id616	name616	midlands	1660	36	9	490617.000	rapeseed	arable
10	id620	name620	north	880	74	6	426988.000	rapeseed	arable
11	id621	name621	southwest	1160	105	4	299274.000	maize	arable
12	id622	name622	southeast	1500	61	7	687736.000	wheat	arable
13	id623	name623	southeast	1260	17	8	170279.000	maize	arable
14	id626	name626	midlands	1580	109	8	1286430.0...	wheat	arable
15	id627	name627	southeast	500	93	3	102720.000	rapeseed	arable
16	id628	name628	southeast	880	15	5	70439.800	wheat	arable
17	id630	name630	midlands	680	81	4	221391.000	potatoes	arable
18	id636	name636	southeast	1160	21	8	185939.000	potatoes	arable
19	id637	name637	midlands	940	106	6	622450.000	maize	arable
20	id638	name638	midlands	1480	64	6	586185.000	wheat	arable

Una serie de nodos, incluidos los nodos Tabla, Matriz y Medias, generan un resultado tabular. Estas tablas de resultados se pueden ver y manipular de formas similares, incluyendo seleccionar casillas, copiar toda o parte de la tabla al Portapapeles, generar nuevos nodos a partir de la selección actual y guardar e imprimir la tabla.

**Selección de casillas.** Para seleccionar una casilla, pulse en ella. Para seleccionar un intervalo rectangular de casillas, pulse en una esquina del intervalo que desee, arrastre el ratón a la otra esquina del intervalo y suelte el botón del ratón. Para seleccionar toda una columna, pulse en su encabezado. Para seleccionar varias columnas, pulse la tecla Mayús o Ctrl mientras pulsa en los encabezados de las columnas.

Cuando realiza una nueva selección, la vieja selección se borra. Si mantiene pulsada la tecla Ctrl mientras realiza la selección, podrá añadir la nueva selección a cualquier selección existente en lugar de borrar la antigua. Puede utilizar este método para seleccionar varias regiones no contiguas de la tabla. El menú Edición también contiene las opciones Seleccionar todo y Borrar selección.

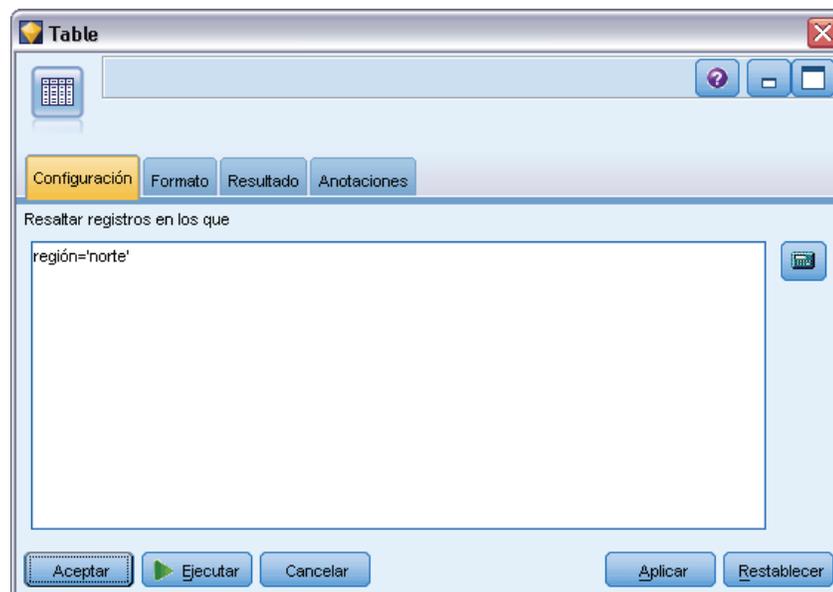
**Reordenación de columnas.** Los exploradores de resultados de los nodos Medias y Tabla permiten mover las columnas en la tabla pulsando en el encabezado de columna y arrastrándolo hasta la ubicación deseada. Sólo puede mover una columna a la vez.

## Nodo Tabla

El nodo Tabla crea una tabla que enumera los valores de sus datos. Se incluyen todos los campos y todos los valores de la ruta, lo que facilita la inspección de sus valores de datos o su exportación en un formato fácilmente legible. También puede resaltar registros que cumplan ciertas condiciones.

### Pestaña Configuración del nodo Tabla

Figura 6-5  
Nodo Tabla: Pestaña Configuración



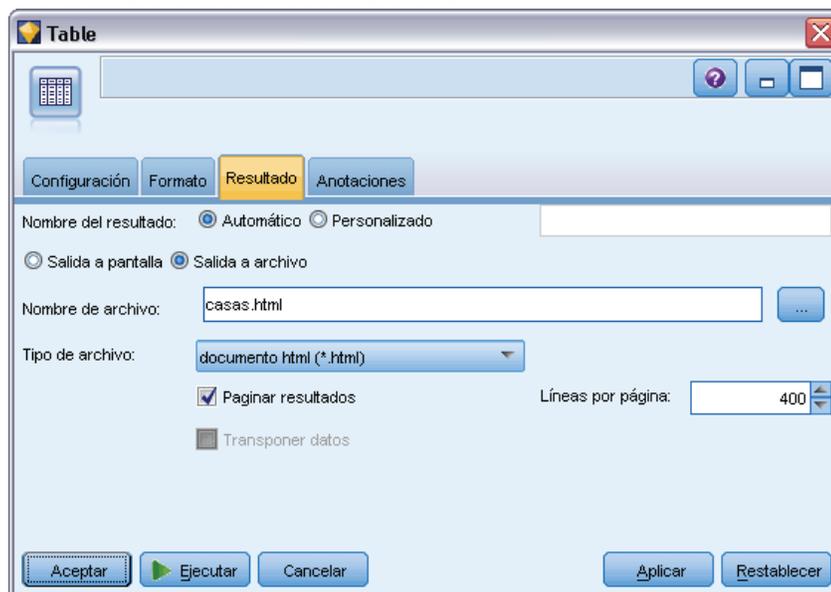
**Resaltar registros en los que.** Puede resaltar registros en la tabla introduciendo una expresión CLEM verdadera para los registros que se van a resaltar. Esta opción se activa solo cuando se selecciona Salida a pantalla.

### ***Pestaña de formato del nodo Tabla***

La pestaña Formato contiene opciones utilizadas para especificar el formato por campos. Esta pestaña se comparte con el nodo Tipo. Si desea obtener más información, consulte el tema [Pestaña Configuración de Formato de campo](#) el p. 151.

### ***Pestaña Resultado del nodo de resultados***

Figura 6-6  
Pestaña Resultado del nodo de resultados



Para los nodos que generan resultados en estilo de tabla, la pestaña Resultados permite especificar el formato y ubicación de los resultados.

**Nombre del resultado.** Especifica el nombre del resultado generado cuando se ejecuta el nodo. Automático selecciona un nombre según el nodo que genera el resultado. De forma opcional, puede seleccionar Personalizado para especificar un nombre diferente.

**Salida a pantalla** (por defecto). Crea un objeto de resultados para ver en línea. El objeto de resultados aparecerá en la pestaña Resultados de la ventana del administrador al ejecutar el nodo de resultados.

**Salida a archivo.** Almacena el resultado en un archivo cuando se ejecuta el nodo. Si selecciona esta opción, introduzca un nombre de archivo (o navegue a un directorio y especifique un nombre de archivo mediante el botón Buscar el archivo) y seleccione un tipo de archivo. Tenga en cuenta que algunos tipos de archivos pueden no estar disponibles para ciertos tipos de resultados.

Los datos son de resultado en el formato de codificación por defecto del sistema, que se especifica en el Panel de control de Windows o, si lo ejecuta en modo distribuido, en el equipo servidor.

- **Datos (delimitados por tabulaciones) (\*.tab).** Esta opción genera un archivo de texto con formato que contiene los valores de datos. Este estilo suele ser útil para generar una representación de texto sin formato de la información que se puede importar a otras aplicaciones. Esta opción está disponible para los nodos Tabla, Matriz y Medias.
- **Datos (delimitados por comas) (\*.dat).** Esta opción genera un archivo de texto delimitado por comas que contiene los valores de datos. Este estilo suele ser útil como una manera rápida de generar un archivo de datos que se puede importar a hojas de cálculo u otras aplicaciones de análisis de datos. Esta opción está disponible para los nodos Tabla, Matriz y Medias.
- **HTML (\*.html)** Esta opción escribe resultado con formato HTML en un archivo o archivos. Para resultados tabulares (de los nodos Tabla, Matriz o Medias), un conjunto de archivos HTML contiene un panel de contenido que muestra nombres de campos y los datos en una tabla HTML. La tabla se puede dividir en varios archivos HTML si el número de filas de la tabla supera la especificación de Líneas por página. En ese caso, el panel de contenido incluye vínculos a todas las páginas de la tabla y ofrece medios para desplazarse por ella. Para resultados no tabulares, se crea un solo archivo HTML que contiene los resultados del nodo.

*Nota:* Si los resultados HTML contienen sólo formato para la primera página, seleccione **Paginar resultados** y ajuste la especificación **Líneas por página** para incluir todos los resultados en una sola página. O, si la plantilla de resultados para los nodos como el nodo Informe contiene etiquetas HTML personalizadas, asegúrese de que ha especificado **Personalizado** como tipo de formato.

- **Archivo de texto (\*.txt)** Esta opción genera un archivo de texto que contiene los resultados. Este estilo suele ser útil para generar resultados que se pueden importar a otras aplicaciones, como procesadores de texto o software de presentación. Esta opción no se encuentra disponible para algunos nodos.
- **Objeto de resultados (\*.cou).** Los objetos de resultados guardados en este formato se pueden abrir y ver en IBM® SPSS® Modeler, añadir a los proyectos y publicar y seguir mediante IBM® SPSS® Collaboration and Deployment Services Repository.

**Vista de resultado.** Para el nodo Medias, puede especificar si se ve por defecto el resultado avanzado o simple. Tenga en cuenta que también puede cambiar de una vista a otra al explorar el resultado generado. Si desea obtener más información, consulte el tema [Explorador de resultados del nodo Medias](#) el p. 440.

**Formato.** Para el nodo Informe, puede seleccionar si al resultado se aplica el formato automáticamente o utilizando el HTML incluido en la plantilla. Seleccione **Personalizado** para permitir formato HTML en la plantilla.

**Título.** Para el nodo Informe, puede especificar un texto de título opcional que aparecerá en la parte superior del resultado del informe.

**Resaltar texto insertado.** Para el nodo Informe, seleccione esta opción para resaltar el texto generado por expresiones CLEM en la plantilla Informe. Si desea obtener más información, consulte el tema [Pestaña Plantilla del nodo Informe](#) el p. 443. Esta opción no es recomendable cuando se utiliza el formato Personalizado.

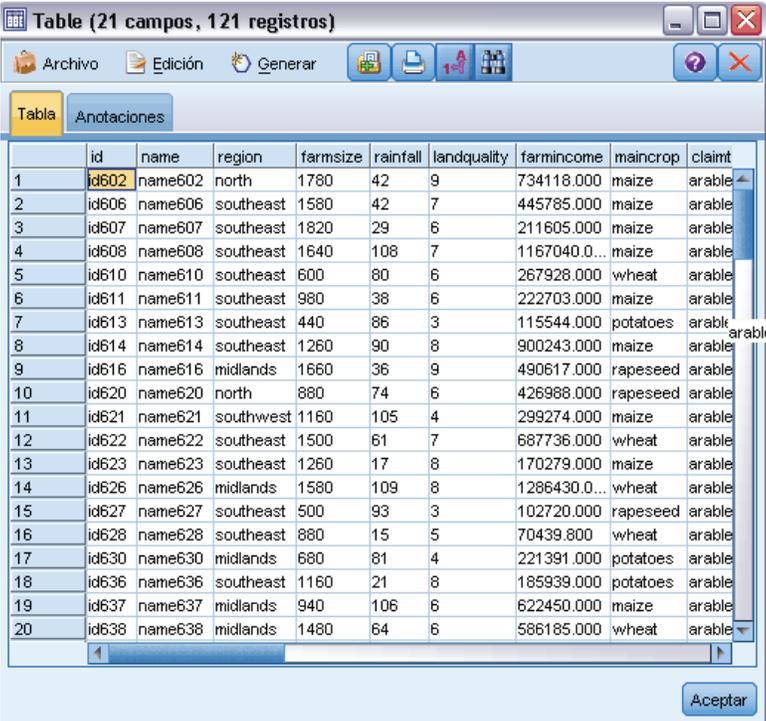
**Líneas por página.** Para el nodo Informe, especifique un número de líneas para incluir en cada página durante el formato Automático del informe de resultados.

**Transponer datos.** Esta opción transpone los datos antes de exportarlos, de manera que las filas representan campos y las columnas registros.

*Nota:* para tablas grandes, las opciones anteriores pueden ser algo ineficaces, especialmente cuando se trabaja con un servidor remoto. En tales casos, utilizar un nodo de resultado de archivo proporciona un rendimiento mucho mayor. Si desea obtener más información, consulte el tema [nodo de exportación Archivo plano](#) en el capítulo 7 el p. 472.

## Explorador de tablas

Figura 6-7  
Ventana del explorador de tablas



	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	claimt
1	id602	name602	north	1780	42	9	734118.000	maize	arable
2	id606	name606	southeast	1580	42	7	445785.000	maize	arable
3	id607	name607	southeast	1820	29	6	211605.000	maize	arable
4	id608	name608	southeast	1640	108	7	1167040.0...	maize	arable
5	id610	name610	southeast	600	80	6	267928.000	wheat	arable
6	id611	name611	southeast	980	38	6	222703.000	maize	arable
7	id613	name613	southeast	440	86	3	115544.000	potatoes	arable
8	id614	name614	southeast	1260	90	8	900243.000	maize	arable
9	id616	name616	midlands	1660	36	9	490617.000	rapeseed	arable
10	id620	name620	north	880	74	6	426988.000	rapeseed	arable
11	id621	name621	southwest	1160	105	4	299274.000	maize	arable
12	id622	name622	southeast	1500	61	7	687736.000	wheat	arable
13	id623	name623	southeast	1260	17	8	170279.000	maize	arable
14	id626	name626	midlands	1580	109	8	1286430.0...	wheat	arable
15	id627	name627	southeast	500	93	3	102720.000	rapeseed	arable
16	id628	name628	southeast	880	15	5	70439.800	wheat	arable
17	id630	name630	midlands	680	81	4	221391.000	potatoes	arable
18	id636	name636	southeast	1160	21	8	185939.000	potatoes	arable
19	id637	name637	midlands	940	106	6	622450.000	maize	arable
20	id638	name638	midlands	1480	64	6	586185.000	wheat	arable

El explorador de tablas muestra datos tabulares y permite realizar operaciones estándar incluyendo seleccionar y copiar casillas, reordenar columnas y guardar e imprimir la tabla. Si desea obtener más información, consulte el tema [Selección de casillas y columnas](#), el p. 397. Son las mismas operaciones que puede realizar cuando visualiza la vista previa de los datos en un nodo.

**Exportación de datos de la tabla.** Puede exportar datos desde el navegador de tablas seleccionando: File > Exportar

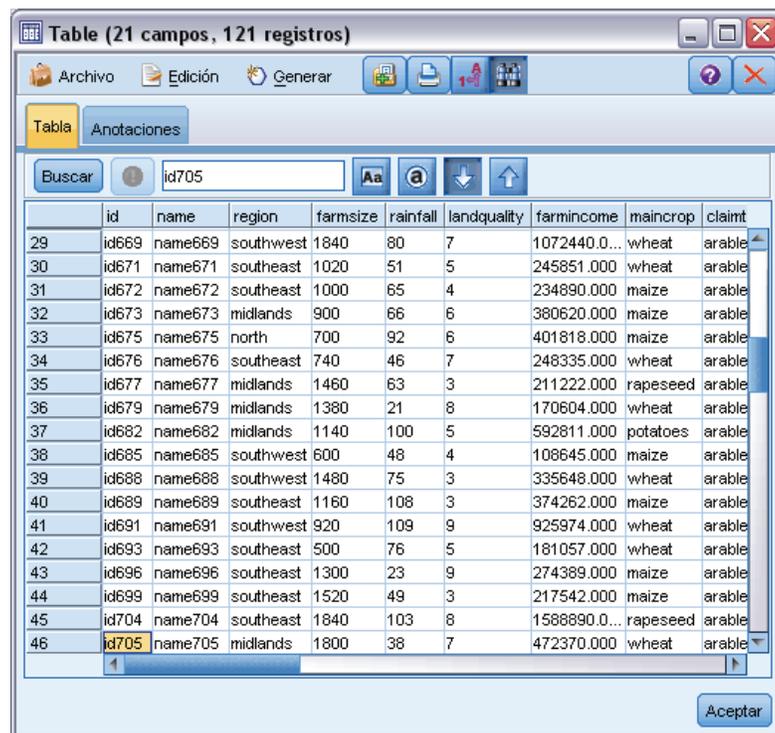
Si desea obtener más información, consulte el tema [Exportación de resultados](#) el p. 396.

Los datos se exportan en el formato de codificación por defecto del sistema, que se especifica en el Panel de control de Windows o, si lo ejecuta en modo distribuido, en el equipo servidor.

**Búsqueda en la tabla.** El botón de búsqueda (con el icono de los prismáticos) de la barra de herramientas principal activa la barra de herramientas de búsqueda, permitiéndole buscar valores específicos en la tabla. Puede buscar hacia adelante o hacia atrás en la tabla, puede especificar una búsqueda que distinga entre mayúsculas y minúsculas (el botón Aa) y puede interrumpir el proceso de búsqueda con el botón de interrupción de búsqueda.

Figura 6-8

Tabla con los controles de búsqueda activados



**Generación de nuevos nodos.** El menú Generar contiene operaciones de generación de nodos.

- **Nodo Seleccionar ("Registros").** Genera un nodo Seleccionar que selecciona los registros para los que se selecciona cualquier casilla de la tabla.
- **Seleccionar ("Y").** Genera un nodo Seleccionar que selecciona los registros que contienen *todos* los valores seleccionados en la tabla.
- **Seleccionar "O").** Genera un nodo Seleccionar que selecciona los registros que contienen *cualquier* valor seleccionado en la tabla.
- **Derivar ("Registros")** Genera un nodo Derivar para crear un nuevo campo de marca. El campo de marca contiene *T* para los registros para los que se selecciona cualquier casilla de la tabla y *F* para los registros restantes.

- **Derivar ("Y").** Genera un nodo Derivar para crear un nuevo campo de marca. El campo de marca contiene *T* para los registros que contienen *todos* los valores seleccionados en la tabla y *F* para los registros restantes.
- **Derivar ("O").** Genera un nodo Derivar para crear un nuevo campo de marca. El campo de marca contiene *T* para los registros que contienen *cualquiera* de los valores seleccionados en la tabla y *F* para los registros restantes.

## Nodo Matriz

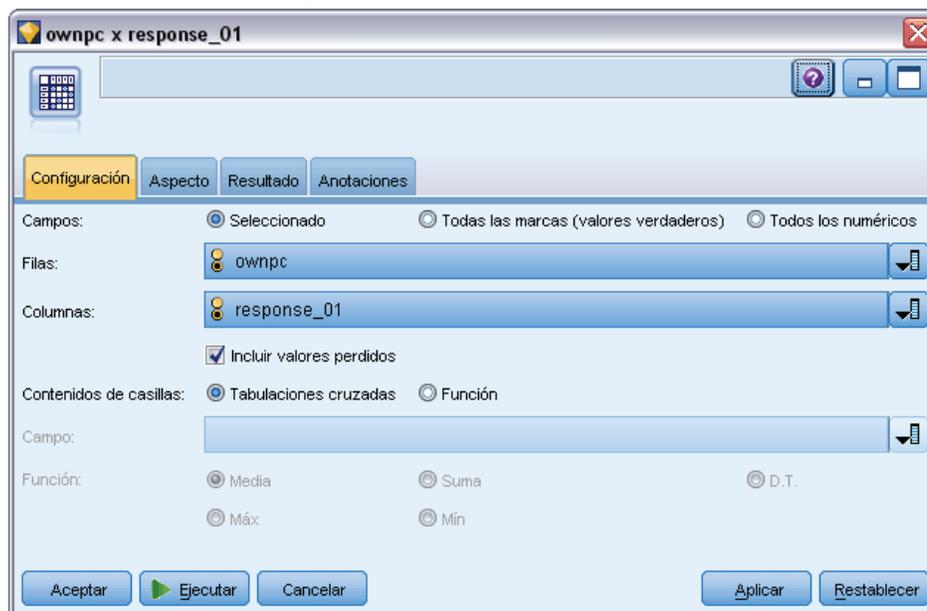
El nodo Matriz permite crear una tabla que muestra las relaciones entre campos. Se suele utilizar normalmente para mostrar las relaciones entre dos campos categóricos (marca, nominal u ordinal), pero también se puede utilizar para mostrar relaciones entre campos continuos (rango numérico).

### Pestaña Configuración del nodo Matriz

La pestaña Configuración permite especificar opciones para la estructura de la matriz.

Figura 6-9

Nodo Matriz: Pestaña Configuración



**Campos.** Seleccione un tipo de selección de campo de entre las siguientes opciones:

- **Seleccionado.** Esta opción permite seleccionar un campo categórico para las filas y otro para las columnas de la matriz. Las filas y columnas de la matriz están definidas por la lista de valores para el campo categórico seleccionado. Las casillas de la matriz contienen los estadísticos de resumen seleccionados más abajo.
- **Todas las marcas (valores verdaderos).** Esta opción solicita una matriz con una fila y una columna para cada campo de marca en los datos. Las casillas de la matriz contienen los recuentos de positivos dobles para cada combinación de marcas. En otras palabras, para una

fila correspondiente a *pan comprado* y una columna correspondiente a *queso comprado*, la casilla en la intersección de esa fila y esa columna contiene el número de registros para los que *pan comprado* y *queso comprado* son verdaderos.

- **Todos los numéricos.** Esta opción solicita una matriz con una fila y una columna para cada campo numérico. Las casillas de la matriz representan la suma de los productos cruzados para el par de campos correspondiente. En otras palabras, para cada casilla de la matriz, los valores para el campo de fila y el campo de columna se multiplican por cada registro y, a continuación, se suman en los registros.

**Incluir valores perdidos.** Incluye los valores perdidos por el sistema (\$null\$) o por el usuario (vacíos) en los resultados de fila y columna. Por ejemplo, si el valor *N/A* se ha definido como perdido por el usuario para el campo de columna seleccionado, se incluirá en la tabla una columna diferente etiquetada con *N/A* (suponiendo que este valor realmente aparezca en los datos) al igual que cualquier otra categoría. Si se anula la selección de esta opción, la columna *N/A* se excluirá independientemente de la frecuencia con que aparezca.

*Nota:* la opción de incluir valores perdidos sólo se aplica cuando los campos seleccionados son de tabulación cruzada. Los valores vacíos se asignan a \$null\$ y se excluyen de la agregación para el campo de la función cuando el modo está Seleccionado y el contenido está definido como Función, y para todos los campos numéricos cuando el modo está definido como Todos los numéricos.

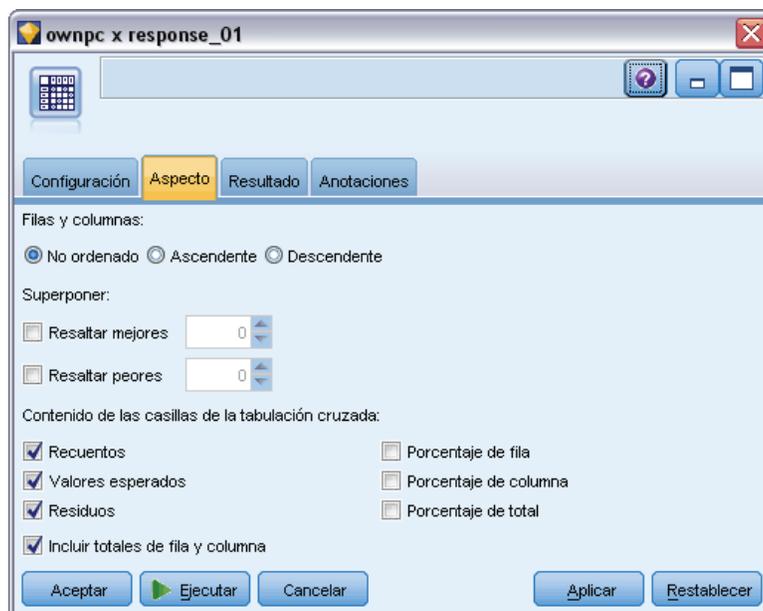
**Contenidos de casillas.** Si ha seleccionado los campos Seleccionado anteriormente, puede especificar el estadístico que se va a utilizar en las casillas de la matriz. Seleccione un estadístico basado en los recuentos o seleccione un campo de superposición para resumir los valores de un campo numérico según los valores de los campos de fila y columna.

- **Tabulaciones cruzadas.** Los valores de las casillas son recuentos y/o porcentajes de los recuentos con la combinación de valores correspondiente. Puede especificar qué resúmenes de tabulación cruzada desea utilizando las opciones de la pestaña Aspecto. El valor chi-cuadrado global es muestra junto con la significación. Si desea obtener más información, consulte el tema [Explorador de resultados de nodo Matriz](#). el p. 406.
- **Función.** Si selecciona una función de resumen, los valores de las casillas son una función de los valores del campo de superposición seleccionados para los casos con valores de fila y columna adecuados. Por ejemplo, si el campo de fila es *Región*, el campo de columna es *Producto* y el campo de superposición es *Ingresos*, entonces la casilla de la fila *Nordeste* y de la columna *Artilugio* contendrán la suma (o media, mínimo o máximo) de los ingresos provenientes de la venta de artilugios en la región nordeste. La función de resumen por defecto es Media. Puede seleccionar otra función para resumir el campo de la función. Las opciones incluyen Media, Suma, D.T. (desviación típica), Máx (máximo) y Mín (mínimo).

## ***Pestaña Aspecto del nodo Matriz***

La pestaña Aspecto permite controlar las opciones de ordenación y resaltado de la matriz, así como los estadísticos presentados para las matrices de tabulación cruzada.

Figura 6-10  
Nodo Matriz: Pestaña Aspecto



**Filas y columnas.** Controla la ordenación de los encabezados de fila y columna en la matriz. El valor por defecto es No ordenado. Seleccione Ascendente o Descendente para ordenar los encabezados de fila y columna en la dirección especificada.

**Superponer.** Permite resaltar los valores extremos de la matriz. Los valores se resaltan según los recuentos de las casillas (para matrices de tabulación cruzada) o los valores calculados (para matrices de función).

- **Resaltar mejores.** Puede solicitar que se resalten (en rojo) los valores más altos de la matriz. Especifique el número de valores que desee resaltar.
- **Resaltar peores.** Puede solicitar que se resalten (en verde) los valores más bajos de la matriz. Especifique el número de valores que desee resaltar.

*Nota:* para las dos opciones de resaltado, los empates pueden hacer que se resalten más valores de los solicitados. Por ejemplo, si tiene una matriz con seis ceros en las casillas y solicita Resaltar peores 5, se resaltarán los seis ceros.

**Contenido de las casillas de la tabulación cruzada.** Para las tabulaciones cruzadas, puede especificar los estadísticos de resumen contenidos en la matriz para las matrices de tabulación cruzada. Estas opciones no están disponibles si se selecciona la opción Todos los numéricos o la opción Función en la pestaña Configuración.

- **Recuentos.** Las casillas incluyen el número de registros con el valor de la fila que tiene el valor de la columna correspondiente. Éste es sólo el contenido por defecto de la casilla.
- **Valores esperados.** El valor esperado para el número de registros de la casilla, suponiendo que no haya relación entre las filas y las columnas. Los valores esperados se basan en la siguiente fórmula:

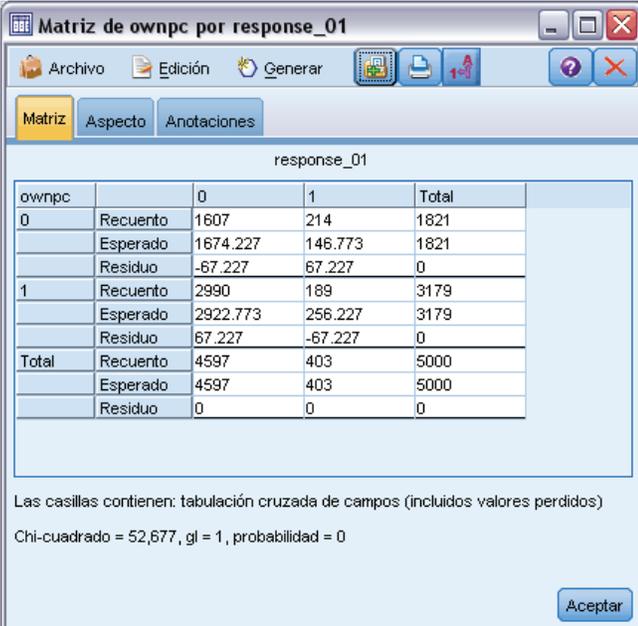
$$p(\text{valor de fila}) * p(\text{valor de columna}) * \text{número total de registros}$$

- **Residuos.** Diferencia entre los valores observados y esperados.
- **Porcentaje de fila.** Porcentaje de todos los registros con el valor de la fila que tiene el valor de la columna correspondiente. La suma de porcentajes es 100 en las filas.
- **Porcentaje de columna.** Porcentaje de todos los registros con el valor de la columna que tiene el valor de la fila correspondiente. La suma de porcentajes es 100 en las columnas.
- **Porcentaje de total.** Porcentaje de todos los registros con la combinación de valor de columna y valor de fila. La suma de porcentajes es 100 en toda la matriz.
- **Incluir totales de fila y columna.** Añade una fila y una columna a la matriz para los totales de columna y fila.
- **Aplicar configuración.** (Sólo en el explorador de resultados). Permite cambiar el aspecto de los resultados del nodo Matriz sin necesidad de cerrar y volver a abrir el explorador de resultados. Realice los cambios necesarios en esta pestaña del explorador de resultados, pulse en este botón y, a continuación, seleccione la pestaña Matriz para ver el efecto de los cambios.

### Explorador de resultados de nodo Matriz.

El explorador de matrices muestra datos de tabulación cruzada y permite realizar operaciones en la matriz, incluida la selección de casillas, copia de la matriz en el Portapapeles en parte o en su totalidad, generación de nuevos nodos basados en la selección de la matriz y almacenamiento e impresión de la matriz. El explorador de matrices también se puede utilizar para mostrar resultados de determinados modelos, como los modelos bayesianos Naive de Oracle.

Figura 6-11  
Explorador de matrices



Matriz de ownpc por response\_01

Archivo Edición Generar

Matriz Aspecto Anotaciones

response\_01

ownpc		0	1	Total
0	Recuento	1607	214	1821
	Esperado	1674.227	146.773	1821
	Residuo	-67.227	67.227	0
1	Recuento	2990	189	3179
	Esperado	2922.773	256.227	3179
	Residuo	67.227	-67.227	0
Total	Recuento	4597	403	5000
	Esperado	4597	403	5000
	Residuo	0	0	0

Las casillas contienen: tabulación cruzada de campos (incluidos valores perdidos)

Chi-cuadrado = 52,677, gl = 1, probabilidad = 0

Aceptar

Los menús Edición y Archivo proporcionan las opciones habituales para impresión, almacenamiento y exportación de resultados, así como selección y copia de datos. Si desea obtener más información, consulte el tema [Visualización de resultados](#) el p. 393.

**Chi-cuadrado.** Para una tabulación cruzada de dos campos categóricos, el chi-cuadrado de Pearson global aparece también bajo la tabla. Esta prueba indica la probabilidad de que los dos campos no estén relacionados, a partir de la diferencia entre los recuentos observados y los recuentos que esperaría si no existiera relación. Por ejemplo, si no existiera relación entre la satisfacción del cliente y la ubicación de la tienda, se esperarían unas tasas de satisfacción similares en todas las tiendas. Pero si los clientes de ciertas tiendas presentasen sistemáticamente tasas superiores a las de otras, se podría sospechar que no es coincidencia. A mayor diferencia, menor probabilidad de que el resultado sea un error de muestreo aleatorio solo.

- La prueba de chi-cuadrado indica la probabilidad de que dos campos no estén relacionados, en cuyo caso cualquier diferencia entre las frecuencias observadas y esperadas es el resultado del azar solo. Si esta probabilidad es muy pequeña (normalmente menor del 5%) la relación entre los dos campos se considera significativa.
- Si sólo hay una columna o una fila (una prueba de chi-cuadrado unidireccional), el número de grados de libertad es el número de casillas menos uno. En una chi-cuadrado bidireccional, el número de grados de libertad es el número de filas menos el número de columnas menos uno.
- Tenga cuidado al interpretar el estadístico chi cuadrado si alguna de las frecuencias de las casillas esperadas es menor de cinco.
- La prueba de chi-cuadrado sólo está disponible para una tabulación cruzada de dos campos. (Cuando en la pestaña Configuración se selecciona Todas las marcas o Todos los numéricos, esta prueba no se muestra.)

**Menú Generar.** El menú Generar contiene operaciones de generación de nodos. Estas operaciones sólo están disponibles para matrices de tabulación cruzada y debe haber al menos una casilla seleccionada en la matriz.

- **Nodo Seleccionar** Genera un nodo Seleccionar que selecciona los registros que coinciden con cualquier casilla de la matriz.
- **Nodo Derivar (Marca).** Genera un nodo Derivar para crear un nuevo campo de marca. El campo de marca contiene  $T$  para los registros que coinciden con cualquier casilla seleccionada en la matriz y  $F$  para los registros restantes.
- **Nodo Derivar (Conjunto).** Genera un nodo Derivar para crear un nuevo campo nominal. El campo nominal contiene una categoría para cada conjunto contiguo de casillas seleccionadas de la matriz.

## ***Nodo Análisis***

El nodo Análisis permite evaluar la capacidad de un modelo para generar pronósticos precisos. Los nodos Análisis realizan varias comparaciones entre los valores pronosticados y los valores (su campo de objetivo) reales para uno o más nugget de modelo. Los nodos Análisis también se pueden utilizar para comparar modelos predictivos con otros modelos predictivos.

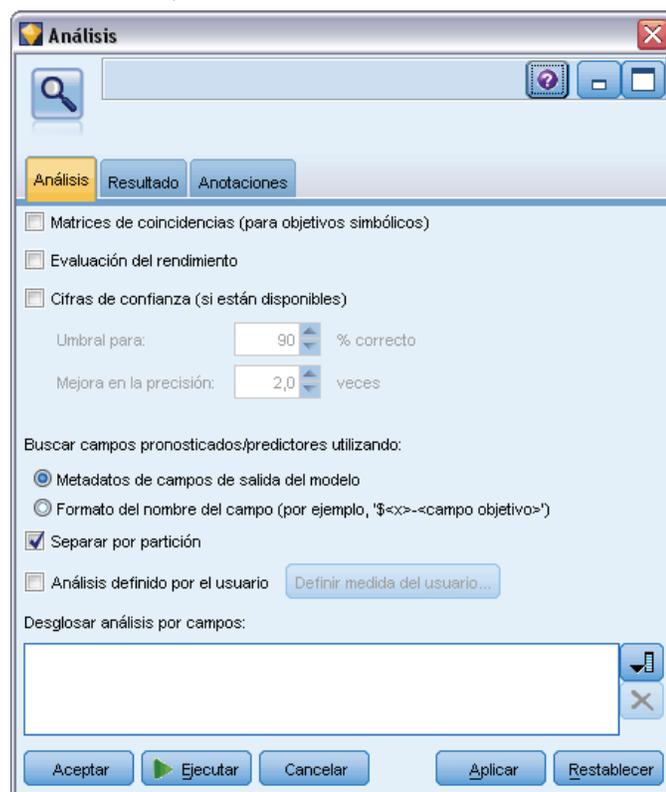
Cuando ejecuta un nodo Análisis, se añade automáticamente un resumen de los resultados de análisis a la sección Análisis de la pestaña Resumen para cada nugget de modelo de la ruta ejecutada. Los resultados de análisis detallados aparecen en la pestaña Resultados de la ventana del administrador o se pueden escribir directamente en un archivo.

*Nota:* ya que los nodos Análisis comparan valores pronosticados con valores reales, sólo son útiles con modelos supervisados (aquellos que requieren un campo de objetivo). Para los modelos sin supervisar como los algoritmos de conglomerados, no existen resultados reales disponibles como base de comparación.

## Pestaña Análisis del nodo Análisis

La pestaña análisis permite especificar los detalles del análisis.

Figura 6-12  
Nodo Análisis: pestaña Análisis



**Matrices de coincidencias (para objetivos simbólicos o categóricos).** Muestra el patrón de coincidencias entre cada campo generado (pronosticado) y su campo objetivo para objetivos categóricos (marca, nominal u ordinal). Se muestra una tabla con filas definidas por valores reales y columnas definidas por valores pronosticados, con el número de registros que tienen ese patrón en cada casilla. Esto es útil para identificar errores sistemáticos en los pronósticos. Si existe más de un campo generado relacionado con el mismo campo de salida pero generado por modelos distintos, los casos en los que estos campos concuerdan y no concuerdan se cuentan y se

muestran los totales. En los casos en los que concuerdan, se muestra otro conjunto de estadísticos correcto/incorrecto.

**Evaluación del rendimiento.** Muestra estadísticos de evaluación del rendimiento para modelos con resultados categóricos. Este estadístico, mostrado para cada categoría de los campos de salida, es una medida del contenido de información medio (en bits) del modelo para pronosticar registros pertenecientes a dicha categoría. Se tiene en cuenta la dificultad del problema de clasificación, de forma que los pronósticos precisos para categorías inusuales obtendrán un índice de evaluación del rendimiento mayor que los pronósticos precisos para categorías comunes. Si el modelo no hace más que adivinar una categoría, el índice de evaluación del rendimiento para esa categoría será 0.

**Cifras de confianza (si están disponibles).** Para los modelos que generan un campo de confianza, esta opción muestra estadísticos de los valores de confianza y su relación con los pronósticos. Existen dos configuraciones para esta opción:

- **Umbral para.** Muestra el nivel de confianza por encima del cual la precisión será el porcentaje especificado.
- **Mejora en la precisión.** Muestra el nivel de confianza por encima del cual la precisión se mejorará en el valor especificado. Por ejemplo, si la precisión total es del 90% y esta opción está definida como 2,0, el valor mostrado será la confianza necesaria para una precisión del 95%.

**Buscar campos pronosticados/predictores utilizando.** Determina cómo se establece la correspondencia entre los campos pronosticados y el campo objetivo original.

- **Metadatos de campos de salida del modelo.** Establece la correspondencia entre los campos pronosticados y el objetivo basándose en información del campo Modelo, lo que permite coincidencias incluso si se ha cambiado el nombre de un archivo pronosticado. También se puede acceder a la información del campo Modelo para cualquier campo pronosticado desde el cuadro de diálogo Valores mediante un nodo Tipo. Si desea obtener más información, consulte el tema [Utilización del cuadro de diálogo de valores](#) en el capítulo 4 el p. 143.
- **Formato del nombre del campo.** Establece la correspondencia entre campos basándose en la convención de nomenclatura. Por ejemplo, los valores pronosticados generados por un nugget de modelo C5.0 para un objetivo denominado *respuesta* deben encontrarse en un campo denominado *\$C-respuesta*.

**Separar por partición.** Si se utiliza un campo de partición para dividir los registros en muestreos de entrenamiento, comprobación y validación, seleccione esta opción para mostrar los resultados de forma separada en cada partición. Si desea obtener más información, consulte el tema [Nodo Partición](#) en el capítulo 4 el p. 205.

*Nota:* al separar en particiones, los registros con valores nulos en el campo de partición se excluyen del análisis. Esto nunca debería ser un problema si se utiliza un nodo Partición ya que este tipo de nodos no genera valores nulos.

**Análisis definido por el usuario.** Puede especificar su propio cálculo de análisis para utilizarlo al evaluar sus modelos. Utilice expresiones CLEM para especificar lo que debe calcularse para cada registro y cómo combinar las puntuaciones de nivel de registro en una puntuación global. Utilice las funciones @TARGET y @PREDICTED para hacer referencia al valor objetivo (resultado real) y al valor pronosticado, respectivamente.

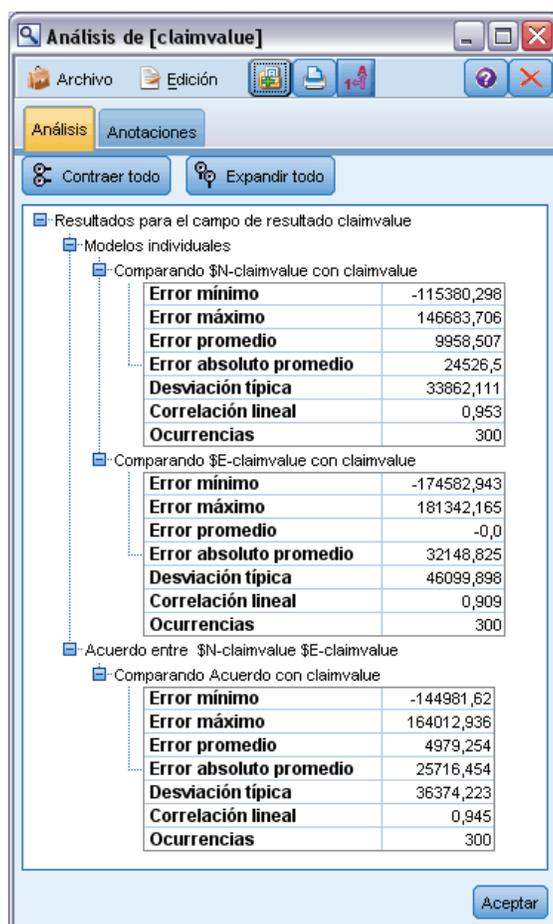
- **Si.** Especifique una expresión condicional si necesita utilizar cálculos distintos dependiendo de alguna condición.
- **Entonces.** Especifique el cálculo si la condición Si es verdadera.
- **En caso contrario.** Especifique el cálculo si la condición Si es falsa.
- **Utilizar.** Seleccione un estadístico para calcular una puntuación global a partir de puntuaciones individuales.

**Desglosar análisis por campos.** Muestra los campos categóricos disponibles para desglosar el análisis. Además del análisis global, se mostrará un análisis independiente para cada categoría de campo de desglose.

## Explorador de resultados de análisis

El explorador de resultados de análisis permite ver los resultados de ejecutar un nodo Análisis. Las opciones normales de almacenamiento, exportación e impresión están disponibles en el menú Archivo. Si desea obtener más información, consulte el tema [Visualización de resultados](#) el p. 393.

Figura 6-13  
Explorador de resultados de análisis



Cuando se examina el resultado del análisis por primera vez, los resultados se expanden. Para ocultar los resultados cuando haya terminado de consultarlos, utilice el control de expansión situado a la izquierda del elemento para contraer los resultados específicos que desee ocultar o pulse en el botón Contraer todo para contraer todos los resultados. Para volver a ver los resultados tras contraerlos, utilice el control de expansión situado a la izquierda del elemento para mostrar los resultados o pulse en el botón Expandir todo para mostrar todos los resultados.

**Resultados para el campo de salida.** El resultado del análisis contiene una sección para cada campo de salida con un campo de pronóstico correspondiente creado por un modelo generado.

**Comparando.** Dentro de la sección del campo de salida hay una subsección para cada campo de pronóstico asociado con dicho campo de salida. Para campos de salida categóricos, el nivel superior de esta sección contiene una tabla que muestra el número y el porcentaje de pronósticos correctos e incorrectos y el número total de registros en la ruta. Para los campos de salida numéricos, esta sección muestra la siguiente información:

- **Error mínimo.** Muestra el error mínimo (diferencia entre los valores observados y los pronosticados).
- **Error máximo.** Muestra el error máximo.
- **Error promedio.** Muestra la media (promedio) de errores en todos los registros. Esto indica si hay un **sesgo** sistemático (una tendencia mayor a sobrestimar que a subestimar, o viceversa) en el modelo.
- **Error absoluto promedio.** Muestra la media de valores absolutos de los errores en todos los registros. Indica la magnitud de error media, independientemente de la dirección.
- **Desviación típica.** Muestra la desviación típica de los errores.
- **Correlación lineal.** Muestra la correlación lineal entre los valores reales y los pronosticados. Este estadístico varía entre  $-1,0$  y  $1,0$ . Los valores cercanos a  $+1,0$  indican una fuerte asociación positiva, de forma que los valores pronosticados altos están asociados con valores reales altos y los valores pronosticados bajos están asociados con valores reales bajos. Los valores cercanos a  $-1,0$  indican una fuerte asociación negativa, de forma que los valores pronosticados altos están asociados con valores reales bajos y viceversa. Los valores cercanos a  $0,0$  indican una asociación débil, de forma que los valores pronosticados son más o menos independientes de los valores reales. *Nota:* Una entrada en blanco aquí indica que la correlación lineal no se puede computar en este caso porque los valores reales o los valores pronosticados son constantes.
- **Ocurrencias.** Muestra el número de registros utilizados en el análisis.

**Matriz de coincidencias.** Para los campos de salida categóricos, si ha solicitado una matriz de coincidencias en las opciones de análisis, aparecerá una subsección aquí que contenga la matriz. Las filas representan valores observados reales y las columnas valores pronosticados. La casilla de la tabla indica el número de registros para cada combinación de valores reales y pronosticados.

**Evaluación del rendimiento.** Para los campos de salida categóricos, si ha solicitado estadísticos de evaluación del rendimiento en las opciones de análisis, los resultados de evaluación del rendimiento aparecerán aquí. Cada categoría de resultado aparece con su estadístico de evaluación del rendimiento.

**Informe de valores de confianza.** Para los campos de salida categóricos, si ha solicitado valores de confianza en las opciones de análisis, los valores aparecen aquí. Los siguientes estadísticos se muestran para los valores de confianza modelo:

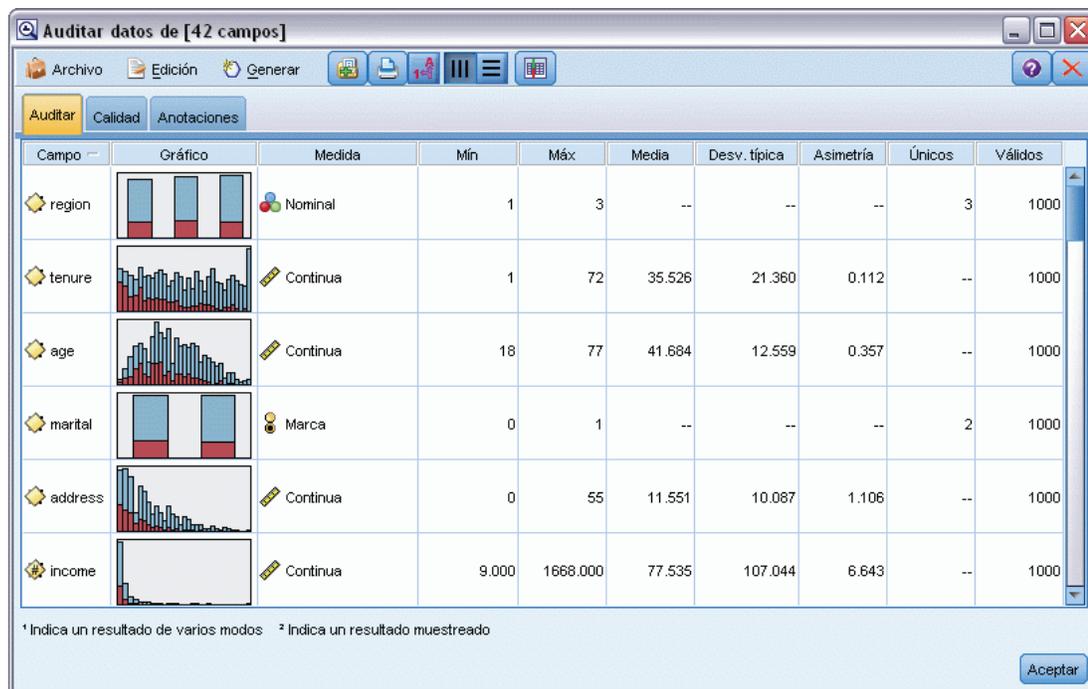
- **Rango.** Muestra el rango (los valores máximos y mínimos) de valores de confianza para registros en los datos de la ruta.
- **Media para correctos.** Muestra la confianza media para los registros que se han clasificado correctamente.
- **Media para incorrectos.** Muestra la confianza media para los registros que se han clasificado de forma incorrecta.
- **Siempre correctos por encima de.** Muestra el umbral de confianza por encima del cual los pronósticos son siempre correctos y el porcentaje de casos que cumplen este criterio.
- **Siempre incorrectos por debajo de.** Muestra el umbral de confianza por debajo del cual los pronósticos son siempre incorrectos y muestra el porcentaje de casos que cumple este criterio.
- **X% Precisión por encima de.** Muestra el nivel de confianza en el que la precisión es X%. X es aproximadamente el valor especificado para Umbral para en las opciones de Análisis. Para algunos modelos y conjuntos de datos, no se puede elegir un valor de confianza que ofrezca el umbral exacto especificado en las opciones (normalmente debido a los conglomerados de casos similares con el mismo valor confianza cerca del umbral). El umbral que se muestra es el valor más cercano al criterio de precisión especificado que se puede obtener con un solo umbral de valor de confianza.
- **X Veces correctas por encima de.** Muestra el valor de confianza en el cual la precisión es X veces mejor de lo que es para el conjunto de datos global. X es aproximadamente el valor especificado para Mejora en la precisión en las opciones de Análisis.

**Acuerdo entre.** Si dos o más modelos generados que pronostican el mismo campo de salida se incluyen en la ruta, también verá estadísticos en el **acuerdo** entre pronósticos generados por los modelos. Esto incluye el número y el porcentaje de registros en los que concuerdan los pronósticos (para campos de resultado categóricos) o estadísticos de resumen de errores (para campos de salida continuos). Para campos categóricos, incluye un análisis de pronósticos comparado con valores reales para el subconjunto de registros en los que concuerdan los modelos (generan el mismo valor pronosticado).

## ***Nodo Auditar datos***

El nodo Auditar datos permite echar un primer vistazo exhaustivo a los datos introducidos en IBM® SPSS® Modeler, presentados en una matriz fácil de leer que se puede ordenar y utilizar para generar nodos de preparación de datos y gráficos de tamaño completo.

Figura 6-14  
explorador de auditoría de datos

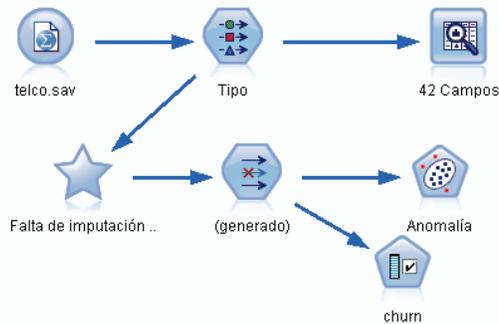


- La pestaña Auditar muestra un informe que proporciona estadísticos de resumen, histogramas y gráficos de distribución que pueden ser útiles para obtener una primera idea de los datos. El informe también muestra el icono de almacenamiento delante del nombre de campo.
- La pestaña Calidad del informe de auditoría muestra información sobre valores extremos, atípicos y perdidos, y proporciona herramientas para tratar dichos valores.

### Uso del nodo Auditar datos

El nodo Auditar datos se puede conectar directamente a un nodo de origen o a nodos posteriores a un nodo Tipo instanciado. También puede generar varios nodos de preparación de datos en base a los resultados. Por ejemplo, puede generar un nodo Filtro que excluya los campos con demasiados valores perdidos como para que sean útiles para el modelado, y generar un Supernodo que impute los valores perdidos de uno de los campos que permanecen o de todos ellos. Aquí es donde se demuestra la potencia real de la auditoría, que permite no sólo evaluar el estado actual de los datos, sino también realizar acciones en función de la evaluación.

Figura 6-15  
Ruta con Supernodo de valores perdidos

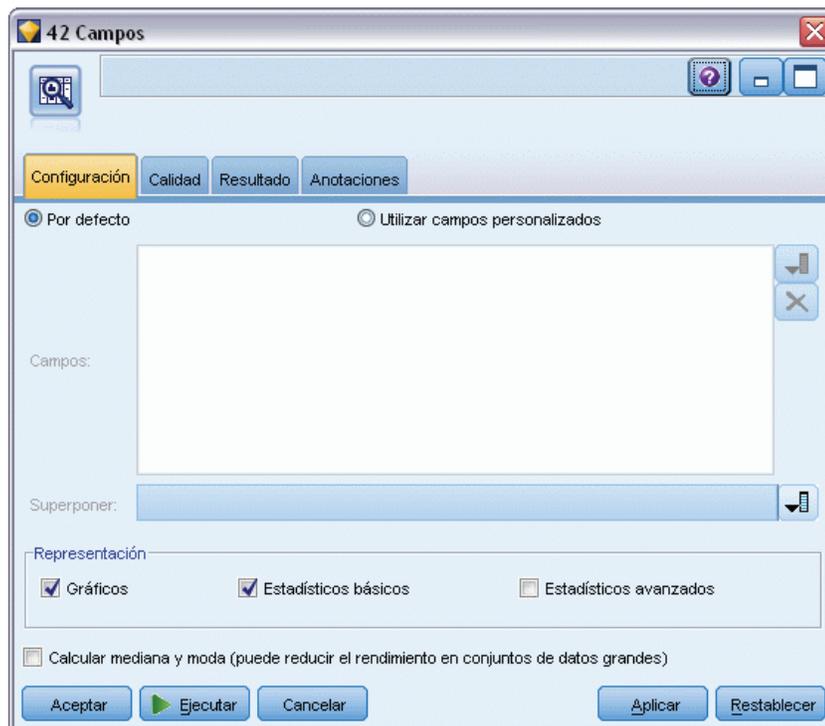


**Cribado o muestreo de datos.** Ya que una auditoría inicial es especialmente efectiva cuando se trabaja con “datos de gran tamaño”, puede utilizarse un nodo Muestrear para reducir el tiempo de procesamiento durante la exploración inicial, mediante la selección de un solo subconjunto de registros. El nodo Auditar datos también se puede utilizar junto con nodos como Selección de características y Detección de anomalías en las fases de exploración del análisis.

### ***Pestaña Configuración del nodo Auditar datos***

La pestaña Configuración permite especificar los parámetros básicos para la auditoría.

Figura 6-16  
Nodo Auditar datos: Pestaña Configuración



**Por defecto.** Puede conectar fácilmente el nodo a la ruta y pulsar en Ejecutar para generar un informe de auditoría para todos los campos con la configuración por defecto, de la siguiente forma:

- Si no se ha configurado el nodo Tipo, todos los campos se incluyen en el informe.
- Si hay configuración del nodo Tipo (independientemente de si está o no instanciada) todos los campos *Entrada*, *Objetivo* y *Ambos* se incluyen en la representación. Si hay un solo campo *Objetivo*, se utiliza como campo de superposición. Si hay más de un campo *Objetivo* especificado, no se especifica ninguna superposición por defecto.

**Utilizar campos personalizados.** Elija esta opción para seleccionar manualmente los campos. Utilice el botón de selección de campos de la derecha para seleccionar los campos de forma individual o por tipo.

**Campo de superposición.** El campo de superposición se emplea para dibujar los gráficos en miniatura que aparecen en el informe de auditoría. En los campos continuos (rango numérico), también se calculan los estadísticos bivariados (covarianza y correlación). Si se presenta un solo campo *Objetivo* basado en la configuración del nodo Tipo, se utiliza como campo de superposición por defecto, tal como se describe arriba. Si lo prefiere, puede seleccionar Utilizar campos personalizados para especificar una superposición.

**Mostrar.** Permite especificar si los gráficos van a estar disponibles en los resultados y seleccionar los estadísticos que se mostrarán por defecto.

- **Gráficos.** Muestra un gráfico para cada campo seleccionado; un gráfico de distribución (de barras), un histograma o un diagrama de dispersión, según los datos. Los gráficos aparecen en miniatura en el informe inicial, pero también se pueden generar gráficos a tamaño completo y nodos de gráficos. Si desea obtener más información, consulte el tema [Explorador de resultados de auditoría de datos](#) el p. 417.
- **Estadísticos básicos/avanzados.** Especifica el nivel de estadísticos que se muestran por defecto en los resultados. Esta configuración determina la representación inicial, ya que todos los estadísticos están disponibles en los resultados, independientemente de esta configuración. Si desea obtener más información, consulte el tema [Mostrar estadísticos](#) el p. 419.

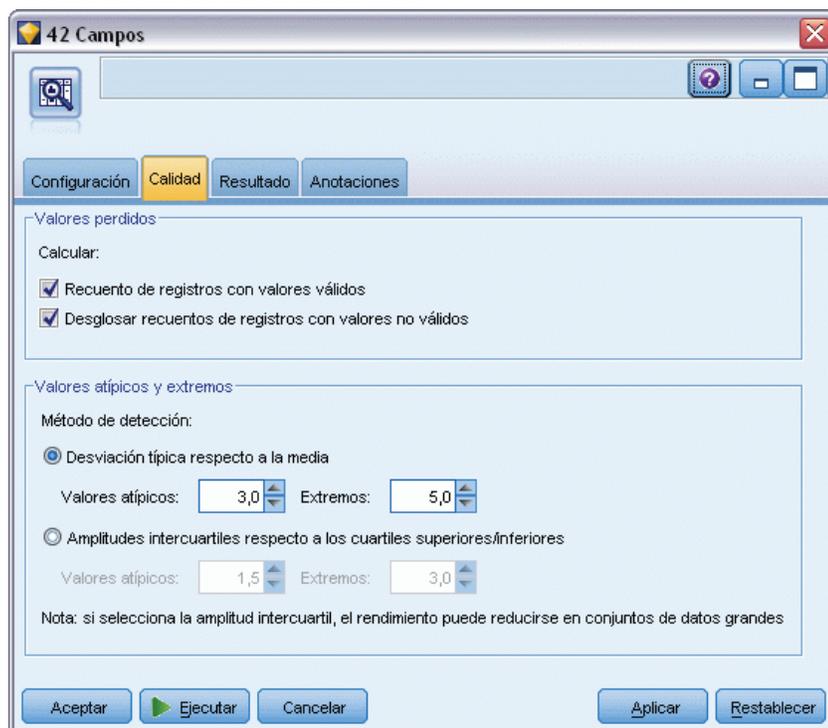
**Mediana y moda.** Calcula la mediana y la moda para todos los campos del informe. Recuerde que cuando se trabaja con grandes conjuntos de datos, estos estadísticos pueden aumentar el tiempo de procesamiento, puesto que su cálculo requiere más tiempo. Sólo en el caso de la mediana, ocasionalmente el valor mostrado puede estar basado en una muestra de 2.000 registros (en lugar del conjunto de datos completo). Este muestreo se realiza por campos cuando, de otro modo, podrían excederse los límites de memoria. Cuando se realice el muestreo, los resultados se etiquetarán como tales en la salida (*Mediana de muestras* en lugar de sólo *Mediana*). El resto de estadísticos se calcula siempre utilizando el conjunto de datos completo.

**Campos vacíos o sin tipo.** Cuando se utilizan con datos instanciados, los campos sin tipo no se incluyen en el informe de auditoría. Para incluir los campos sin tipo (incluidos los campos vacíos), seleccione Borrar todos los valores en los nodos Tipo anteriores de la ruta. De esta forma se garantiza que los datos no estén instanciados, lo que hace que todos los campos se incluyan en el informe. Por ejemplo, puede resultar útil para obtener una lista completa de todos los campos o para generar un nodo Filtro que excluya los campos vacíos. Si desea obtener más información, consulte el tema [Filtrado de campos con datos perdidos](#) el p. 424.

## Pestaña Calidad de Auditar datos

La pestaña Calidad del nodo Auditar datos proporciona opciones para tratar los valores perdidos, atípicos y extremos.

Figura 6-17  
Pestaña Calidad del nodo Auditar datos



### Valores perdidos

- **Recuento de registros con valores válidos.** Seleccione esta opción para mostrar el número de registros con valores válidos para cada campo evaluado. Recuerde que los valores nulos (no definidos), los valores vacíos, los espacios en blanco y las cadenas vacías siempre se tratan como valores no válidos.
- **Desglosar recuentos de registros con valores no válidos.** Seleccione esta opción para mostrar el número de registros con cada tipo de valor no válido para cada campo.

### Valores atípicos y extremos

Métodos de detección de valores atípicos y extremos. Se admiten dos métodos:

**Desviación típica de la media.** Detecta los valores atípicos y extremos a partir del número de desviaciones típicas de la media. Por ejemplo, si tiene un campo con una media de 100 y una desviación típica de 10, podría especificar 3,0 para indicar que cualquier valor inferior a 70 o superior a 130 debe tratarse como atípico.



- Ver valores y rangos para los campos pulsando dos veces en un campo de las columnas Medición y Únicos.
- Utilice la barra de herramientas o el menú Edición para mostrar u ocultar etiquetas de valor, o seleccione los estadísticos que desea mostrar. Si desea obtener más información, consulte el tema [Mostrar estadísticos](#) el p. 419.
- Verifique los iconos de almacenamiento a la izquierda de los nombres de campo. Almacenamiento describe la forma en la que los datos se almacenan en un campo. Por ejemplo, un campo con valores 1 y 0 almacena datos enteros. Esto es diferente al nivel de medición, que describe el uso de los datos y no afecta al almacenamiento. Si desea obtener más información, consulte el tema [Configuración de almacenamiento de campos y formato](#) en el capítulo 2 el p. 32.

### Visualización y generación de gráficos

Si no hay ninguna superposición seleccionada, la pestaña Auditar muestra gráficos de barras (para campos nominales o marca) o histogramas (campos continuos).

Figura 6-19  
Extracto de resultados de auditoría sin un campo de superposición

Campo	Gráfico	Tipo	Mín	Máx	Media	Desv. típica	Asimetría	Únicos	Válidos
region		Conjunto	1	3	--	--	--	3	1000
tenure		Rango	1	72	35.526	21.360	0.112	--	1000

Para la superposición de un campo nominal o marca, los gráficos se colorean con los valores de la superposición.

Figura 6-20  
Extracto de resultados de auditoría con una superposición de campo nominal

Campo	Gráfico	Tipo	Mín	Máx	Media	Desv. típica	Asimetría	Únicos	Válidos
region		Conjunto	1	3	--	--	--	3	1000
tenure		Rango	1	72	35.526	21.360	0.112	--	1000

Para una superposición de campo nominal, se generan diagramas de dispersión bidimensionales en lugar de histogramas y barras unidimensionales. En ese caso, el eje x establece correspondencia con el campo de superposición, permitiendo ver la misma escala en todos los ejes x conforme lee la tabla.

Figura 6-21  
Extracto de resultados de auditoría con una superposición de campo continuo

Campo	Gráfico	Tipo	Mín	Máx	Media	Correlation	Correlation T	Correlation T df.
region		Conjunto	1	3	--	--	--	--
tenure		Rango	1	72	35.526	0.490	17.768	998.000

- Para los campos nominales o marca, mantenga el cursor del ratón sobre una barra para mostrar el valor subyacente o la etiqueta en un texto de información.
- Para los campos nominales o marca, utilice la barra de herramientas para cambiar la orientación de los gráficos en miniatura entre horizontal y vertical.
- Para generar un gráfico a tamaño completo a partir de una miniatura, pulse dos veces en la miniatura o seleccione la miniatura y elija Resultado gráfico en el menú Generar. *Nota:* si hay un gráfico en miniatura basado en datos muestreados, el gráfico generado contendrá todos los casos si la ruta de datos original todavía está abierta.

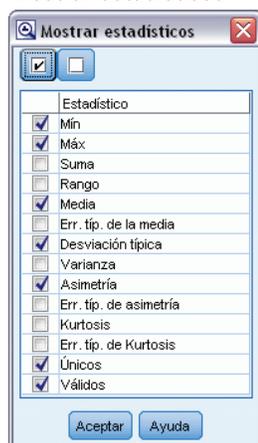
Sólo puede generar un gráfico si el nodo Auditar datos que creó el resultado está conectado a la ruta.

- Para generar el correspondiente nodo de gráficos, seleccione uno o más campos en la pestaña Auditar y seleccione Nodo Gráfico en el menú Generar. El nodo resultante se añade al lienzo de rutas y se puede utilizar para volver a crear el gráfico cada vez que se ejecute la ruta.
- si un conjunto de superposiciones tiene más de 100 valores, aparece una advertencia y la superposición no se incluye.

### Mostrar estadísticos

El cuadro de diálogo Mostrar estadísticos permite seleccionar los estadísticos que se muestran en la pestaña Auditar. La configuración inicial se especifica en el nodo Auditar datos. Si desea obtener más información, consulte el tema [Pestaña Configuración del nodo Auditar datos](#) el p. 414.

Figura 6-22  
Mostrar estadísticos



**Mínimo.** Valor más pequeño de una variable numérica.

**Máximo.** El mayor valor de una variable numérica.

**Suma.** Suma o total de todos los valores, a lo largo de todos los casos que no tengan valores perdidos.

**Rango.** Diferencia entre los valores mayor y menor de una variable numérica; el máximo menos el mínimo.

**Media.** Una medida de tendencia central. El promedio aritmético, la suma dividida por el número de casos.

**Error típico de la media.** Es una medida de cuánto puede variar el valor de la media entre varias muestras tomadas de la misma distribución. Puede utilizarse para comparar de forma aproximada la media observada respecto a un valor hipotetizado (es decir, se puede concluir que los dos valores son distintos si la diferencia entre ellos, dividida por el error típico, es menor que -2 o mayor que +2).

**desviación típica.** Una medida de la dispersión en torno a la media, igual a la raíz cuadrada de la varianza. La desviación típica se mide en las mismas unidades que la variable original.

**Varianza.** Es una medida de dispersión en torno a la media, igual a la suma de las desviaciones al cuadrado respecto a la media, dividida por el número de casos menos 1. La varianza se mide en unidades que son el cuadrado de las de la variable en cuestión.

**Asimetría.** Medida de la asimetría de una distribución La distribución normal es simétrica y tiene un valor de asimetría igual a 0. Una distribución que tenga una asimetría positiva significativa tiene una cola derecha larga. Una distribución que tenga una asimetría negativa significativa tiene una cola izquierda larga. Como regla aproximada, un valor de la asimetría mayor que el doble de su error típico se asume que indica una desviación de la simetría.

**Error típico de la asimetría.** La razón de la asimetría sobre su error típico puede utilizarse como contaste de la normalidad (es decir, se puede rechazar la normalidad si la razón es menor que -2 o mayor que +2). Un valor grande y positivo para la asimetría indica una cola larga a la derecha; un valor extremo y negativo indica una cola larga por la izquierda

**Curtosis.** Medida del grado en que las observaciones están agrupadas en torno al punto central. Para una distribución normal, el valor del estadístico de curtosis es 0. Una curtosis positiva indica que, con respecto a una distribución normal, las observaciones se concentran más en el centro de la distribución y presentan colas más estrechas hasta los valores extremos de la distribución, en cuyo punto las colas de la distribución leptocúrtica son más gruesas con respecto a una distribución normal. Una curtosis negativa indica que, con respecto a una distribución normal, las observaciones se concentran menos y presentan colas más gruesas hasta los valores extremos de la distribución, en cuyo punto las colas de la distribución platicúrtica son más estrechas con respecto a una distribución normal.

**Error típico de la curtosis.** La razón de la curtosis sobre su error típico puede utilizarse como contaste de la normalidad (es decir, se puede rechazar la normalidad si la razón es menor que -2 o mayor que +2). Un valor grande y positivo para la curtosis indica que las colas son más largas que las de una distribución normal; por el contrario, un valor extremo y negativo indica que las colas son más cortas (llegando a tener forma de caja como en la distribución uniforme).

**Único.** Evalúa todos los efectos simultáneamente, corrigiendo cada efecto por todos los demás efectos de cualquier tipo.

**Válido.** Los casos válidos que no tienen un valor perdido por el sistema ni un valor definido como perdido por el usuario.

**Mediana.** Es el valor por encima y por debajo del cual se encuentran la mitad de los casos, el percentil 50. Si hay un número par de casos, la mediana es la media de los dos valores centrales, cuando los casos se ordenan en orden ascendente o descendente. La mediana es una medida de tendencia central que no es sensible a los valores atípicos (a diferencia de la media, que puede resultar afectada por unos pocos valores extremadamente altos o bajos).

**Modo.** El valor que ocurre con mayor frecuencia. Si varios valores comparten la mayor frecuencia de aparición, cada uno de ellos es un modo.

Recuerde que la mediana y la moda se eliminan por defecto para mejorar el rendimiento, pero se pueden seleccionar en la pestaña Configuración del nodo Auditar datos. Si desea obtener más información, consulte el tema [Pestaña Configuración del nodo Auditar datos](#) el p. 414.

### **Estadísticos para superposiciones**

Si se está utilizando un campo de superposición continuo (rango numérico), también están disponibles los siguientes estadísticos:

**Covarianza.** Medida no tipificada del grado de asociación entre dos variables igual al producto cruzado de las diferenciales, dividido por N-1.

### **Pestaña Calidad del explorador de auditoría de datos**

Figura 6-23  
Informe de calidad en el explorador de auditoría de datos

Campo	Medida	Valores atípicos	Extremos	Acción	Imputar perdidos	Método	% Completc
region	Nominal	--	--		Nunca	Fijo	
tenure	Continua	0	0 Ninguna		Nunca	Fijo	
age	Continua	0	0 Ninguna		Nunca	Fijo	
marital	Marca	--	--		Nunca	Fijo	
address	Continua	12	0 Ninguna		Nunca	Fijo	
income	Continua	9	6 Ninguna		Nunca	Fijo	
ed	Ordinal	--	--		Nunca	Fijo	
employ	Continua	8	0 Ninguna		Nunca	Fijo	
retire	Nominal	--	--		Nunca	Fijo	
gender	Nominal	--	--		Nunca	Fijo	
reside	Ordinal	--	--		Nunca	Fijo	
tollfree	Marca	--	--		Nunca	Fijo	
equip	Marca	--	--		Nunca	Fijo	
calcard	Marca	--	--		Nunca	Fijo	
wireless	Marca	--	--		Nunca	Fijo	
longmon	Continua	18	4 Ninguna		Nunca	Fijo	
tollmon	Continua	9	1 Ninguna		Nunca	Fijo	
equipmon	Continua	2	0 Ninguna		Nunca	Fijo	
cardmon	Continua	11	3 Ninguna		Nunca	Fijo	

La pestaña Calidad del explorador de auditoría de datos muestra los resultados del análisis de calidad de los datos y permite especificar los tratamientos para los valores atípicos, extremos y perdidos.

### **Introducción de valores perdidos**

El informe de auditoría muestra el porcentaje de registros completos para cada campo, junto con el número de valores válidos, nulos y vacíos. Puede seleccionar imputar valores perdidos para campos específicos, como considere oportuno y, a continuación, generar un Supernodo para aplicar estas transformaciones.

- ▶ En la columna Imputar perdidos, especifique el tipo de valores que desea imputar, si los hay. Puede elegir entre imputar valores vacíos, nulos o ambos, o especificar una condición o expresión personalizada que seleccione los valores que se van a imputar.

Hay varios tipos de valores perdidos reconocidos por IBM® SPSS® Modeler:

- **Valores nulos o perdidos por el sistema.** Son valores que no son de cadena y que se han dejado en blanco en la base de datos o el archivo fuente y que no se han definido específicamente como “perdidos” en un nodo Tipo o de origen. Los valores perdidos del sistema se muestran como \$null\$. Tenga en cuenta que las cadenas vacías no se consideran nulas en SPSS Modeler, aunque se pueden tratar como tales en algunas bases de datos.
  - **Cadenas vacías y espacios en blanco.** Los valores de cadenas vacías y los espacios en blanco (cadenas con caracteres no visibles) se tratan como diferentes de los valores nulos. Las cadenas vacías se tratan como equivalentes al espacio en blanco en la mayoría de los casos. Por ejemplo, si selecciona la opción de tratar los espacios en blanco como vacíos en un nodo Tipo o de origen, esta configuración se aplica a las cadenas vacías también.
  - **Valores vacíos o perdidos definidos por el usuario.** Estos valores, como desconocidos, -99, o -1, que se definen explícitamente como perdidos en el nodo Tipo o de origen. Si lo prefiere, también puede elegir tratar los valores nulos o espacios en blanco como vacíos, lo que permite que se les marque para un tratamiento especial y que se excluyan de la mayoría de los cálculos. Por ejemplo, puede utilizar la función @BLANK para tratar estos valores junto con otros tipos de valores perdidos, como vacíos. Si desea obtener más información, consulte el tema [Utilización del cuadro de diálogo de valores](#) en el capítulo 4 el p. 143.
- ▶ En la columna Método, especifique el método que desea utilizar.

Los siguientes métodos están disponibles para introducir valores perdidos:

**Fijas.** Sustituye un valor fijo (ya sea la media de campo, el punto medio del rango o un constante que especifique).

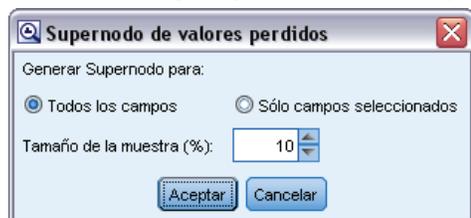
**Aleatorios.** Sustituye un valor aleatorio basado en una distribución uniforme o normal.

**Expresión.** Permite especificar una expresión personalizada. Por ejemplo, podría sustituir los valores con una variable global creada por el nodo Val. globales.

**Algoritmo.** Sustituye un valor pronosticado por un modelo basado en el algoritmo C&RT. En cada campo imputado con este método, habrá un modelo C&RT independiente, junto con un nodo Rellenar que sustituye valores vacíos y nulos con el valor que predice el modelo. A continuación, se utiliza un nodo Filtro para eliminar los campos de pronóstico generados por el modelo.

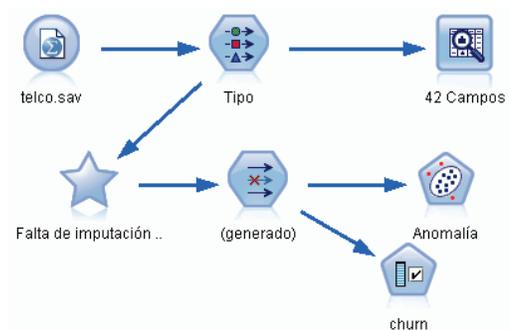
- ▶ Para generar un Supernodo de valores perdidos, elija en los menús:  
Generar > Supernodo de valores perdidos

Figura 6-24  
Cuadro de diálogo Supernodo de valores perdidos



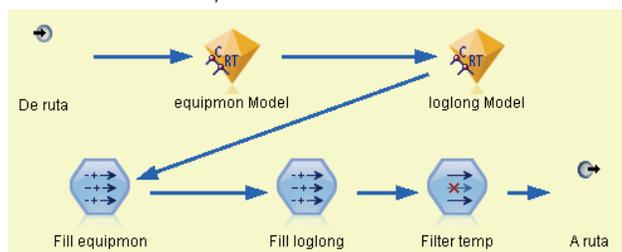
- ▶ Seleccione Todos los campos o Sólo campos seleccionados, y especifique un tamaño de muestra si lo desea. (La muestra especificada es un porcentaje, por defecto se muestrea el 10 % de todos los registros).
- ▶ Pulse en Aceptar para añadir el Supernodo generado al lienzo de rutas.
- ▶ Conecte el Supernodo a la ruta para aplicar las transformaciones.

Figura 6-25  
Adición del Supernodo a la ruta



Dentro del Supernodo, se utiliza una combinación de nugget de modelo, Rellenar y Filtro, según sea apropiado. Para comprender cómo funciona, puede editar el Supernodo y pulsar en Acercar, tras lo que podrá añadir, modificar o eliminar nodos específicos dentro del Supernodo para ajustar el comportamiento.

Figura 6-26  
Acercamiento al Supernodo



### **Tratamiento de los valores atípicos y extremos**

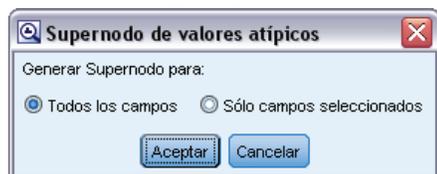
El informe de auditoría muestra el número de valores atípicos y extremos para cada campo según las opciones de detección especificadas en el nodo Auditar datos. Si desea obtener más información, consulte el tema [Pestaña Calidad de Auditar datos](#) el p. 416. Puede seleccionar forzar, descartar o anular estos valores para campos específicos, según considere oportuno y, a continuación, generar un Supernodo para aplicar las transformaciones.

- ▶ En la columna Acción, especifique cómo desea tratar los valores atípicos y extremos para campos específicos.

Hay disponibles las siguientes acciones para el tratamiento de valores atípicos y extremos:

- **Forzar.** Sustituye los valores atípicos y extremos por el valor más cercano que no se consideraría extremo. Por ejemplo, si un valor atípico se define como todo lo que queda por encima o por debajo de tres desviaciones típicas, todos los valores atípicos se sustituirían por el valor mayor o menor de este intervalo.
  - **Descartar.** Descarta los registros con valores atípicos o extremos para el campo especificado.
  - **Anular.** Sustituye los valores atípicos y extremos con valores nulos o perdidos por el sistema.
  - **Forzar valores atípicos/descartar extremos.** Descarta sólo los valores extremos.
  - **Forzar valores atípicos/anular extremos.** Anula sólo los valores extremos.
- ▶ Para generar el Supernodo, seleccione en los menús:  
Generar > Supernodo de valores atípicos y extremos

Figura 6-27  
Cuadro de diálogo Supernodo de valores atípicos



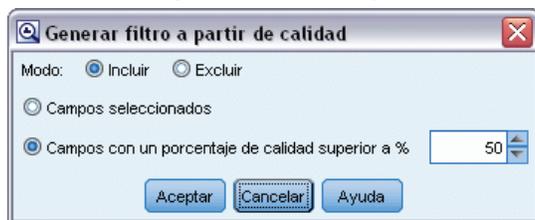
- ▶ Seleccione Todos los campos o Sólo campos seleccionados y, a continuación, pulse en Aceptar para añadir el Supernodo generado al lienzo de rutas.
- ▶ Conecte el Supernodo a la ruta para aplicar las transformaciones.

Si lo desea, puede editar el Supernodo y utilizar la opción de acercar para examinarlo o realizar cambios. Dentro del Supernodo, los valores se descartan, fuerzan o anulan mediante varios nodos Seleccionar o Rellenar, según proceda.

### **Filtrado de campos con datos perdidos**

Desde el explorador de auditoría de datos, puede crear un nuevo nodo Filtro según los resultados del análisis de calidad.

Figura 6-28  
Cuadro de diálogo Generar filtro a partir de calidad



**Moda.** Seleccione la operación que desea para los campos especificados, ya sea Incluir o Excluir.

- **Campos seleccionados.** El nodo Filtro incluirá o excluirá los campos seleccionados en la pestaña Calidad. Por ejemplo, puede ordenar la tabla por la columna % Completo. Para ello, mantenga pulsada la tecla Mayús mientras pulsa en los campos menos completos y, a continuación, genere un nodo Filtro que excluya dichos campos.
- **Campos con un porcentaje de calidad superior a %.** El nodo Filtro incluirá/excluirá campos en los que el porcentaje de registros completos sea mayor que el umbral especificado. El umbral por defecto es 50%.

#### **Filtrado de campos vacíos o sin tipo**

Recuerde que una vez instanciados los valores de datos, los campos vacíos o sin tipo se excluyen de los resultados de auditoría y de la mayoría de los otros resultados de IBM® SPSS® Modeler. Estos campos se ignoran con fines de modelado, pero pueden recargar o hinchar los datos. Si es así, puede utilizar el explorador de auditoría de datos para generar un nodo Filtro que elimine estos campos de la ruta.

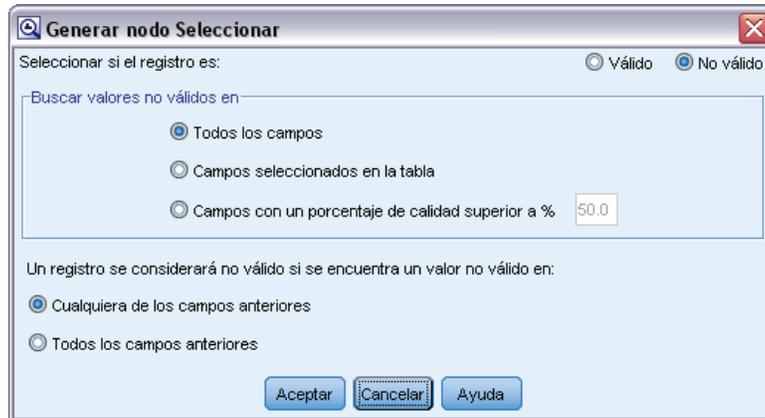
- ▶ Para asegurarse de que todos los campos se incluyen en la auditoría, incluidos los campos vacíos o sin tipo, pulse en Borrar todos los valores en el nodo Tipo o de origen anterior de la ruta, o establezca Valores en <Pasar> para todos los campos.
- ▶ En el explorador de auditoría de datos, ordene por la columna % Completo, seleccione los campos que tienen cero valores válidos (u otro umbral) y utilice el menú Generar para generar un nodo Filtro que se pueda añadir a la ruta.

#### **Selección de registros con datos perdidos**

Desde el explorador de auditoría de datos, puede crear un nuevo nodo Seleccionar según los resultados del análisis de calidad.

- ▶ En el navegador Auditar datos, seleccione la pestaña Calidad.
- ▶ Elija en el menú:  
Generar > Nodo Selección de valores perdidos

Figura 6-29  
Cuadro de diálogo Generar nodo Seleccionar



**Seleccionar si el registro es.** Especifique si el registro debe mantenerse cuando es Válido o No válido.

**Buscar valores no válidos en.** Especifique dónde deben buscarse los valores no válidos.

- **Todos los campos.** El nodo Seleccionar buscará valores no válidos en todos los campos.
- **Campos seleccionados en la tabla.** El nodo Seleccionar comprobará sólo los campos seleccionados actualmente en la tabla de resultados de calidad.
- **Campos con un porcentaje de calidad superior a %.** El nodo Seleccionar comprobará los campos en los que el porcentaje de registros completos es mayor que el umbral especificado. El umbral por defecto es 50%.

**Un registro se considerará no válido si se encuentra un valor no válido en.** Especifique la condición para identificar un registro no válido.

- **Cualquiera de los campos anteriores.** El nodo Seleccionar considerará un registro como no válido si *cualquiera* de los campos especificados anteriormente contiene un valor no válido para ese registro.
- **Todos los campos anteriores.** El nodo Seleccionar considerará un registro como no válido sólo si *todos* los campos especificados anteriormente contienen un valor no válido para ese registro.

### **Generación de otros nodos para la preparación de datos**

Se pueden generar varios nodos utilizados en la preparación de datos directamente desde el explorador de auditoría de datos, incluidos los nodos Intervalos, Reclasificar y Derivar. Por ejemplo:

- Puede derivar un nuevo campo basado en los valores de *claimvalue* y *farmincome* seleccionando ambos en el informe de auditoría y seleccionando Derivar en el menú Generar. El nuevo nodo se añade al lienzo de rutas.
- De forma similar, puede determinar, según los resultados de auditoría, si recodificar *farmincome* en intervalos de percentiles ofrece un análisis más centrado. Para generar un nodo Intervalos, seleccione la fila de campo de la representación y seleccione Intervalos del menú Generar.

Una vez generado el nodo y añadido al lienzo de rutas, debe adjuntarlo a la ruta y abrirlo para especificar opciones para los campos seleccionados.

## Nodo Transformación

La normalización de los campos de entrada es un importante paso previo a la utilización de las técnicas de puntuación tradicionales, como la regresión, la regresión logística y el análisis discriminante. Estas técnicas implican ciertos supuestos sobre las distribuciones normales de datos que pueden no ser ciertas para muchos archivos de datos sin procesar. Un método para trabajar con datos reales es aplicar transformaciones para convertir un elemento de los datos brutos hacia una distribución más normal. Además, los campos normalizados se pueden comparar fácilmente entre sí: por ejemplo, los ingresos y la edad son escalas completamente diferentes en un archivo de datos sin procesar pero, al normalizarse, se puede interpretar fácilmente el impacto relativo sobre cada uno.

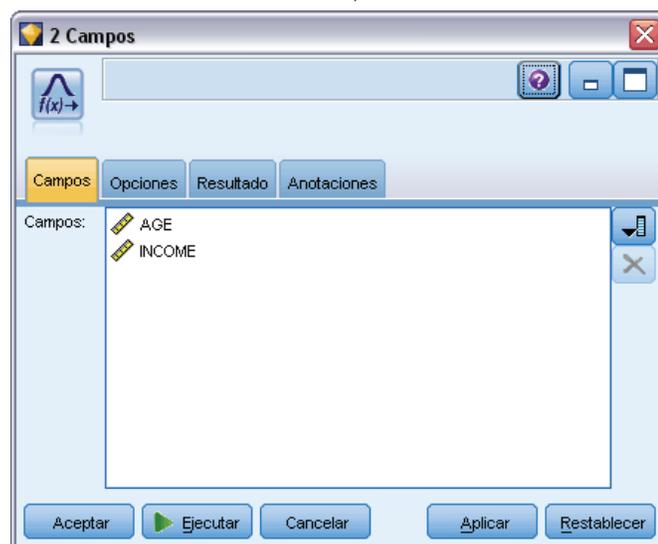
El nodo Transformación proporciona un visor de resultados que permite realizar una rápida valoración visual de la mejor transformación que se puede utilizar. Con sólo un vistazo, puede ver si las variables están distribuidas normalmente y, si es necesario, seleccionar la transformación necesaria y aplicarla. Puede elegir varios campos y realizar una transformación en cada uno.

Tras seleccionar las transformaciones preferidas para los campos, puede generar nodos Derivar o Rellenar para realizar las transformaciones y conectar estos nodos a la ruta. El nodo Derivar crea nuevos campos, mientras que el nodo Rellenar transforma los campos ya existentes. Si desea obtener más información, consulte el tema [Generación de gráficos](#) el p. 431.

### Pestaña Campos del nodo Transformar

En la pestaña Campos, puede especificar los campos de los datos que desea utilizar para ver las posibles transformaciones y aplicarlas. Sólo es posible transformar los campos numéricos. Pulse en el botón selector de campos y seleccione uno o varios campos numéricos de la lista.

Figura 6-30  
Nodo Transformar: Pestaña campos



## Pestaña Opciones del nodo Transformar

La pestaña Opciones le permite especificar el tipo de transformaciones que desea incluir. Puede elegir entre incluir todas las transformaciones disponibles o seleccionarlas individualmente.

En este último caso, también puede introducir un número con el fin de desplazar los datos para las transformaciones inversas y logarítmicas. Esto resulta útil en situaciones en las que una gran proporción de ceros en los datos sesgaría los resultados de la media y la desviación típica.

Por ejemplo, supongamos que tiene un campo llamado *SALDO* que tiene algunos valores cero y desea utilizar la transformación inversa sobre él. Para evitar sesgos no deseados, tendría que seleccionar Inversa ( $1/x$ ) y escribir un 1 en el campo Usar desplazamiento de datos. (Recuerde que este desplazamiento no está relacionado con el de la función de secuencia @OFFSET de IBM® SPSS® Modeler.)

Figura 6-31  
Nodo Transformar: Pestaña Opciones



**Todas las fórmulas.** Indica que todas las transformaciones disponibles deben calcularse y mostrarse en el resultado.

**Fórmulas seleccionadas.** Permite seleccionar las diferentes transformaciones que se van a calcular y mostrar en el resultado.

- **Inversa ( $1/x$ ).** Indica que la transformación inversa debe mostrarse en el resultado.
- **Log ( $\log n$ ).** Indica que la transformación del  $\log_n$  debe mostrarse en el resultado.
- **Log ( $\log 10$ ).** Indica que la transformación del  $\log_{10}$  debe mostrarse en el resultado.
- **Exponencial.** Indica que la transformación exponencial ( $e^x$  debe mostrarse en el resultado.
- **Raíz cuadrada.** Indica que la transformación de raíz cuadrada debe mostrarse en el resultado.

## Pestaña Resultado del nodo Transformar

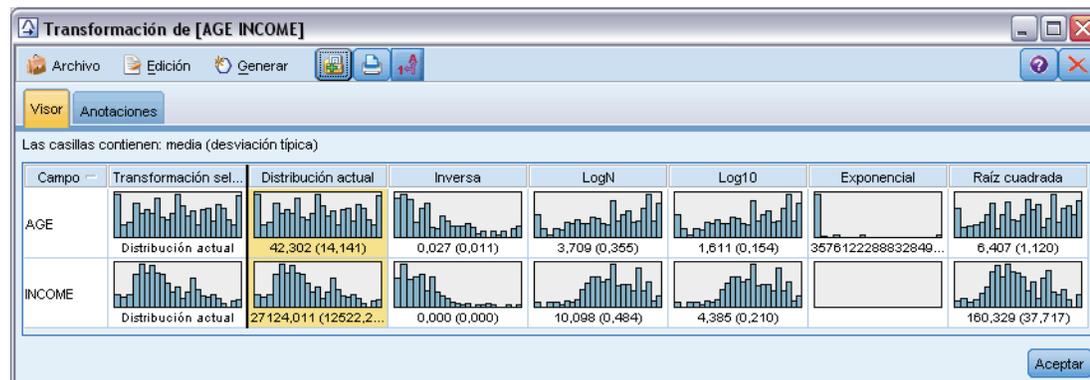
La pestaña Resultado permite especificar el formato y la ubicación del resultado. Puede elegir entre mostrar los resultados en pantalla o enviarlos a uno de los tipos de archivo estándar. Si desea obtener más información, consulte el tema [Pestaña Resultado del nodo de resultados](#) el p. 399.

## Visor de resultados del nodo Transformar

El visor de resultados permite ver los resultados obtenidos tras ejecutar el nodo Transformar. El visor es una potente herramienta que muestra múltiples transformaciones por campo en vistas en miniatura de la transformación, lo que le permite comparar rápidamente los campos. Puede utilizar las opciones del menú Archivo para guardar, exportar o imprimir el resultado. Si desea obtener más información, consulte el tema [Visualización de resultados](#) el p. 393.

Figura 6-32

Visualización de transformaciones disponibles por campo



Para cada transformación (distinta a la seleccionada), se muestra una leyenda debajo con el formato:

Mean (Standard deviation)

## Generación de nodos para las transformaciones

El visor de resultados ofrece un punto de partida útil para la preparación de datos. Por ejemplo, puede desear normalizar el campo *EDAD* con el fin de utilizar una técnica de puntuación (como la regresión logística o el análisis discriminante) que suponga una distribución normal. Según los gráficos iniciales y a los estadísticos de resumen, es posible que decida transformar el campo *EDAD* de acuerdo con una distribución determinada (por ejemplo, log). Tras seleccionar la distribución deseada, puede generar un nodo Derivar con una transformación estandarizada para utilizarla para puntuar.

Desde el visor de resultados puede generar los siguientes nodos de operaciones con campos:

- Derivar
- Rellenar

El nodo Derivar crea nuevos campos con las transformaciones deseadas, mientras que el nodo Rellenar transforma los campos existentes. Los nodos se colocan en el lienzo en forma de un Supernodo.

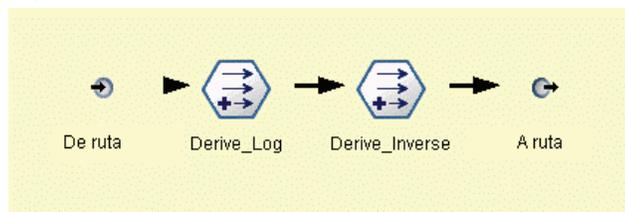
Si selecciona la misma transformación para campos diferentes, un nodo Derivar o Rellenar contiene las fórmulas de dicho tipo de transformación para todos los campos a los que se aplica la transformación. Por ejemplo, supongamos que ha seleccionado los siguientes campos y transformaciones para generar un nodo Derivar:

Campo	Transformación
EDAD	Distribución actual
INGRESOS	Log
SALDO_PENDIENTE	Inverso
SALDO	Inverso

El Supernodo contiene los siguientes nodos:

Figura 6-33

Supernodo en el lienzo



En este ejemplo, el nodo Derive\_Log tiene la fórmula de log para el campo *INGRESOS* y el nodo Derive\_Inverse tiene las fórmulas inversas para los campos *SALDO\_PENDIENTE* y *SALDO*.

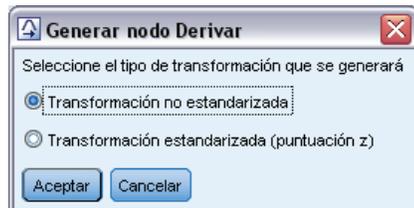
### Generación de un nodo

- ▶ Para cada campo del visor de resultados, seleccione la transformación deseada.
- ▶ En el menú Generar, seleccione Nodo Derivar o Nodo Rellenar como desee.

Al hacerlo, se muestra el cuadro de diálogo Generar nodo Derivar o Generar nodo Rellenar, según corresponda.

Figura 6-34

Selección de transformación estandarizada o no estandarizada



Seleccione Transformación no estandarizada o Transformación estandarizada (puntuación z), según sus preferencias. La segunda opción aplica una puntuación z a la transformación; las puntuaciones z representan valores como la distancia desde la media de la variable en desviaciones típicas. Por

ejemplo, si aplica la transformación logarítmica al campo *EDAD* y selecciona una transformación estandarizada, la ecuación final para el nodo generado será:

$$(\log(\text{AGE})-\text{Mean})/\text{SD}$$

Una vez generado el nodo y representado en el lienzo de ruta:

- ▶ Conéctelo con la ruta.
- ▶ Para un Supernodo, si lo desea pulse dos veces en el nodo para ver su contenido.
- ▶ Si lo prefiere, pulse dos veces en un nodo Derivar o Rellenar para modificar las opciones del campo seleccionado.

### **Generación de gráficos**

Puede generar resultados como histogramas a tamaño completo a partir de los histogramas en miniatura del visor de resultados.

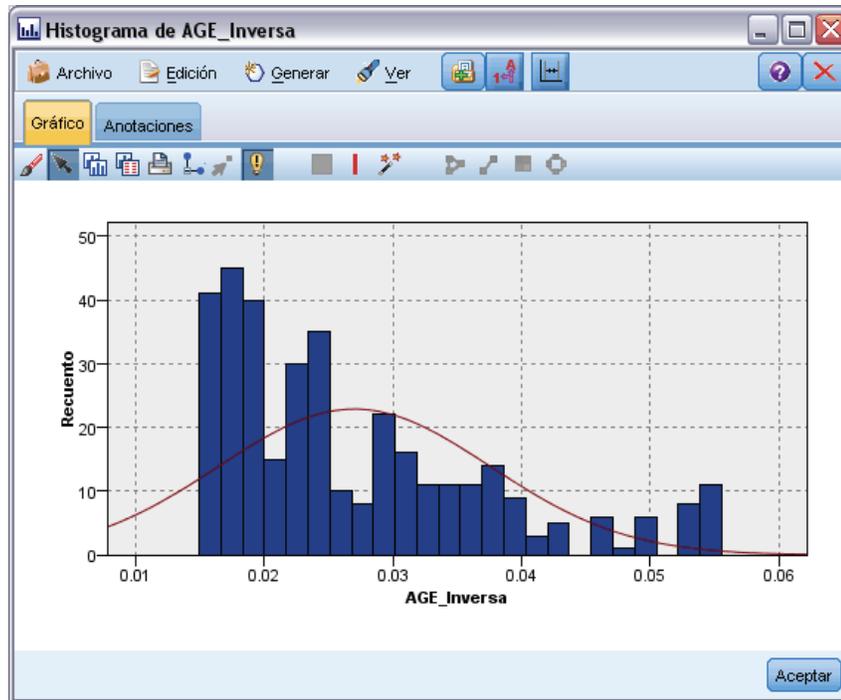
#### **Generación de un gráfico**

- ▶ Pulse dos veces en un gráfico en miniatura del visor de resultados.  
*o*
- ▶ Seleccione un gráfico en miniatura del visor de resultados.
- ▶ En el menú Generar, elija Resultado gráfico.

Al hacerlo, se muestra el histograma con una curva de distribución normal superpuesta, lo que le permite comparar en qué medida se ajusta cada transformación disponible a una distribución normal.

*Nota:* Sólo puede generar un gráfico si el nodo Transformar que creó el resultado está conectado a la ruta.

Figura 6-35  
Histograma de transformación con curva de distribución normal superpuesta



### Otras operaciones

Desde el visor de resultados, también puede:

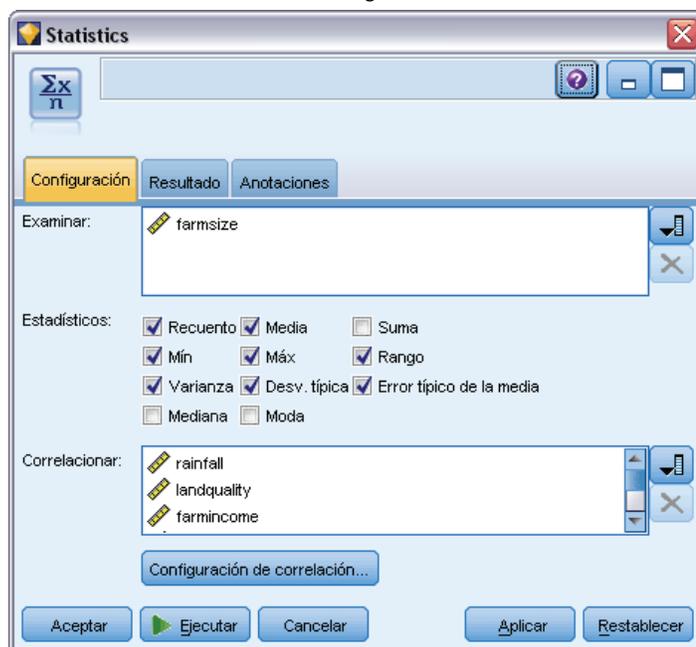
- Ordenar la cuadrícula de salida por la columna Campo.
- Exportar los resultados a un archivo HTML. Si desea obtener más información, consulte el tema [Exportación de resultados](#) el p. 396.

## Nodo Estadísticos

El nodo Estadísticos ofrece información de resumen básica acerca de los campos numéricos. Puede obtener estadísticos de resumen para campos individuales y correlaciones entre campos.

## Pestaña Configuración del nodo Estadísticos

Figura 6-36  
Nodo Estadísticos: Pestaña Configuración



**Examinar.** Seleccione el campo o los campos para los que desee estadísticos de resumen individuales. Puede seleccionar varios campos.

**Estadísticos.** Seleccione los estadísticos para el informe. Las opciones disponibles incluyen Recuento, Media, Suma, Mín, Máx, Rango, Varianza, Desv. típica, Error típico de la media, Mediana y Moda.

**Correlacionar.** Seleccione el campo o los campos que desee correlacionar. Puede seleccionar varios campos. Cuando los campos de correlación están seleccionados, la correlación entre cada campo Examinar y los campos de correlación se mostrará en el resultado.

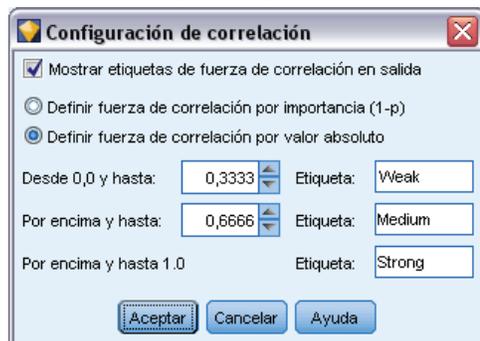
**Configuración de correlación.** Puede especificar las opciones de visualización de la fuerza de correlación de los resultados.

### Configuración de correlación.

IBM® SPSS® Modeler puede caracterizar correlaciones con etiquetas descriptivas para ayudarle a resaltar relaciones importantes. La **correlación** mide la fuerza de la relación entre dos campos continuos (rango numérico). Toma valores entre  $-1,0$  y  $1,0$ . Los valores cercanos a  $+1,0$  indican una fuerte asociación positiva, de forma que los valores altos de un campo están asociados con los valores altos del otro y los valores bajos están asociados con los valores bajos. Los valores cercanos a  $-1,0$  indican una asociación negativa fuerte, de forma que los valores altos para un campo están asociados con los valores bajos para el otro y viceversa. Los valores cercanos a  $0,0$  indican una asociación débil, de forma que los valores para los dos campos son más o menos independientes.

Puede controlar la visualización de las etiquetas de correlación, cambiar los umbrales que definen las categorías y cambiar las etiquetas utilizadas para cada rango. Ya que la forma de caracterizar los valores de correlación depende en gran parte del dominio del problema, puede que desee personalizar los rangos y las etiquetas para que se ajusten a su situación en particular.

Figura 6-37  
Cuadro de diálogo Configuración de correlación



**Mostrar etiquetas de fuerza de correlación en salida.** Esta opción está seleccionada por defecto. Anule la selección para anular las etiquetas descriptivas del resultado.

**Fuerza de correlación.** Hay dos opciones para definir y etiquetar la fuerza de las correlaciones:

- **Definir fuerza de correlación por importancia (1-p).** Etiqueta las correlaciones según la importancia, que se define como 1 menos la significación o 1 menos la probabilidad de que la diferencia de las medias se pueda explicar por el azar solo. Cuanto más cerca está dicho valor a 1, mayor es la probabilidad de que los dos campos *no* sean independientes; es decir, que exista alguna relación entre ellos. El etiquetado de correlaciones por importancia se recomienda generalmente en vez del valor absoluto ya que tiene en cuenta la variabilidad de los datos, por ejemplo, un coeficiente de 0,6 puede ser altamente significativo en un conjunto de datos y nada significativo en otro. Por defecto, los valores de importancia 0,0 y 0,9 se etiquetan como *Débil*, los valores entre 0,9 y 0,95 se etiquetan como *Media* y entre 0,95 y 1,0 se etiquetan como *Fuerte*.
- **Definir fuerza de correlación por valor absoluto.** Etiqueta las correlaciones según el valor absoluto del coeficiente de correlación de Pearson, que varía entre  $-1$  y  $1$ , como se ha descrito anteriormente. Cuanto más cerca de 1 está el valor de esta medida, más fuerte es la correlación. Por defecto, las correlaciones entre 0,0 y 0,3333 (en valores absolutos) se etiquetan como *Débil*, entre 0,3333 y 0,6666 se etiquetan como *Media* y entre 0,6666 y 1,0 se etiquetan como *Fuerte*. Tenga en cuenta, sin embargo, que la significación de cualquier valor es difícil de generalizar de un conjunto de datos a otro. Por ello, se recomienda en la mayoría de los casos definir correlaciones según la probabilidad en lugar de por el valor absoluto.

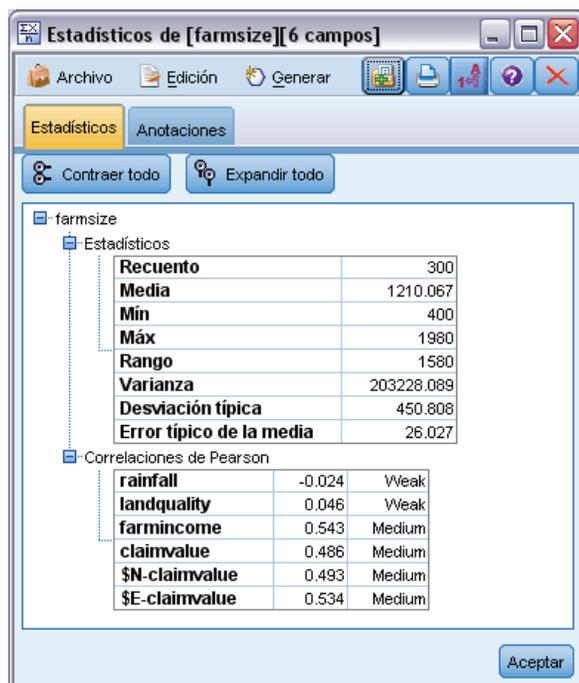
## Explorador de resultados de estadísticos

El explorador de resultados del nodo Estadísticos muestra los resultados de análisis estadísticos y permite realizar operaciones, incluida la selección de campos, la generación de nuevos nodos basados en la selección y el almacenamiento y la impresión de resultados. Las opciones normales de almacenamiento, exportación e impresión están disponibles en el menú Archivo y las opciones

de edición normales están disponibles en el menú Edición. Si desea obtener más información, consulte el tema [Visualización de resultados](#) el p. 393.

Cuando se examina el resultado de estadísticos por primera vez, los resultados se expanden. Para ocultar los resultados cuando haya terminado de consultarlos, utilice el control de expansión situado a la izquierda del elemento para contraer los resultados específicos que desee ocultar o pulse en el botón Contraer todo para contraer todos los resultados. Para volver a ver los resultados tras contraerlos, utilice el control de expansión situado a la izquierda del elemento para mostrar los resultados o pulse en el botón Expandir todo para mostrar todos los resultados.

Figura 6-38  
Explorador de resultados de estadísticos



El resultado contiene una sección para cada campo *Examinar* que contiene una tabla de los estadísticos solicitados.

- **Recuento.** Número de registros con valores válidos para el campo.
- **Media.** Valor medio (promedio) para el campo en todos los registros.
- **Suma.** Suma de valores para el campo en todos los registros.
- **Mín.** Valor mínimo para el campo.
- **Máx.** Valor máximo para el campo.
- **Rango.** Diferencia entre los valores mínimos y máximos.
- **Varianza.** Medida de la variabilidad en los valores de un campo. Se calcula tomando la diferencia entra cada valor y la media global, calculando el cuadrado, sumando todos los valores y dividiendo por el número de registros.
- **Desviación típica.** Otra medida de variabilidad en los valores de un campo, calculada como la raíz cuadrada de la varianza.

- **Error típico de la media.** Medida de la incertidumbre en la estimación de la media del campo si se asume que la media se aplica a los datos nuevos.
- **Mediana.** El valor del “centro” para el campo, es decir, el valor que separa la mitad superior de los datos de la mitad inferior de los datos (según los valores del campo).
- **Moda.** El valor único más común de los datos.

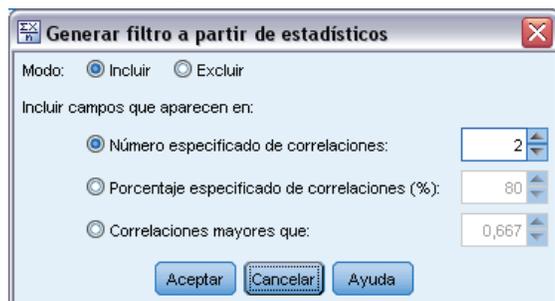
**Correlaciones.** Si ha especificado campos correlacionados, el resultado también contiene una sección que muestra la correlación de Pearson entre el campo Examinar y cada campo correlacionado, así como etiquetas descriptivas opcionales que describen los valores de correlación. Si desea obtener más información, consulte el tema [Configuración de correlación](#), el p. 433.

**Menú Generar.** El menú Generar contiene operaciones de generación de nodos.

- **Filtro.** Genera un nodo Filtro que filtra campos que no están correlacionados o que están débilmente correlacionados con otros campos.

### **Generación de un nodo Filtro a partir de Estadísticos**

Figura 6-39  
Cuadro de diálogo Generar filtro a partir de estadísticos



El nodo Filtro generado a partir del explorador de resultados de estadísticos filtrará los campos según sus correlaciones con otros campos. Funciona ordenando las correlaciones según el valor absoluto, tomando las correlaciones mayores (según el criterio establecido en el cuadro de diálogo) y creando un filtro que pasa todos los campos que aparecen en cualquiera de esas correlaciones de gran tamaño.

**Moda.** Decida cómo seleccionar las correlaciones. Incluir hace que los campos que aparecen en las correlaciones especificadas se mantengan. Excluir hace que se filtren los campos.

**Incluir/excluir campos que aparecen en.** Definir el criterio para seleccionar correlaciones.

- **Número especificado de correlaciones.** Selecciona el número especificado de correlaciones e incluye/excluye los campos que aparecen en cualquiera de esas correlaciones.
- **Porcentaje especificado de correlaciones (%).** Selecciona el porcentaje especificado ( $n\%$ ) de correlaciones e incluye/excluye los campos que aparecen en cualquiera de esas correlaciones.
- **Correlaciones mayores que.** Selecciona correlaciones con un valor absoluto mayor que el umbral especificado.

## **Nodo Medias**

El nodo Medias compara las medias de grupos independientes o de pares de campos relacionados para probar si existen diferencias significativas. Por ejemplo, puede comparar los ingresos medios antes y después de llevar a cabo una promoción o comparar los ingresos de los clientes que no recibieron esa promoción con los que sí lo hicieron.

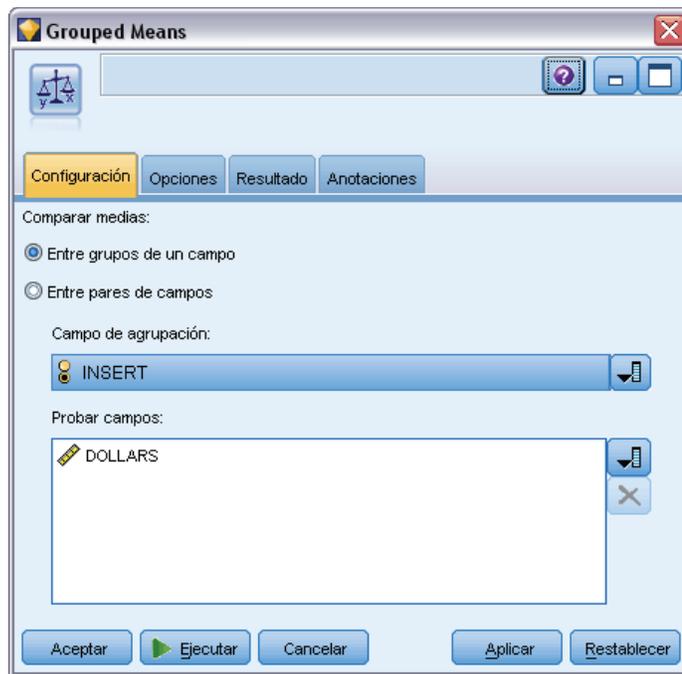
Puede comparar medias de dos maneras distintas, dependiendo de los datos:

- **Entre grupos dentro de un campo.** Para comparar grupos independientes, seleccione un campo de prueba y un campo de agrupación. Por ejemplo, se puede excluir una muestra de clientes “reservados” al enviar una promoción y comparar los ingresos medios de ese grupo con los del resto. En este caso, podría especificar un campo de prueba único que indique el ingreso de cada cliente, con un campo nominal o marca que indique si recibieron la oferta. Las muestras son independientes en el sentido de que cada registro se asigna a un grupo o a otro, y no hay manera de vincular un miembro específico de un grupo con un miembro específico de otro. También puede especificar un campo nominal con más de dos valores para comparar las medias de varios grupos. Cuando se ejecuta, el nodo calcula una prueba ANOVA unidireccional de los campos seleccionados. En los casos en que sólo hay dos grupos de campos, los resultados de ANOVA unidireccional son básicamente los mismos que los de una prueba  $t$  de muestras independientes. Si desea obtener más información, consulte el tema [Comparación de medias para grupos independientes](#) el p. 437.
- **Entre pares de campos.** Al comparar las medias de dos campos relacionados, los grupos deben estar relacionados de alguna manera para que los resultados tengan sentido. Por ejemplo, podría comparar los ingresos medios de un mismo grupo de clientes antes y después de llevar a cabo una promoción o comparar las tasas de uso de un servicio entre pares marido-mujer para ver si son diferentes. Cada registro contiene dos medidas diferentes pero relacionadas que se pueden comparar de manera lógica. Cuando se ejecuta, el nodo calcula una prueba  $t$  de muestras relacionadas en cada par de campos seleccionado. Si desea obtener más información, consulte el tema [Comparación de medias entre campos relacionados](#) el p. 438.

### **Comparación de medias para grupos independientes**

Seleccione Entre grupos de un campo en el nodo Medias para comparar la media de dos o más grupos independientes.

Figura 6-40  
Comparación de medias entre grupos dentro de un campo



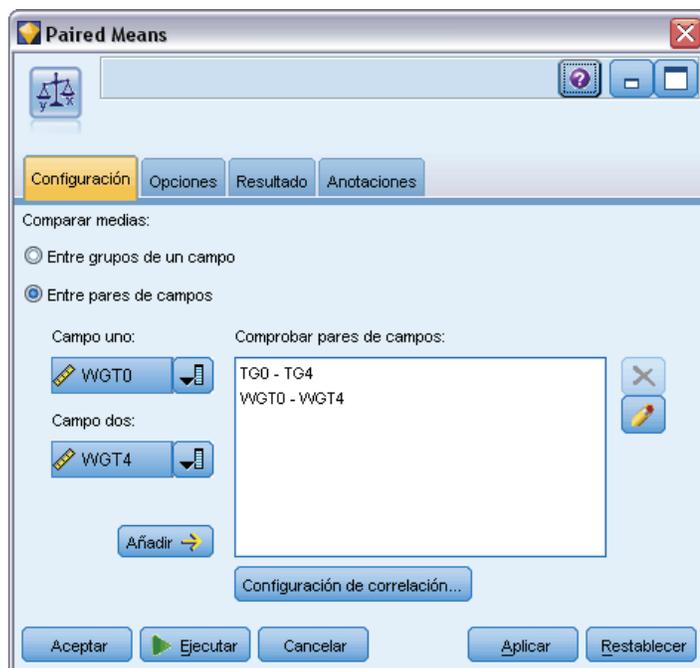
**Campo de agrupación.** Seleccione una marca numérica o un campo nominal con dos o más valores distintos que divida los registros en los grupos que se quieren comparar, por ejemplo, los que recibieron la oferta frente a los que no la recibieron. Independientemente del número de campos de prueba, sólo se puede seleccionar un campo de agrupamiento.

**Probar campos.** Seleccione uno o más campos numéricos que contengan las medidas que desea analizar. Se realizará una prueba diferente para cada campo seleccionado. Por ejemplo, podría analizar el impacto de una promoción determinada sobre el uso, ingresos y pérdida de clientes.

### ***Comparación de medias entre campos relacionados***

Seleccionar Entre pares de campos en el nodo Medias para comparar medias entre campos distintos. Los campos deben estar relacionados de alguna manera para que los resultados tengan sentido, como los ingresos antes y después de una promoción. También se pueden seleccionar varios pares de campos.

Figura 6-41  
Comparación de medias entre campos relacionados



**Campo uno.** Seleccione un campo numérico que contenga la primera medida que se desea comparar. En un estudio de antes y después, este sería el campo “antes”.

**Campo dos.** Seleccione el segundo campo que desea comparar.

**Añadir.** Añade el par seleccionado a la lista Comprobar pares de campos.

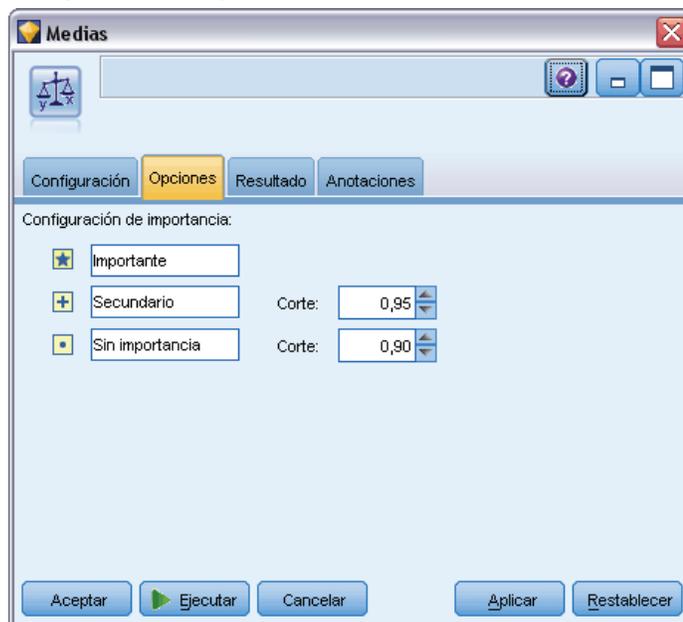
Repita la selección de campos según sea necesario para añadir varios pares a la lista.

**Configuración de correlación.** Permite especificar las opciones para etiquetar la fuerza de las correlaciones. Si desea obtener más información, consulte el tema [Configuración de correlación](#), el p. 433.

### ***Opciones del nodo Medias***

La pestaña Opciones permite establecer el umbral de los valores  $p$  que se utiliza para etiquetar los resultados como importantes, secundarios o sin importancia. También puede editar la etiqueta de cada rango. La importancia se mide en una escala de porcentaje y se puede definir en líneas generales como 1 menos la probabilidad de obtener un resultado (como la diferencia en las medias de dos campos) tan extremo o más que el resultado observado debido únicamente al azar. Por ejemplo, un valor  $p$  superior a 0,95 indica menos de un 5% de probabilidad de que el resultado se pueda explicar únicamente por el azar.

Figura 6-42  
Configuración de importancia



**Etiquetas de importancia.** Puede editar las etiquetas utilizadas para etiquetar cada par de campos o grupo del resultado. Las etiquetas por defecto son *importante*, *secundario* y *sin importancia*.

**Valores de corte.** Especifica el umbral para cada rango. Normalmente, los valores  $p$  superiores a 0,95 se considerarían como importantes, mientras que los valores inferiores a 0,9 se considerarían como sin importancia, pero estos umbrales se pueden ajustar según sea necesario.

*Nota:* Las medidas de importancia están disponibles en una serie de nodos. Los cálculos específicos dependen del nodo y del tipo de objetivo y campos de entrada utilizados, aunque los valores se pueden seguir comparando ya que todos se miden en una escala de porcentajes.

## Explorador de resultados del nodo Medias

El explorador de resultados de medias muestra datos de tabulación cruzada y permite realizar operaciones estándar, incluyendo seleccionar y copiar la tabla fila por fila, ordenar por cualquier columna y guardar e imprimir la tabla. Si desea obtener más información, consulte el tema [Visualización de resultados](#) el p. 393.

La información específica de la tabla depende del tipo de comparación (grupos de un campo o diferentes campos).

**Ordenar por.** Permite ordenar los resultados por una columna específica. Pulse en la flecha de arriba o abajo para cambiar la dirección del orden. También puede pulsar en cualquier encabezado de columna para ordenar por esa columna. (Para cambiar la dirección del orden dentro de esa columna, vuelva a pulsar.)

**Ver.** Puede elegir Simple o Avanzado para controlar el nivel de detalle de la visualización. La vista avanzada incluye toda la información de la vista simple pero con detalles adicionales.

### Resultados de las medias al comparar los grupos de un campo

Al comparar los grupos de un campo, el nombre del campo de agrupamiento aparece encima de la tabla de resultados y se presentan las medias y los estadísticos relacionados de forma independiente para cada grupo. La tabla incluye una fila independiente para cada campo de prueba.

Figura 6-43  
Comparación de grupos dentro de un campo.

Campo	Standard*	New Promoti...	Importancia
\$ spent during promotional period	1566.389	1637.500	0.976 ★ Important

Se muestran las siguientes columnas:

- **Campo.** Muestra los nombres de los campos de prueba seleccionados.
- **Medias por grupo.** Muestra la media para cada categoría del campo de agrupamiento. Por ejemplo, puede comparar los grupos que recibieron una oferta especial (*Nueva promoción*) con aquellos que no lo hicieron (*Estándar*). En la vista avanzada también aparecen la desviación típica, el error típico y el recuento.
- **Importancia.** Muestra la etiqueta y el valor de importancia. Si desea obtener más información, consulte el tema [Opciones del nodo Medias](#) el p. 439.

#### Salida avanzada

En la vista avanzada, se muestran las siguientes columnas adicionales.

- **Prueba F.** Esta prueba se basa en el cociente de la varianza entre los grupos y la varianza de cada grupo. Si las medias son las mismas para todos los grupos, se esperaría una razón *F* cercana a 1 ya que son estimaciones de la misma varianza poblacional. Cuanto mayor es dicha razón, mayor es la variación entre grupos y mayor la probabilidad de que exista una diferencia significativa.
- **gl.** Muestra los grados de libertad.

### Resultados de las medias al comparar pares de campos

Al comparar campos diferentes, la tabla de resultados incluye una fila para cada par de campos seleccionado.

Figura 6-44  
Comparación de pares de campos.

Campo uno	Campo dos	Media uno*	Media dos*	Correlación	Diferencia d...	Importancia
Triglyceride	Final triglyce...	138.438	124.375	-0.286 Weak	14.062	0.751 Unimportant
Weight	Final weight	198.375	190.312	0.996 Strong	8.062	1.000 Important

- **Campo uno/dos.** Muestra el nombre del primer y segundo campo de cada par. En la vista avanzada también aparecen la desviación típica, el error típico y el recuento.
- **Media uno/dos.** Muestra la media de cada campo, respectivamente.
- **Correlación.** Mide la fuerza de la relación entre dos campos continuos (rango numérico). Los valores cercanos a +1,0 indican una fuerte asociación positiva y los valores cercanos a -1,0 indican una fuerte asociación negativa. Si desea obtener más información, consulte el tema [Configuración de correlación](#) el p. 433.
- **Diferencia de medias.** Muestra la diferencia entre las medias de los dos campos.
- **Importancia.** Muestra la etiqueta y el valor de importancia. Si desea obtener más información, consulte el tema [Opciones del nodo Medias](#) el p. 439.

### Salida avanzada

Un resultado avanzado añade las siguientes columnas:

**Intervalo de confianza de 95%.** Límites inferior y superior del intervalo en el que la media verdadera se encuentra dentro del 95% de todas las muestras posibles de este tamaño a partir de esta población.

**Prueba T.** El estadístico  $t$  se obtiene dividiendo la diferencia de medias por su error típico. Cuanto mayor es el valor absoluto de este estadístico, mayor es la probabilidad de que las medias no sean iguales.

**gl.** Muestra los grados de libertad del estadístico.

## Nodo Informe

El nodo Informe permite crear informes con formato que contienen texto fijo, así como datos y otras expresiones derivadas de los datos. Puede especificar el formato del informe utilizando plantillas de texto para definir el texto fijo y las construcciones de resultados de datos. Puede proporcionar formato de texto personalizado utilizando etiquetas HTML de la plantilla y

configurando opciones en la pestaña Resultado. Los valores de los datos y otros resultados condicionales se incluyen en el informe utilizando expresiones CLEM de la plantilla.

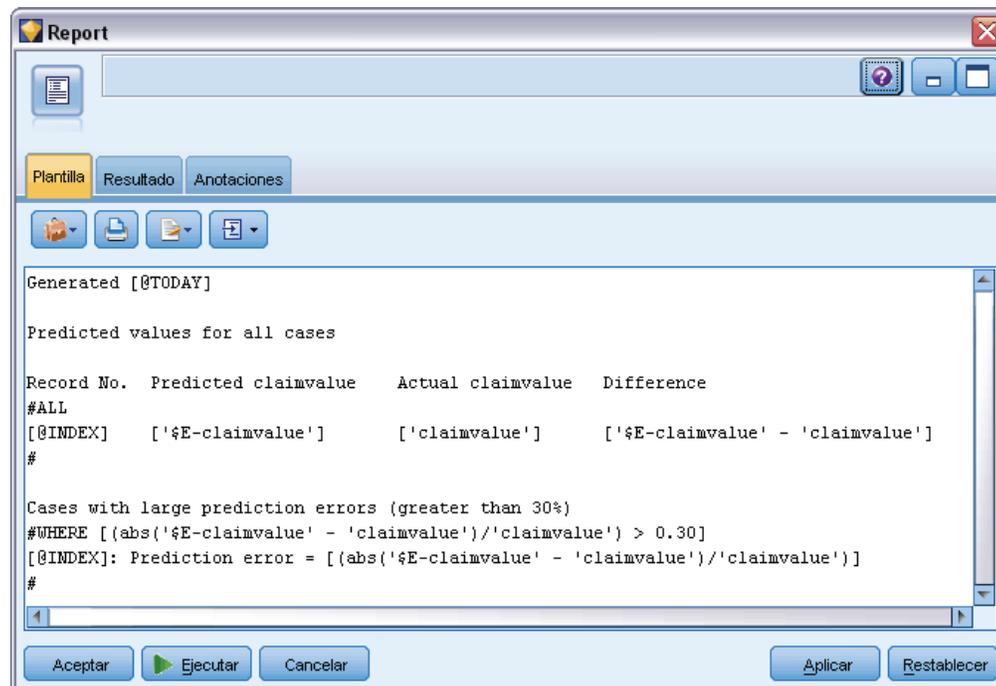
### Alternativas al nodo Informe

El nodo Informe se suele utilizar para enumerar registros o resultados de casos de una ruta, como todos los registros que cumplan una condición determinada. A este respecto, se puede considerar como una alternativa menos estructurada que el nodo Tabla.

- Si desea obtener un informe que enumere la información de campo o cualquier otro dato que esté definido en la ruta en lugar de los propios datos (como definiciones de campo especificadas en un nodo Tipo) puede utilizar un proceso en su lugar.
- Para generar un informe que incluya varios objetos de resultados (como una colección de modelos, tablas y gráficos generados por una o más rutas) y que se pueden mostrar en varios formatos (incluido texto, HTML y Microsoft Word/Office) se puede utilizar un proyecto de IBM® SPSS® Modeler.
- Para generar una lista de nombres de campos sin utilizar un proceso, puede utilizar un nodo Tabla precedido por un nodo Muestrear que descarte todos los registros. Así, se producirá una tabla sin filas, que se puede transponer al exportarla para generar una lista de nombres de campos en una única columna. (Seleccione Transponer datos en la pestaña Resultado del nodo Tabla para realizar esta acción.)

### Pestaña Plantilla del nodo Informe

Figura 6-45  
Nodo Informe: pestaña Plantilla



**Creación de una plantilla.** Para definir el contenido del informe, cree una plantilla en la pestaña Plantilla del nodo Informe. La plantilla consta de líneas de texto, cada una de las cuales especifica algo sobre el contenido del informe y algunas líneas de etiqueta especiales utilizadas para indicar el ámbito de las líneas de contenido. Con cada línea de contenido, las expresiones CLEM entre corchetes ([]) se evalúan antes de que la línea se envíe al informe. Existen tres ámbitos posibles para una línea en la plantilla:

**Fija.** Las líneas que no están marcadas se consideran fijas. Las líneas fijas se copian en el informe sólo una vez, después de la evaluación de cualquier expresión que contengan. Por ejemplo, la línea

```
Este informe fue generado: [@TODAY]
```

copiará una sola línea que contiene el texto y la fecha actual en el informe.

**Globales (iterar ALL).** Las líneas que se encuentran entre las etiquetas especiales #ALL y # se copian en el informe una vez para cada registro de datos de entrada. Las expresiones CLEM (entre paréntesis) se evalúan según el registro actual para cada línea de resultado. Por ejemplo, las líneas

```
#ALL
Para el registro [@INDEX], el valor de EDAD es [AGE]
#
```

incluiría una línea para cada registro indicando el número del registro y la edad.

Para generar una lista de todos los registros:

```
#ALL
[Edad] [Sexo] [Colesterol] [PS]
#
```

**Condicional (iterar WHERE).** Las líneas que se encuentran entre las etiquetas especiales #WHERE <condición> y # se copian en el informe una vez para cada registro en el que se cumple la condición especificada. La condición es una expresión CLEM. (En la condición WHERE, los paréntesis son opcionales.) Por ejemplo, las líneas

```
#WHERE [SEX = 'M']
El hombre del registro nº [@INDEX] tiene [AGE] años.
#
```

escribirá una línea en el archivo para cada registro con un valor *M* para sexo. El informe completo contendrá las líneas fijas, globales y condicionales definidas al aplicar la plantilla a los datos de entrada.

Puede especificar opciones para mostrar o guardar los resultados mediante la pestaña Resultado, común a varios tipos de nodos de resultados. Si desea obtener más información, consulte el tema [Pestaña Resultado del nodo de resultados](#) el p. 399.

### **Salida de datos en formato XML o HTML**

Puede incluir etiquetas XML o HTML directamente en la plantilla para escribir informes en cualquiera de estos formatos. Por ejemplo, la siguiente plantilla genera una tabla HTML.

Este informe está escrito en HTML.  
Sólo se incluyen los registros donde la edad es superior a 60.

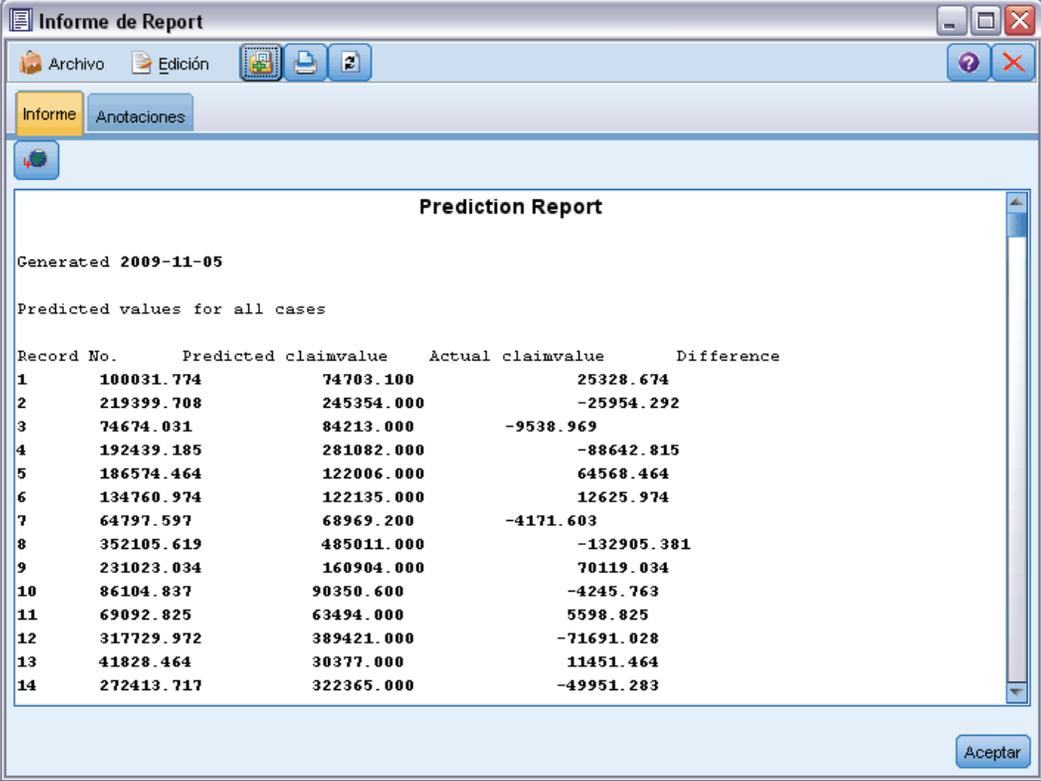
```
<HTML>
<TABLE border="2">
  <TR>
    <TD>Edad</TD>
    <TD>PS</TD>
    <TD>Colesterol</TD>
    <TD>Medicamento</TD>
  </TR>

#WHERE Edad > 60
<TR>
  <TD>[Edad]</TD>
  <TD>[PS]</TD>
  <TD>[Colesterol]</TD>
  <TD>[Medicamento]</TD>
</TR>
#
</TABLE>
</HTML>
```

### ***Explorador de resultados del nodo Informe***

El explorador de informes muestra el contenido del informe generado. Las opciones normales de almacenamiento, exportación e impresión están disponibles en el menú Archivo y las opciones de edición normales están disponibles en el menú Edición. Si desea obtener más información, consulte el tema [Visualización de resultados](#) el p. 393.

Figura 6-46  
Explorador de informes



The screenshot shows a window titled 'Informe de Report' with a menu bar containing 'Archivo' and 'Edición'. Below the menu bar are two tabs: 'Informe' (selected) and 'Anotaciones'. The main content area displays a 'Prediction Report' with the following text and table:

Generated 2009-11-05

Predicted values for all cases

Record No.	Predicted claimvalue	Actual claimvalue	Difference
1	100031.774	74703.100	25328.674
2	219399.708	245354.000	-25954.292
3	74674.031	84213.000	-9538.969
4	192439.185	281082.000	-88642.815
5	186574.464	122006.000	64568.464
6	134760.974	122135.000	12625.974
7	64797.597	68969.200	-4171.603
8	352105.619	485011.000	-132905.381
9	231023.034	160904.000	70119.034
10	86104.837	90350.600	-4245.763
11	69092.825	63494.000	5598.825
12	317729.972	389421.000	-71691.028
13	41828.464	30377.000	11451.464
14	272413.717	322365.000	-49951.283

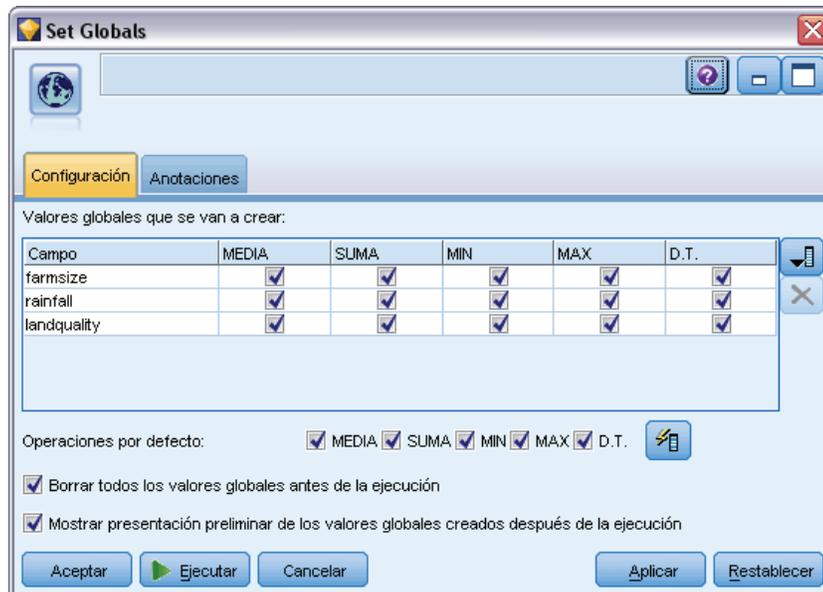
An 'Aceptar' button is located at the bottom right of the window.

## Nodo Val. globales

El nodo Val. globales explora los datos y calcula los valores de resumen que se pueden utilizar en expresiones CLEM. Por ejemplo, puede utilizar un nodo Val. globales para calcular estadísticos para un campo denominado *edad* y, a continuación, utilizar la media global de *edad* en expresiones CLEM insertando la función @GLOBAL\_MEAN(age).

## Pestaña Configuración del nodo Val. globales

Figura 6-47  
Nodo Val. globales: Pestaña Configuración



**Valores globales que se van a crear.** Seleccione el campo o los campos para los que desea que haya valores globales disponibles. Puede seleccionar varios campos. Para cada campo, especifique los estadísticos para calcular asegurándose de que los estadísticos que desea están seleccionados en las columnas junto al nombre del campo.

- **MEDIA.** Valor medio (promedio) para el campo en todos los registros.
- **SUMA.** Suma de valores para el campo en todos los registros.
- **MÍN.** Valor mínimo para el campo.
- **MÁX.** Valor máximo para el campo.
- **D.T.** La desviación típica, medida de variabilidad en los valores de un campo, se calcula como la raíz cuadrada de la varianza.

**Operaciones por defecto.** Las opciones seleccionadas aquí se utilizarán cuando se añadan nuevos campos a la lista de valores globales anterior. Para cambiar el conjunto de estadísticos por defecto, seleccione o anule la selección de los estadísticos de la forma adecuada. También puede utilizar el botón Aplicar para aplicar las operaciones por defecto a todos los campos de la lista.

**Borrar todos los valores globales antes de la ejecución.** Seleccione esta opción para quitar todos los valores globales antes de calcular los valores nuevos. Si esta opción no está seleccionada, los valores recién calculados sustituirán a los valores antiguos, pero los valores globales que no se han vuelto a calcular permanecerán, también, disponibles.

**Mostrar presentación preliminar de los valores globales creados después de la ejecución.** Si selecciona esta opción, la pestaña Globales del cuadro de diálogo de propiedades de la ruta aparecerá tras la ejecución para mostrar los valores globales calculados.

## Complementos de IBM SPSS Statistics

Si tiene instalada una versión compatible de IBM® SPSS® Statistics en su ordenador con la licencia necesaria, puede configurar IBM® SPSS® Modeler para procesar datos con funcionalidad SPSS Statistics mediante los nodos Transformación de Statistics, Modelo de Statistics Resultados de Statistics o Exportar de Statistics.

- Para configurar SPSS Modeler para que funcione con SPSS Statistics y otras aplicaciones, seleccione:

Herramientas > Opciones > Complementos

**IBM SPSS Statistics Interactive.** Introduzca la ruta completa y el nombre del comando (por ejemplo, *C:\Archivos de programa\IBM\SPSS\Statistics\<nn>\stats.exe*) que desea utilizar al iniciar SPSS Statistics directamente en un archivo de datos producido por el nodo de exportación de Statistics. Si desea obtener más información, consulte el tema [Nodo Exportación de Statistics](#) en el capítulo 8 el p. 501.

**Conexión.** Si el servidor SPSS Statistics se encuentra en el mismo servidor que el servidor de IBM® SPSS® Modeler Server, puede activar una conexión entre las dos aplicaciones, lo que aumenta la eficacia al dejar los datos en el servidor durante el análisis. Seleccione Servidor para activar la opción Puerto más abajo. La configuración por defecto es Local.

**Puerto.** Especifique el puerto del servidor para el servidor SPSS Statistics.

**Utilidad de localización de licencias de IBM SPSS Statistics.** Para permitir que SPSS Modeler utilice los nodos Transformación de Statistics, Modelo de Statistics y Resultados de Statistics, debe tener instalada una copia de SPSS Statistics con licencia en el ordenador donde se ejecute la ruta. Además, si está ejecutando en modo distribuido en un SPSS Modeler Server remoto, también debe tener una copia del cliente de SPSS Statistics instalado y con licencia en su equipo cliente de SPSS Modeler.

- Si se está ejecutando SPSS Modeler en modo local (independiente), la copia con licencia de SPSS Statistics debe estar en el equipo local. Pulse en este botón para especificar la ubicación de la instalación de SPSS Statistics local que desee utilizar para la licencia.
- Además, si se ejecuta en modo distribuido en un SPSS Modeler Server remoto, también necesita una versión con licencia de SPSS Statistics en el equipo servidor y la configuración de la licencia también debe realizarse en el servidor. Para ello, en el símbolo del sistema, cambie al directorio *bin* de SPSS Modeler Server y en Windows, ejecute:

```
statisticsutility -location=<ruta al archivo de licencia de IBM SPSS Statistics Server>/bin
```

De forma alternativa, en UNIX, ejecute:

```
./statisticsutility -location=<ruta al archivo de licencia de IBM SPSS Statistics Server>/bin
```

donde *<ruta al archivo de licencia de SPSS Statistics Server>* es el directorio de instalación de un servidor SPSS Statistics con licencia.

Si no tiene una copia con licencia de SPSS Statistics en su equipo local, todavía puede ejecutar el nodo Archivo Statistics en un servidor SPSS Statistics con licencia, pero cualquier intento de ejecutar otro nodo de SPSS Statistics mostrará un mensaje de error.

**Comentarios**

Si tiene problemas para ejecutar los nodos Procedimiento de SPSS Statistics, tenga en cuenta las siguientes sugerencias:

- Si los nombres de campos utilizados en SPSS Modeler tienen más de ocho caracteres (para versiones anteriores a SPSS Statistics 12.0), más de 64 caracteres (para SPSS Statistics 12.0 y versiones posteriores) o contienen caracteres no válidos, es necesario cambiarles el nombre o truncarlos antes de leerlos en SPSS Statistics. Si desea obtener más información, consulte el tema [Cambio de nombre o filtrado de campos para IBM SPSS Statistics](#) en el capítulo 8 el p. 503.
- Si SPSS Statistics se instaló después que SPSS Modeler, puede que necesite especificar la ubicación de la licencia de SPSS Statistics, tal y como se explicó anteriormente.

# Nodos de exportación

## Conceptos básicos sobre nodos de exportación

Los nodos de exportación proporcionan un mecanismo para exportar datos en varios formatos y poder interactuar con otras herramientas de software.

Los siguientes nodos de exportación están disponibles:



El nodo Opciones de exportación escribe datos en orígenes de datos relacionales compatibles con ODBC. Para escribir en un origen de datos ODBC, el origen de datos debe existir y debe tener permiso para escribir en él. Si desea obtener más información, consulte el tema [Nodo Exportar base de datos](#) el p. 451.



El nodo Archivo plano produce datos en un archivo de texto delimitado. Esto es útil para exportar datos que se pueden leer con otro software de hoja de cálculo o de análisis. Si desea obtener más información, consulte el tema [nodo de exportación Archivo plano](#) el p. 472.



El nodo Exportar Statistics ofrece los resultados en formato IBM® SPSS® Statistics.sav *de*. Los archivos .sav se pueden leer con SPSS Statistics Base y otros productos. Este es también el formato utilizado para los archivos caché de IBM® SPSS® Modeler. Si desea obtener más información, consulte el tema [Nodo Exportación de Statistics](#) en el capítulo 8 el p. 501.



El nodo de exportación IBM® SPSS® Data Collection abre los datos en el formato utilizado por el software de investigación de mercados Data Collection. Se debe instalar la biblioteca de datos de Data Collection para utilizar este nodo. Si desea obtener más información, consulte el tema [Nodos de exportación IBM SPSS Data Collection](#) el p. 473.



El nodo Exportar SAS produce datos en formato SAS, para leerlos en SAS o en un paquete de software compatible con SAS. Hay disponibles tres formatos de archivos SAS: SAS para Windows/OS2, SAS para UNIX o SAS versión 7/8. Si desea obtener más información, consulte el tema [Nodo Exportar SAS](#) el p. 479.



El nodo de exportación Excel ofrece los datos resultantes en formato de Microsoft Excel (.xls). Si lo desea, puede elegir iniciar automáticamente Excel y abrir el archivo exportado cuando se ejecute el nodo. Si desea obtener más información, consulte el tema [Nodo de exportación Excel](#) el p. 480.



El nodo de exportación XML exporta datos a un archivo en formato XML. También puede crear un nodo de origen XML para leer los datos exportados a la ruta. Si desea obtener más información, consulte el tema [Nodo de exportación XML](#) el p. 481.

## Nodo Exportar base de datos

Puede usar nodos Base de datos para escribir datos en fuentes de datos relacionales compatibles con ODBC, que se explican en la descripción del nodo de fuente Base de datos. Si desea obtener más información, consulte el tema [Nodo de fuente Base de datos](#) en el capítulo 2 el p. 15.

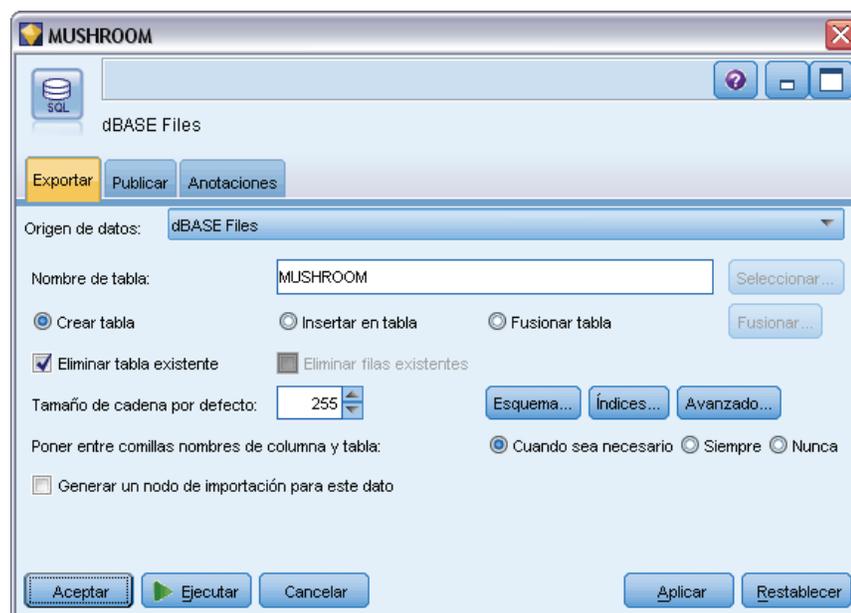
Siga estos pasos generales para escribir datos en una base de datos:

- ▶ Instale un controlador ODBC y configure una fuente de datos en la base de datos que desee utilizar.
- ▶ En la pestaña Exportar del nodo Base de datos, especifique la fuente de datos y la tabla en la que desea escribir. Puede crear una tabla nueva o insertar datos en una existente.
- ▶ Especifique opciones adicionales según sea necesario.

Estos pasos se describen de forma más detallada en los siguientes temas.

### Pestaña Exportar del nodo Base de datos

Figura 7-1  
Nodo Exportar base de datos, pestaña Exportar



**Origen de datos.** Muestra el origen de datos seleccionado. Introduzca el nombre o selecciónelo de la lista desplegable. Si no ve en la lista la base de datos que desea, seleccione Añadir nueva conexión de base de datos y busque la base de datos desde el cuadro de diálogo Conexiones a la base de datos. Si desea obtener más información, consulte el tema [Adición de conexión a la base de datos](#) en el capítulo 2 el p. 18.

**Nombre de tabla.** Introduzca el nombre de la tabla a la que desee enviar los datos. Si selecciona la opción Insertar en tabla, puede seleccionar una tabla existente de la base de datos pulsando en el botón Seleccionar.

**Crear tabla.** Seleccione esta opción para crear una nueva tabla de base de datos o para sobrescribir una tabla de base de datos existente.

**Insertar en tabla.** Seleccione esta opción para insertar los datos como nuevas filas en una tabla de base de datos existente.

**Tabla Fusión.** (Donde esté disponible) Seleccione esta opción para actualizar columnas de base de datos seleccionadas con valores de los campos de datos de origen correspondiente. Si selecciona esta opción activará el botón Fusión, que muestra un cuadro de diálogo donde podrá asignar campos de orígenes de datos a columnas de base de datos.

**Eliminar tabla existente.** Seleccione esta opción para eliminar cualquier tabla existente con el mismo nombre al crear una tabla nueva.

**Eliminar filas existentes.** Seleccione esta opción para eliminar filas existentes de la tabla antes de exportar cuando se inserte en una tabla.

*Nota:* si está seleccionada alguna de las dos opciones anteriores, recibirá el mensaje Advertencia de sobrescritura cuando ejecute el nodo. Para suprimir las advertencias, anule la selección de Avisar cuando un nodo sobrescriba una tabla de la base de datos en la pestaña Notificaciones del cuadro de diálogo Opciones de usuario.

**Tamaño de cadena por defecto.** Los campos marcados como sin tipo en un nodo Tipo anterior de la ruta se escriben en la base de datos como campos de cadena. Especifique el tamaño de las cadenas que se van a utilizar para campos sin tipo.

Pulse en Esquema para abrir un cuadro de diálogo donde podrá establecer diversas opciones de exportación (para bases de datos que admitan esta función), establecer los tipos de datos de SQL para los campos y especificar la clave primaria para la indización de la base de datos. Si desea obtener más información, consulte el tema [Opciones de Esquema de exportación de base de datos](#) el p. 455.

Pulse en Índices para especificar las opciones de indización de la tabla exportada para mejorar el rendimiento de la base de datos. Si desea obtener más información, consulte el tema [Opciones de índice de exportación de base de datos](#) el p. 459.

Pulse en Avanzado para especificar opciones de carga masiva y de confirmación de base de datos. Si desea obtener más información, consulte el tema [Opciones avanzadas de exportación de base de datos](#) el p. 461.

**Poner entre comillas nombres de columna y tabla.** Seleccione opciones utilizadas al enviar una instrucción CREATE TABLE a la base de datos. Las tablas o columnas con espacios o caracteres que no sean estándar deben ponerse entre comillas.

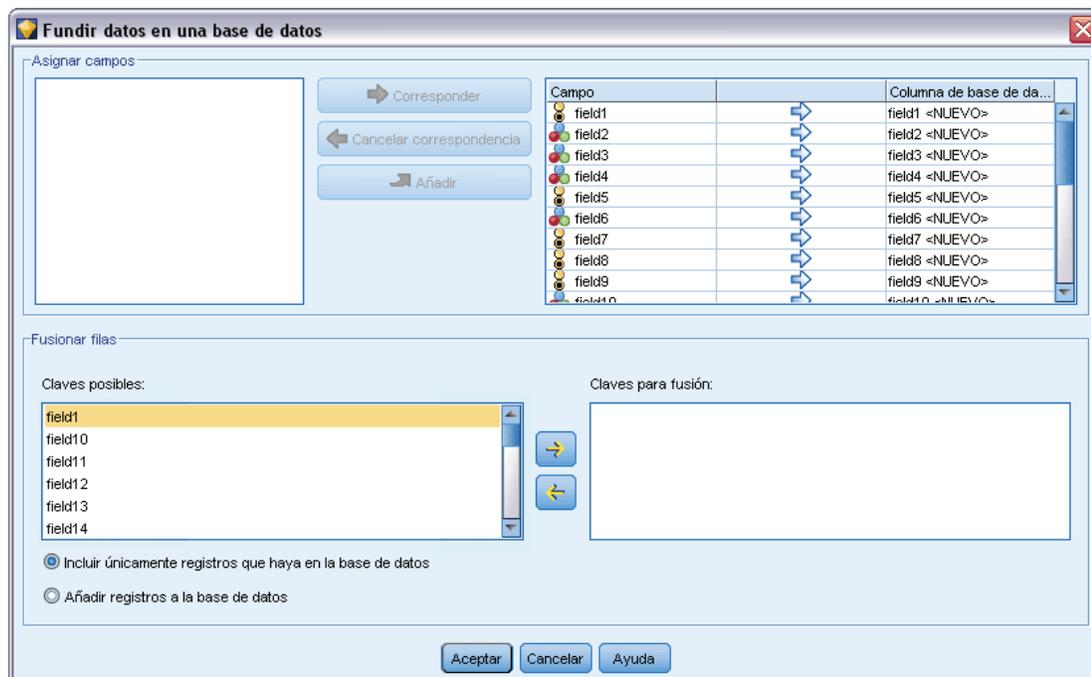
- **Cuando sea necesario.** Selecciónelo para permitir que IBM® SPSS® Modeler determine automáticamente cuándo son necesarias las comillas de forma individual.
- **Siempre.** Para poner siempre entre comillas los nombres de tabla y columna.
- **Nunca.** Para desactivar la utilización de comillas.

**Generar un nodo de importación para este dato.** Para generar un nodo de fuente de Base de datos para los datos tal y como se exportan en la tabla y la fuente de datos especificadas. Al ejecutarse, este nodo se añade al lienzo de rutas.

## Opciones de fusión Exportar base de datos

Este cuadro de diálogo le permite asignar campos desde los datos de fuente a columnas en la tabla de base de datos de destino. Si un campo de datos de fuente se asigna a una columna de base de datos, el valor de columna se sustituye por el valor de los datos de fuente cuando se ejecuta la ruta. Los campos de fuente sin asignar se quedan sin cambiar en la base de datos.

Figura 7-2  
Asignación de campos de datos de fuente a columnas de base de datos



**Corresponder campos.** Aquí puede especificar la asignación entre campos de datos de fuente y columnas de base de datos. Los campos de datos de fuente con el mismo nombre que las columnas de la base de datos se asignan automáticamente.

- **Corresponder.** Asigna un campo de datos de fuente en la lista de campos en la parte izquierda del botón a una columna de base de datos seleccionada en la lista de la derecha. Puede asignar uno o más campos a la vez, pero el número de entradas seleccionadas en ambas listas deben ser las mismas.
- **Cancelar correspondencia.** Elimina la asignación para una o más columnas de base de datos seleccionadas. Este botón se activa cuando se selecciona una columna de base de datos o campo en la tabla que aparece a la derecha del cuadro de diálogo.
- **Añadir.** Añade uno o más campos de datos de fuente seleccionados en la lista de campos a la izquierda del botón a la lista de la derecha, lista para la asignación. Este botón se activa cuando selecciona un campo de la lista de la izquierda y no existe ningún campo con ese nombre en la lista de la derecha. Al pulsar en este botón se asigna el campo seleccionado a una nueva columna de base de datos con el mismo nombre. La palabra <NEW> aparece tras el nombre de columna de base de datos para indicar que se trata de un nuevo campo.

**Fusionar filas.** Usa un campo clave, como el *ID de transacción*, para fusionar registros que compartan el mismo valor en este campo. Equivale a una “equi-uniión” de base de datos. Los valores clave deben ser aquellos de las claves primarias; es decir, deben ser únicos y no pueden contener valores nulos.

- **Claves posibles.** Enumera todos los campos encontrados en todas las fuentes de datos de entrada. Seleccione uno o más campos de esta lista y utilice el botón de flecha para añadirlos como campos clave para fusionar los registros. Cualquier campo asignado con una columna de base de datos asignada está disponible como clave, salvo que los campos se añadan como nuevas columnas de bases de datos (se muestran con <NEW> tras el nombre) no están disponibles.
- **Claves para fusión.** Enumera todos los campos usados para fusionar registros desde todas las fuentes de datos de entrada basándose en valores de los campos clave. Para quitar una clave de la lista, selecciónela y use el botón de flecha para devolverla a la lista Claves posibles. Cuando se seleccionan varios campos clave, la opción de la parte inferior se desactiva.
- **Incluir únicamente registros que hay en la base de datos.** Realiza una unión parcial; si el registro está en la base de datos y la ruta, se actualizarán los campos asignados.
- **Añade registros a la base de datos.** Realiza una unión exterior; se fusionarán todos los registros de la ruta (si existe el mismo registro en la base de datos) o se añadirán (si el registro no existe en la base de datos).

***Para asignar un campo de datos de fuente a una columna de base de datos nueva***

- ▶ Pulse en el nombre de campo de fuente en la lista de la izquierda, en Corresponder campos.
- ▶ Pulse en el botón Añadir para comprobar la asignación.

***Para asignar un campo de datos de fuente a una columna de base de datos existente***

- ▶ Pulse en el nombre de campo de fuente en la lista de la izquierda, en Corresponder campos.
- ▶ Pulse en el nombre de la columna en Columna de base de datos en la parte derecha.
- ▶ Pulse en el botón Corresponder para completar la asignación.

***Para eliminar una asignación***

- ▶ En la lista de la derecha, debajo de Campo, pulse en el nombre del campo cuya asignación desea eliminar.
- ▶ Pulse en el botón Cancelar correspondencia.

***Para cancelar la selección de un campo en cualquiera de las listas***

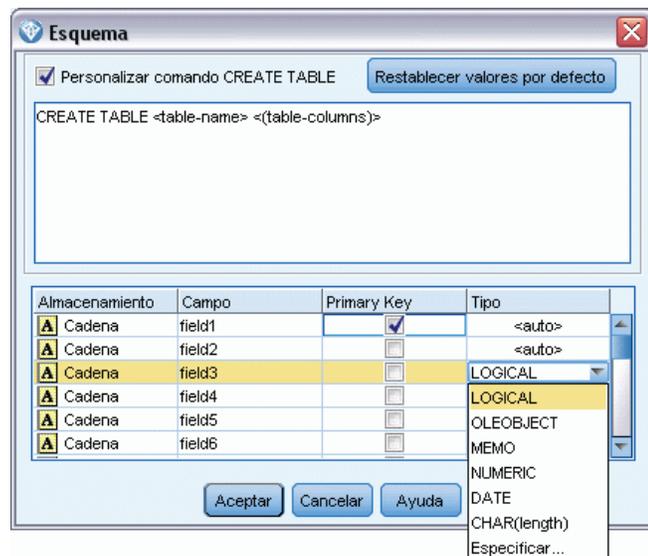
- ▶ Mantenga pulsada la tecla CTRL y pulse el nombre del campo.

## Opciones de Esquema de exportación de base de datos

En el cuadro de diálogo Esquema de exportación de base de datos, puede establecer opciones para la exportación de base de datos (para bases de datos que admitan estas opciones), establecer tipos de datos SQL para sus campos, especificar qué campos son las claves primarias y personalizar la instrucción CREATE TABLE generada al exportar.

Figura 7-3

Ejemplo de cuadro de diálogo Esquema de exportación de base de datos



El cuadro de diálogo tiene varias partes:

- La sección de la parte superior (si se muestra) contiene opciones para la exportación a una base de datos que admita estas opciones. Esta sección no se muestra si no está conectado a dicha base de datos.
- El campo de texto de la parte central muestra la plantilla utilizada para generar el comando CREATE TABLE, que por defecto tiene el siguiente formato:  
CREATE TABLE <table-name> <(table columns)>
- La tabla de la parte inferior le permite especificar el tipo de datos SQL para cada campo e indicar qué campos son claves primarias como se explica a continuación. El cuadro de diálogo genera automáticamente los valores de los parámetros de <table-name> y <(table columns)> basados en las especificaciones de la tabla.

### Configuración de opciones de exportación de base de datos

Si esta sección se muestra, puede especificar determinados ajustes para exportar a la base de datos. Los tipos de bases de datos que admiten esta función son los siguientes:

- IBM InfoSphere Warehouse ejecutándose en DB2 9.1 o posterior. Si desea obtener más información, consulte el tema [Opciones para IBM DB2 InfoSphere Warehouse](#) el p. 457.

- SQL Server 2008 o posterior, Enterprise Edition y Developer Edition. Si desea obtener más información, consulte el tema [Opciones para SQL Server](#) el p. 457.
- Oracle 10g y 11gR1 o posterior, Enterprise Edition o Personal Edition. Si desea obtener más información, consulte el tema [Opciones para Oracle](#) el p. 457.

### **Personalización de las instrucciones CREATE TABLE**

Utilizando la parte del campo de texto de este cuadro de diálogo, puede añadir opciones adicionales específicas de la base de datos a la instrucción CREATE TABLE.

- ▶ Seleccione la casilla de verificación Personalizar comando CREAR TABLA para activar la ventana de texto.
- ▶ Añada cualquier opción específica de la base de datos a la instrucción. Asegúrese de mantener el texto de los parámetros <table-name> y (<table-columns>) ya que los sustituye por las definiciones de columna y nombre real de tabla mediante IBM® SPSS® Modeler.

### **Configuración de tipos de datos SQL**

Por defecto, SPSS Modeler permitirá que el servidor de la base de datos asigne tipos de datos SQL automáticamente. Para sobrescribir el tipo automático para un campo, busque la fila correspondiente al campo y seleccione el tipo que desee de la lista desplegable en la columna *Tipo* de la tabla de esquema. Puede utilizar Mayús y pulsar para seleccionar más de una fila.

Para tipos con una longitud, precisión o argumento de escala (BINARY, VARBINARY, CHAR, VARCHAR, NUMERIC y NUMBER), debe especificar una longitud en lugar de permitir que el servidor de la base de datos asigne una longitud automática. Por ejemplo, especificar un valor adecuado, como VARCHAR(25), para la longitud garantiza que el tipo de almacenamiento en SPSS Modeler se sobrescribirá si esa es su intención. Para sobrescribir la asignación automática, seleccione Especificar de la lista desplegable Tipo y sustituya la definición de tipo con la instrucción de definición de tipo de SQL que desee.

Figura 7-4

Cuadro de diálogo Especificar tipo de los resultados de la base de datos



La forma más fácil de hacerlo es seleccionando primero el tipo más cercano a la definición de tipo que desee y, a continuación, seleccionando Especificar para modificar esa definición. Por ejemplo, para establecer un tipo de datos de SQL en VARCHAR(25), establezca primero el tipo en VARCHAR(longitud) desde la lista desplegable Tipo y, a continuación, seleccione Especificar y sustituya la longitud del texto por el valor 25.

### **Claves primarias**

Si una o más columnas de la tabla exportada debe tener un valor o una combinación de valores únicos para cada fila, puede indicarlo seleccionando la casilla de verificación Clave primaria para cada campo al que se aplique. La mayoría de las bases de datos no permiten que se modifique la tabla de ninguna manera que invalide una restricción de clave primaria y crearán automáticamente un índice sobre la clave primaria para ayudar a aplicar dicha restricción. (Si lo desea, puede crear índices para otros campos en el cuadro de diálogo Índices. Si desea obtener más información, consulte el tema [Opciones de índice de exportación de base de datos](#) el p. 459.)

### **Opciones para IBM DB2 InfoSphere Warehouse**

**Espacio de tabla.** El espacio de tabla que debe utilizarse para la exportación. Los administradores de la base de datos pueden crear o configurar espacios de tabla con particiones. Recomendamos que seleccione uno de estos espacios de tabla (en lugar del valor por defecto) para utilizarlo para la exportación de base de datos.

**Dividir los datos por campo.** Especifica el campo de entrada que debe utilizarse en la partición.

**Utilizar compresión.** Si está seleccionado, crea tablas para la exportación con compresión (por ejemplo, el equivalente a CREATE TABLE MYTABLE(...) COMPRESS YES; en SQL).

### **Opciones para SQL Server**

**Utilizar compresión.** Si está seleccionado, crea tablas para la exportación con compresión.

**Compresión para.** Elija el nivel de compresión.

- **Fila.** Activa la compresión a nivel de fila (por ejemplo, el equivalente a CREATE TABLE MYTABLE(...) WITH (DATA\_COMPRESSION = ROW); en SQL).
- **Página.** Activa la compresión a nivel de página (por ejemplo, CREATE TABLE MYTABLE(...) WITH (DATA\_COMPRESSION = PAGE); en SQL).

### **Opciones para Oracle**

#### **Configuración de Oracle 10g**

**Utilizar compresión.** Si está seleccionado, crea tablas para la exportación con compresión. Para esta versión de la base de datos, únicamente está disponible la compresión básica (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS; en SQL).

#### **Configuración de Oracle 11gR1**

**Utilizar compresión.** Si está seleccionado, crea tablas para la exportación con compresión.

**Compresión para.** Elija el nivel de compresión.

- **Por defecto.** Activa la compresión por defecto (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS; en SQL). En este caso, tiene el mismo efecto que la opción Operaciones de carga directa.
- **Operaciones de carga directa.** Activa la compresión únicamente para operaciones de inserción masiva (ruta directa) (por ejemplo, CREATE TABLE MYTABLE(...)COMPRESS FOR DIRECT\_LOAD OPERATIONS; en SQL).
- **Todas las operaciones.** Activa la compresión para todas las operaciones (por ejemplo, CREATE TABLE MYTABLE(...)COMPRESS FOR ALL OPERATIONS; en SQL).

#### **Configuración de Oracle 11gR2: opción Básica**

**Utilizar compresión.** Si está seleccionado, crea tablas para la exportación con compresión.

**Compresión para.** Elija el nivel de compresión.

- **Por defecto.** Activa la compresión por defecto (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS; en SQL). En este caso, tiene el mismo efecto que la opción Básica.
- **Básica.** Activa la compresión básica (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS BASIC; en SQL).

#### **Configuración de Oracle 11gR2: opción Avanzada**

**Utilizar compresión.** Si está seleccionado, crea tablas para la exportación con compresión.

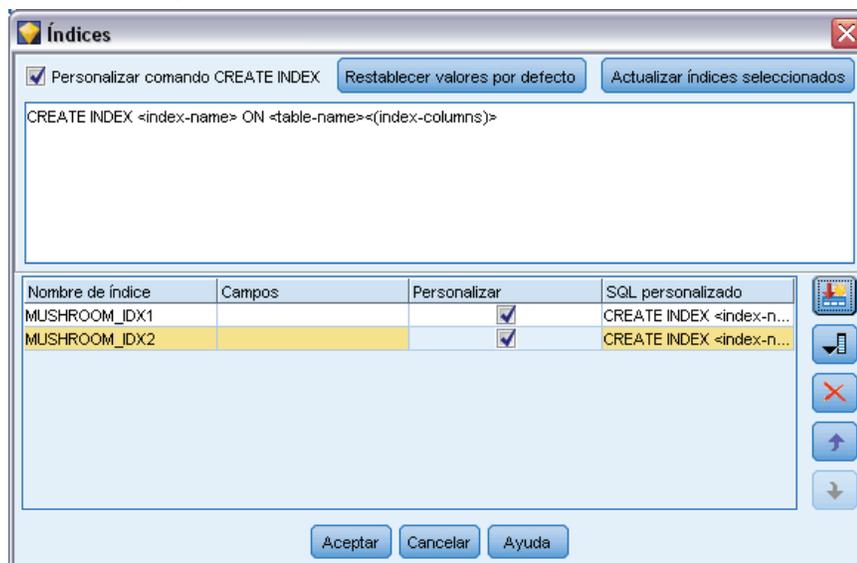
**Compresión para.** Elija el nivel de compresión.

- **Por defecto.** Activa la compresión por defecto (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS; en SQL). En este caso, tiene el mismo efecto que la opción Básica.
- **Básica.** Activa la compresión básica (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS BASIC; en SQL).
- **OLTP.** Activa la compresión OLTP (por ejemplo, CREATE TABLE MYTABLE(...)COMPRESS FOR OLTP; en SQL).
- **Consulta baja/alta.** (Servidores Exadata únicamente) Activa la compresión de columna híbrida para la consulta (por ejemplo, CREATE TABLE MYTABLE(...)COMPRESS FOR QUERY LOW; o CREATE TABLE MYTABLE(...)COMPRESS FOR QUERY HIGH; en SQL). La compresión para consulta es de utilidad en entornos de almacenamiento de datos; HIGH proporciona una relación de compresión mayor que LOW.
- **Archivo bajo/alto.** (Servidores Exadata únicamente) Activa la compresión de columna híbrida para el archivo (por ejemplo, CREATE TABLE MYTABLE(...)COMPRESS FOR ARCHIVE LOW; o CREATE TABLE MYTABLE(...)COMPRESS FOR ARCHIVE HIGH; en SQL). La compresión para archivo es de utilidad para comprimir datos que se almacenarán durante largos periodos de tiempo; HIGH proporciona una relación de compresión mayor que LOW.

## Opciones de índice de exportación de base de datos

El cuadro de diálogo Índices le permite crear índices en tablas de bases de datos exportadas desde IBM® SPSS® Modeler. Puede especificar los conjuntos de campo que desea incluir y personalizar el comando `CREATE INDEX`, según sea necesario.

Figura 7-5  
Cuadro de diálogo Índices de los resultados de la base de datos



El cuadro de diálogo tiene dos partes:

- El campo de texto de la parte superior muestra una plantilla que se puede utilizar para generar uno o más comandos `CREATE INDEX`, que por defecto tienen el formato:  
`CREATE INDEX <index-name> ON <table-name>`
- La tabla de la parte inferior del cuadro de diálogo le permite añadir especificaciones para cada índice que desee crear. En cada índice, especifique el nombre del índice y los campos o columnas que desee incluir. El cuadro de diálogo genera automáticamente los valores de los parámetros `<index-name>` y `<table-name>` en consecuencia.

Por ejemplo, el SQL generado para un índice único de los campos *empid* y *deptid* tendría un aspecto similar a este:

```
CREATE INDEX MYTABLE_IDX1 ON MYTABLE(EMPID,DEPTID)
```

Puede añadir varias filas para crear varios índices. Se generará un comando `CREATE INDEX` diferente para cada fila.

### Personalización del comando `CREATE INDEX`

Si lo desea, puede personalizar el comando `CREATE INDEX` para todos los índices o solamente para un índice específico, lo que le proporciona la flexibilidad para adaptarse a opciones o requisitos específicos de la base de datos y aplicar personalizaciones a todos los índices o solamente a unos determinados, según sea necesario.

- Seleccione Personalizar comando CREATE INDEX en la parte superior del cuadro de diálogo para modificar la plantilla que se utiliza para todos los índices añadidos posteriormente. Tenga en cuenta que los cambios no se aplicarán automáticamente a los índices que ya se hayan añadido a la tabla.
- Seleccione una o más filas de la tabla y, a continuación, pulse en Actualizar índices seleccionados en la parte superior del cuadro de diálogo para aplicar las personalizaciones actuales a todas las filas seleccionadas.
- Seleccione la casilla de verificación Personalizar de cada fila para modificar la plantilla del comando solamente para tal índice.

Tenga en cuenta que los valores de los parámetros <index-name> y <table-name> se generan automáticamente a través del cuadro de diálogo según las especificaciones de la tabla y no se pueden editar directamente.

**Palabra clave BITMAP.** Si está utilizando una base de datos Oracle, puede personalizar la plantilla para crear un índice de mapa de bits en lugar de un índice estándar, de la siguiente manera:

```
CREATE BITMAP INDEX <index-name> ON <table-name>
```

Los índices de mapas de bits pueden ser útiles para la indización de columnas con una pequeña cantidad de valores distintos. El SQL resultante tendría una apariencia similar a esto:

```
CREATE BITMAP INDEX MYTABLE_IDX1 ON MYTABLE(COLOR)
```

**palabra clave UNIQUE.** La mayoría de las bases de datos admiten la clave UNIQUE en el comando CREATE INDEX. Esto aplica una restricción de exclusividad parecida a la restricción de clave primaria de la tabla subyacente.

```
CREATE UNIQUE INDEX <index-name> ON <table-name>
```

Tenga en cuenta que para los campos ya designados como claves primarias, no es necesaria esta especificación. La mayoría de las bases de datos crearán automáticamente un índice para cualquier campo especificado como campo clave primario dentro del comando CREATE TABLE, por lo que no es necesario crear explícitamente índices en estos campos. Si desea obtener más información, consulte el tema [Opciones de Esquema de exportación de base de datos](#) el p. 455.

**palabra clave FILLFACTOR.** Algunos parámetros físicos para el índice se pueden ajustar con precisión. Por ejemplo, SQL Server permite al usuario equilibrar el tamaño del índice (después de la creación inicial) respecto a los costes de mantenimiento a medida que se realizan cambios posteriores a la tabla.

```
CREATE INDEX MYTABLE_IDX1 ON MYTABLE(EMPID,DEPTID) WITH FILLFACTOR=20
```

### **Otros comentarios**

- Si ya existe un índice con el nombre especificado, la creación del índice fallará. Cualquier fallo se tratará en principio como una advertencia, lo que permite que se creen los índices siguientes y que después se vuelvan a informar como errores en el registro de mensajes una vez que se hayan intentado todos los índices.
- Para un rendimiento óptimo, los índices se deberían crear una vez que se hayan cargado los datos en la tabla. Los índices deben contener al menos una columna.

- Antes de ejecutar el nodo, puede obtener una presentación preliminar del SQL generado en el registro de mensajes.
- Las opciones que permiten especificar las claves primarias y los índices no están disponibles para las tablas temporales escritas en la base de datos (es decir, cuando el almacenamiento en caché del nodo está activo). Sin embargo, el sistema puede crear índices en la tabla temporal según sea necesario dependiendo de la manera en que se usen los datos en los nodos siguientes. Por ejemplo, si los datos en caché se unen posteriormente mediante una columna *DEPT*, tendría sentido indizar la tabla en caché en dicha columna.

### **Optimización de consultas e índices**

En algunos sistemas de administración de bases de datos, una vez que se ha creado, cargado e indizado una base de datos, es necesario un paso más para que el optimizador pueda utilizar los índices para acelerar la ejecución de consultas en la tabla nueva. Por ejemplo, en Oracle, el optimizador de consultas basado en costes necesita que se analice una tabla antes de que se puedan utilizar los índices en la optimización de consultas. El archivo de propiedades interno ODBC de Oracle (no visible para el usuario) contiene una opción para que esto suceda, de la siguiente manera:

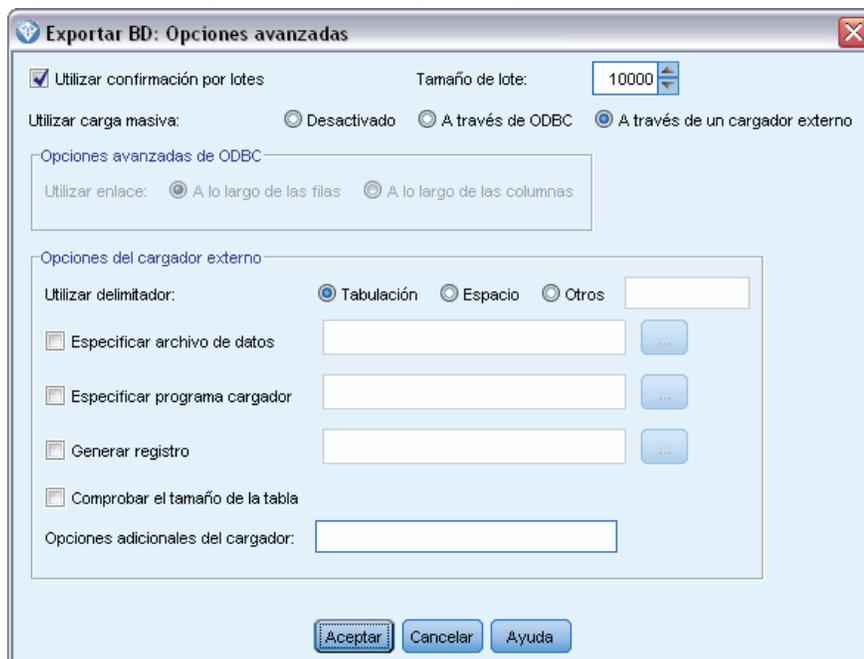
```
# Define el SQL que se ejecutará después de que se hayan creado una tabla y  
# todos los números de índices asociados y se haya  
rellenado table_analysis_sql, 'ANALYZE TABLE <table-name> COMPUTE STATISTICS'
```

Este paso se ejecuta cada vez que se crea una tabla en Oracle (independientemente de que se definan índices o claves primarias). Si fuera necesario, el archivo de propiedades ODBC para bases de datos adicionales se puede personalizar de una forma parecida. Póngase en contacto con la asistencia técnica.

### **Opciones avanzadas de exportación de base de datos**

Al pulsar en el botón Avanzado en los cuadros de diálogo del nodo de exportación de la base de datos, se abre un nuevo cuadro de diálogo en el que puede especificar los detalles técnicos para exportar resultados a una base de datos.

Figura 7-6  
Especificación de opciones avanzadas para la exportación de la base de datos



**Utilizar confirmación por lotes.** Para desactivar las confirmaciones fila por fila en la base de datos.

**Tamaño de lote.** Especifica el número de registros para enviar a la base de datos antes de confirmar en la memoria. Al reducir este número se proporciona una integridad de datos mayor al precio de velocidades de transferencia más lentas. Puede que desee ajustar este número para que su base de datos tenga un rendimiento óptimo.

**Opciones de InfoSphere Warehouse.** Solamente se muestra si está conectado a una base de datos de InfoSphere Warehouse (IBM DB2 9.7 o posterior). No registrar actualizaciones le permite evitar eventos de registro al crear tablas e insertar datos.

**Utilizar carga masiva.** Especifica un método para la carga masiva de datos a la base de datos directamente desde IBM® SPSS® Modeler. Es necesario algo de experimentación para seleccionar las opciones de carga masiva que son apropiadas para un escenario determinado.

- **A través de ODBC.** Seleccione utilizar la API de ODBC para ejecutar inserciones de varias filas con mayor eficacia que la exportación normal a la base de datos. Seleccione entre enlace a lo largo de las filas o de las columnas en las opciones siguientes.
- **A través de un cargador externo.** Seleccione utilizar un programa cargador masivo específico para su base de datos. Si selecciona esta opción se activan varias de las opciones siguientes.

**Opciones avanzadas de ODBC.** Estas opciones solamente están disponibles cuando A través de ODBC está seleccionado. Tenga en cuenta que puede que no todos los controladores ODBC admitan esta funcionalidad.

- **A lo largo de las filas.** Seleccione un enlace a lo largo de las filas para utilizar la llamada `SQLBulkOperations` para cargar datos a la base de datos. El enlace a lo largo de las filas suele mejorar la velocidad en comparación con el uso de inserciones parametrizadas que insertan datos según una base de registro por registro.
- **A lo largo de las columnas.** Seleccione utilizar un enlace a lo largo de las columnas para cargar datos a la base de datos. Un enlace a lo largo de las columnas mejora el rendimiento enlazando cada columna de la base de datos (en una instrucción `INSERT` parametrizada) con una matriz de valores  $N$ . Ejecutar la instrucción `INSERT` una vez hace que las filas  $N$  se inserten en la base de datos. Este método puede aumentar el rendimiento de forma espectacular.

**Opciones del cargador externo.** Cuando se especifica A través de un cargador externo, se muestran varias opciones para exportar el conjunto de datos a un archivo y especificar y ejecutar un programa cargador personalizado para cargar los datos desde ese archivo a la base de datos. SPSS Modeler puede interactuar con los cargadores externos para muchos sistemas de base de datos populares. Se han incluido varios procesos con el software que están disponibles junto con documentación técnica en el subdirectorio *scripts*. Tenga en cuenta que para utilizar esta funcionalidad, Python 2.7 debe estar instalado en el mismo ordenador que SPSS Modeler o IBM® SPSS® Modeler Server y el parámetro `python_exe_path` debe estar establecido en el archivo *options.cfg*. Si desea obtener más información, consulte el tema [Programación del cargador masivo](#) el p. 464.

- **Utilizar delimitador.** Especifica qué carácter delimitador debe utilizarse en el archivo exportado. Seleccione Tabulador para delimitar con el tabulador y Espacio para delimitar con espacios. Seleccione Otros para especificar otro carácter, como una coma (,).
- **Especificar archivo de datos.** Para introducir la ruta que se va a utilizar para el archivo de datos escrito durante la carga masiva. Por defecto, se crea un archivo temporal en el directorio temporal del servidor.
- **Especificar programa cargador.** Para especificar el programa de carga masiva. Por defecto, el software busca el subdirectorio *scripts* de la instalación de SPSS Modeler en un proceso Python para ejecutarlo para una base de datos determinada. Se han incluido varios procesos con el software que están disponibles junto con documentación técnica en el subdirectorio *scripts*.
- **Generar registro.** Para generar un archivo de registro en el directorio especificado. El archivo de registro contiene información de error y es útil si la operación de carga masiva falla.
- **Comprobar el tamaño de la tabla.** Para realizar una comprobación de tabla que garantice que el aumento del tamaño de la tabla se corresponda con el número de filas exportadas desde SPSS Modeler.
- **Opciones adicionales del cargador.** Especifica argumentos adicionales para el programa cargador. Utilice comillas para los argumentos que contengan espacios.

Las comillas se incluyen en los argumentos opcionales con una barra invertida. Por ejemplo, la opción especificada como `-comment "This is a \"comment\""` incluye tanto la marca `-comment` como el comentario en sí representado como `This is a "comment"`.

Se puede incluir una sola barra invertida utilizando otra barra invertida. Por ejemplo, la opción especificada como `-specialdir "C:\\Test Scripts\\"` incluye la marca `-specialdir` y el directorio representado como `C:\Test Scripts\`.

## Programación del cargador masivo

El nodo Exportar base de datos incluye opciones de carga masiva en el cuadro de diálogo Opciones avanzadas. Los programas de cargadores masivos se pueden utilizar para cargar datos desde un archivo de texto a una base de datos.

La opción Utilizar carga masiva (a través de un cargador externo) configura IBM® SPSS® Modeler para que realice tres cosas:

- Crear cualquier tabla de base de datos necesaria.
- Exportar los datos a un archivo de texto.
- Invocar un programa cargador masivo para cargar los datos desde dicho archivo a la tabla de la base de datos.

Normalmente, el programa cargador masivo no es la propia utilidad de carga de la base de datos (por ejemplo, la utilidad sqlldr de Oracle), sino un pequeño proceso o programa que forma los argumentos correctos, crea cualquier archivo auxiliar específico de la base de datos (como un archivo de control) y, a continuación, invoca la utilidad de carga de la base de datos. La información de las siguientes secciones le ayudará a editar un cargador masivo existente.

Si lo prefiere, puede escribir su propio programa de carga masiva. Si desea obtener más información, consulte el tema [Desarrollo de programas de carga masiva](#) el p. 469.

### Procesos de carga masiva

SPSS Modeler se entrega con varios programas cargadores masivos para bases de datos diferentes que se implementan utilizando procesos Python. Al ejecutar una ruta que contenga un nodo Exportar base de datos con la opción A través de un cargador externo seleccionada, SPSS Modeler crea una tabla de base de datos (en caso necesario) a través de ODBC, exporta los datos a un archivo temporal en el host que ejecuta IBM® SPSS® Modeler Server y, a continuación, invoca el proceso de carga masiva. A su vez, el proceso ejecuta utilidades proporcionadas por el proveedor de DBMS para cargar datos de los archivos temporales a la base de datos.

*Nota:* La instalación de SPSS Modeler no incluye un intérprete de tiempo de ejecución de Python, por lo que se requiere una instalación independiente de Python. Si desea obtener más información, consulte el tema [Opciones avanzadas de exportación de base de datos](#) el p. 461.

Se proporcionan procesos (en la carpeta `\scripts` del directorio de instalación de SPSS Modeler) para las siguientes bases de datos.

Tabla 7-1  
Procesos cargadores masivos proporcionados

Base de datos	Nombre de proceso	
IBM DB2	<code>db2_loader.py</code>	Si desea obtener más información, consulte el tema <a href="#">Carga masiva de datos a bases de datos IBM DB2</a> el p. 465.
IBM Netezza	<code>netezza_loader.py</code>	Si desea obtener más información, consulte el tema <a href="#">Carga masiva de datos a bases de datos IBM Netezza</a> el p. 466.
Oracle	<code>oracle_loader.py</code>	Si desea obtener más información, consulte el tema <a href="#">Carga masiva de datos a bases de datos Oracle</a> el p. 466.

Base de datos	Nombre de proceso	
SQL Server	<i>mssql_loader.py</i>	Si desea obtener más información, consulte el tema <a href="#">Carga masiva de datos a bases de datos SQL Server</a> el p. 467.
Teradata	<i>teradata_loader.py</i>	Si desea obtener más información, consulte el tema <a href="#">Carga masiva de datos a bases de datos Teradata</a> el p. 468.

### ***Carga masiva de datos a bases de datos IBM DB2***

Los siguientes puntos pueden ayudarle a configurar la carga masiva de IBM® SPSS® Modeler a una base de datos IBM DB2 utilizando la opción de cargador externo en el cuadro de diálogo Exportar BD: Opciones avanzadas.

#### ***Asegúrese de que la utilidad de procesador de línea de comandos (CLP) de DB2 está instalada***

El proceso *db2\_loader.py* invoca el comando DB2 LOAD. Asegúrese de que el procesador de línea de comandos (*db2* en UNIX, *db2cmd* en Windows) está instalado en el servidor en el que se ejecuta *db2\_loader.py* (normalmente, el host que ejecuta IBM® SPSS® Modeler Server).

#### ***Compruebe si el nombre de alias de la base de datos local coincide con el nombre de la base de datos real***

El alias de la base de datos local DB2 es el nombre usado por el software cliente de DB2 para hacer referencia a una base de datos en una instancia de DB2 local o remota. Si el alias de la base de datos local no es igual que el nombre de la base de datos remota, proporcione la opción adicional del cargador:

```
-alias <local_database_alias>
```

Por ejemplo, la base de datos remota se llama STARS en el host GALAXY, pero el alias de la base de datos local DB2 en el host que ejecuta SPSS Modeler Server es STARS\_GALAXY. Utilice la opción adicional del cargador

```
-alias STARS_GALAXY
```

#### ***Codificación de datos con caracteres que no son ASCII***

Si está realizando una carga masiva de datos que no están en formato ASCII, debe asegurarse de que la variable de página de códigos en la sección de configuración de *db2\_loader.py* está configurada correctamente en su sistema.

#### ***Cadenas en blanco***

Las cadenas en blanco se exportan a la base de datos como valores nulos (NULL).

### ***Carga masiva de datos a bases de datos IBM Netezza***

Los siguientes puntos pueden ayudarle a configurar la carga masiva de IBM® SPSS® Modeler a una base de datos IBM Netezza utilizando la opción de cargador externo en el cuadro de diálogo Exportar BD: Opciones avanzadas.

#### ***Asegúrese de que la utilidad `nzload` de Netezza está instalada***

El proceso `netezza_loader.py` invoca la utilidad `nzload` de Netezza. Asegúrese de que `nzload` está instalado y se ha configurado correctamente en el servidor en el que se va a ejecutar `netezza_loader.py`.

#### ***Exportación de datos que no están en ASCII***

Si exporta contenido que no está en formato ASCII, puede que necesite añadir `-encoding UTF8` al campo de opciones adicionales del cargador en el cuadro de diálogo Exportar BD: Opciones avanzadas. De este modo, se asegura que se cargan correctamente los datos que no están en ASCII.

#### ***Datos de formato de fecha, hora y marca de tiempo***

En las propiedades de ruta, establezca el formato de fecha DD-MM-AAAA y el formato de hora HH:MM:SS.

#### ***Cadenas en blanco***

Las cadenas en blanco se exportan a la base de datos como valores nulos (NULL).

#### ***Orden de columnas diferente en la tabla de destino y la ruta al insertar datos en una tabla existente***

Si el orden de las columnas de la ruta es distinto al de la tabla de destino, los valores de datos se insertarán en columnas equivocadas. Utilice el nodo `Reorg. campos` para garantizar que el orden de las columnas de la ruta coincide con el orden de la tabla de destino. Si desea obtener más información, consulte el tema [Nodo Reorg. campos](#) en el capítulo 4 el p. 239.

#### ***Seguimiento del progreso de `nzload`***

Cuando ejecute SPSS Modeler en modo local, añada `-sts` al campo de opciones adicionales del cargador, en el cuadro de diálogo Exportar BD: Opciones avanzadas, para ver mensajes de estado cada 10.000 filas en la ventana de comando abierta por la utilidad `nzload`.

### ***Carga masiva de datos a bases de datos Oracle***

Los siguientes puntos pueden ayudarle a configurar la carga masiva de IBM® SPSS® Modeler a una base de datos Oracle utilizando la opción de cargador externo en el cuadro de diálogo Exportar BD: Opciones avanzadas.

**Asegúrese de que la utilidad *sqlldr* de Oracle está instalada**

El proceso *oracle\_loader.py* invoca la utilidad *sqlldr* de Oracle. Tenga en cuenta que *sqlldr* no se incluye automáticamente en el cliente de Oracle. Asegúrese de que *sqlldr* está instalado en el servidor en el que se va a ejecutar *oracle\_loader.py*.

**Especifique el SID o el nombre de servicio de la base de datos**

Si está exportando datos a un servidor de Oracle no local o su servidor de Oracle local tiene múltiples bases de datos, debe especificar lo siguiente en el campo de opciones adicionales del cargador, en el cuadro de diálogo Exportar BD: Opciones avanzadas, para pasar el SID o nombre de servicio:

-database <SID>

**Edición de la sección de configuración en *oracle\_loader.py***

En sistemas UNIX (opcional en Windows), edite la sección de configuración al inicio del proceso *oracle\_loader.py*. Aquí, los valores para las variables de entorno ORACLE\_SID, NLS\_LANG, TNS\_ADMIN y ORACLE\_HOME se pueden especificar, si procede, así como la ruta completa de la utilidad *sqlldr*.

**Datos de formato de fecha, hora y marca de tiempo**

En las propiedades de ruta, normalmente debe establecer el formato de fecha AAAA-MM-DD y el formato de hora HH:MM:SS.

Si necesita utilizar formatos de fecha y hora distintos a los anteriores, consulte la documentación de Oracle y edite el archivo de procesos *oracle\_loader.py*.

**Codificación de datos con caracteres que no son ASCII**

Si está realizando una carga masiva de datos que no están en formato ASCII, debe asegurarse de que la variable de entorno NLS\_LANG está configurada correctamente en su sistema. Esto es leído por la utilidad de cargador *sqlldr* de Oracle. Por ejemplo, el valor correcto de NLS\_LANG para Shift-JIS en Windows es Japanese\_Japan.JA16SJIS. Si necesita más información acerca de NLS\_LANG, consulte la documentación de Oracle.

**Cadenas en blanco**

Las cadenas en blanco se exportan a la base de datos como valores nulos (NULL).

**Carga masiva de datos a bases de datos SQL Server**

Los siguientes puntos pueden ayudarle a configurar la carga masiva de IBM® SPSS® Modeler a una base de datos SQL Server utilizando la opción de cargador externo en el cuadro de diálogo Exportar BD: Opciones avanzadas.

**Asegúrese de que la utilidad *bcp.exe* de SQL Server está instalada**

El proceso *mssql\_loader.py* invoca la utilidad *bcp.exe* de SQL Server. Asegúrese de que *bcp.exe* está instalado en el servidor donde se va a ejecutar *mssql\_loader.py*.

**Los espacios no sirven como delimitadores**

Evite elegir espacios como delimitadores en el cuadro de diálogo Exportar BD: Opciones avanzadas.

**Se recomienda seleccionar la opción *Comprobar el tamaño de la tabla***

Le recomendamos que active la opción *Comprobar tamaño de la tabla* en el cuadro de diálogo Exportar BD: Opciones avanzadas. No siempre se detectan los errores en el proceso de carga masiva, y si se activa esta opción, se hará una comprobación adicional para verificar que se ha cargado el número correcto de filas.

**Cadenas en blanco**

Las cadenas en blanco se exportan a la base de datos como valores nulos (NULL).

**Carga masiva de datos a bases de datos Teradata**

Los siguientes puntos pueden ayudarle a configurar la carga masiva de IBM® SPSS® Modeler a una base de datos Teradata utilizando la opción de cargador externo en el cuadro de diálogo Exportar BD: Opciones avanzadas.

**Asegúrese de que la utilidad *fastload* de Teradata está instalada**

El proceso *teradata\_loader.py* invoca la utilidad *fastload* de Teradata. Asegúrese de que *fastload* se ha instalado y configurado correctamente en el servidor en el que se ejecutará *teradata\_loader.py*.

**Solo es posible realizar cargas masivas de datos a tablas vacías**

Solo puede utilizar tablas vacías como destino de una carga masiva. Si una tabla de destino contiene datos previos a la carga masiva, la operación fallará.

**Datos de formato de fecha, hora y marca de tiempo**

En las propiedades de ruta, establezca el formato de fecha AAAA-MM-DD y el formato de hora HH:MM:SS.

**Cadenas en blanco**

Las cadenas en blanco se exportan a la base de datos como valores nulos (NULL).

### ***ID de procesos de Teradata (tdpid)***

Por defecto, *fastload* exporta los datos al sistema Teradata con `tdpid=dbc`. Normalmente, habrá una entrada en el archivo HOSTS que asocia `dbccop1` con la dirección IP del servidor de Teradata. Para utilizar un servidor diferente, especifique lo siguiente en el campo de opciones adicionales del cargador del cuadro de diálogo Exportar BD: Opciones avanzadas para pasar el `tdpid` de este servidor:

`-tdpid <id>`

### ***Espacios en nombres de columnas y tablas***

Si el nombre de una columna o tabla contiene espacios, fallará la operación de carga masiva. Si es posible, cambie el nombre de la columna o tabla eliminando los espacios.

## ***Desarrollo de programas de carga masiva***

Aquí se explica cómo desarrollar un programa cargador masivo que se pueda ejecutar desde IBM® SPSS® Modeler para cargar datos de un archivo de texto a una base de datos.

### ***Utilización de Python para crear programas cargadores masivos***

Por defecto, SPSS Modeler busca un programa cargador masivo por defecto basado en el tipo de base de datos. Consulte [Tabla 7-1](#) el p. 464.

Puede utilizar el proceso `test_loader.py` para ayudar al desarrollo de programas cargadores por lotes. Si desea obtener más información, consulte el tema [Prueba de programas de carga masiva](#) el p. 471.

### ***Objetos pasados al programa cargador masivo***

SPSS Modeler escribe dos archivos que se han pasado al programa cargador masivo.

- **Archivo de datos.** Contiene los datos que se tienen que cargar, en formato de texto.
- **Archivo de esquema.** Es un archivo XML que describe los nombres y tipos de las columnas, y proporciona información acerca del formato del archivo de datos (por ejemplo, qué carácter se emplea como delimitador entre campos).

Además, SPSS Modeler pasa otra información, como el nombre de la tabla, el nombre de usuario y la contraseña como argumentos al invocar el programa cargador masivo.

*Nota:* Para señalar la finalización satisfactoria a SPSS Modeler, el programa cargador masivo debería eliminar el archivo de esquema.

### ***Argumentos pasados al programa cargador masivo***

Los argumentos pasados al programa son los siguientes.

Tabla 7-2  
Argumentos pasados al cargador masivo

Argumento	Descripción
schemafilename	Ruta del archivo de esquema.
data file	Ruta del archivo de datos.
servername	Nombre de servidor de DBMS (puede estar en blanco).
databasename	Nombre de la base de datos dentro del servidor de DBMS (puede estar en blanco).
username	Nombre de usuario para acceder a la base de datos.
password	Contraseña para acceder a la base de datos.
tablename	Nombre de la tabla para cargar.
ownername	Nombre del propietario de la tabla (también se denomina nombre de esquema).
logfile	Nombre del archivo de registro (si está en blanco, no se crea ningún archivo de registro).
rowcount	Número de filas del conjunto de datos.

Cualquier opción especificada en el campo Opciones adicionales del cargador del cuadro de diálogo Exportar BD: Opciones avanzadas se pasa al programa cargador masivo tras estos argumentos estándar.

#### **Formato del archivo de datos**

Los datos se escriben en el archivo de datos con formato de texto; cada campo se separa con un carácter delimitador especificado en el cuadro de diálogo Exportar BD: Opciones avanzadas. A continuación se muestra un ejemplo del aspecto de un archivo delimitado por tabulaciones.

```
48 F HIGH NORMAL 0.692623 0.055369 drugA
15 M NORMAL HIGH 0.678247 0.040851 drugY
37 M HIGH NORMAL 0.538192 0.069780 drugA
35 F HIGH HIGH 0.635680 0.068481 drugA
```

El archivo se escribe con la codificación local utilizada por IBM® SPSS® Modeler Server (o SPSS Modeler si no está conectado a SPSS Modeler Server). Algunos formatos están controlados a través de la configuración de la ruta de SPSS Modeler.

#### **Formato del archivo de esquema**

El archivo de esquema es un archivo XML que describe el archivo de datos. A continuación se muestra un ejemplo que acompaña al archivo de datos previo.

```
<?xml version="1.0" encoding="UTF-8"?>
<DBSCHEMA version="1.0">
  <table delimiter="\t" commit_every="10000" date_format="YYYY-MM-DD" time_format="HH:MM:SS"
  append_existing="false" delete_datafile="false">
    <column name="Age" encoded_name="416765" type="integer"/>
    <column name="Sex" encoded_name="536578" type="char" size="1"/>
    <column name="BP" encoded_name="4250" type="char" size="6"/>
    <column name="Cholesterol" encoded_name="43686F6C65737465726F6C" type="char" size="6"/>
    <column name="Na" encoded_name="4E61" type="real"/>
    <column name="K" encoded_name="4B" type="real"/>
```

```
<column name="Drug" encoded_name="44727567" type="char" size="5"/>
</table>
</DBSCHEMA>
```

Las tablas siguientes enumeran los atributos de los elementos de <table> y <column> del archivo de esquema.

Tabla 7-3  
Atributos del elemento de <tabla>

Atributo	Descripción
delimiter	Carácter delimitador de campo (TAB se representa como \t).
commit_every	Intervalo de tamaño de lote (como en el cuadro de diálogo Exportar BD: Opciones avanzadas).
date_format	Formato usado para representar fechas.
time_format	Formato usado para representar horas.
append_existing	true si la tabla para cargar ya contiene datos; false en caso contrario.
delete_datafile	true si el programa de carga masiva debe eliminar el archivo de datos al completar la carga.

Tabla 7-4  
Atributos del elemento de <columna>

Atributo	Descripción
name	Nombre de columna.
encoded_name	El nombre de columna convertido a la misma codificación que el archivo de datos y la salida como una serie de números hexadecimales de dos dígitos.
type	Tipo de datos de la columna: uno de integer, real, char, time, date y datetime.
size	Para el tipo de datos char, la anchura máxima de la columna en caracteres.

### Prueba de programas de carga masiva

Puede probar la carga masiva utilizando el proceso de prueba *test\_loader.py* incluido en la carpeta *\scripts* del directorio de instalación de IBM® SPSS® Modeler. Resulta útil al intentar desarrollar, depurar o solucionar problemas con programas de carga masiva o procesos para usarlos con SPSS Modeler.

Para utilizar el proceso de prueba, siga estos pasos.

- ▶ Ejecute el proceso *test\_loader.py* para copiar los archivos de esquema y de datos a los archivos *schema.xml* y *data.txt*, y cree un archivo por lotes de Windows (*test.bat*).
- ▶ Edite el archivo *test.bat* para seleccionar el programa cargador masivo o el proceso que desea probar.
- ▶ Ejecute *test.bat* desde un shell de comandos para probar el programa cargador masivo o el proceso elegido.

*Nota:* Al ejecutar *test.bat* no se carga ningún dato en la base de datos.

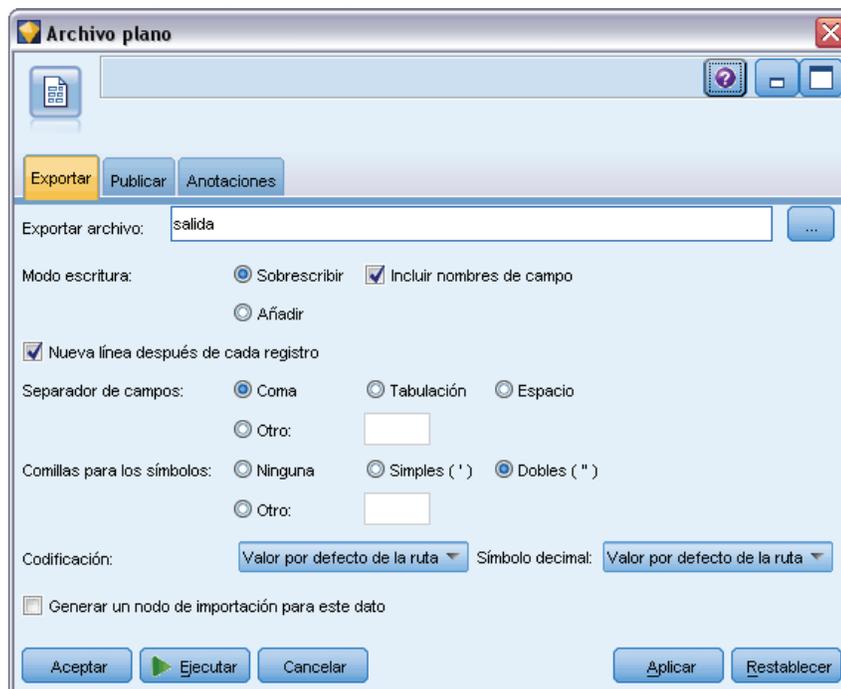
## ***nodo de exportación Archivo plano***

El nodo de exportación de archivo plano permite escribir datos en un archivo de texto delimitado. Esto es útil para exportar datos que se pueden leer con otro software de hoja de cálculo o de análisis.

*Nota:* No puede escribir archivos en el antiguo formato de caché porque IBM® SPSS® Modeler ya no lo utiliza para los archivos caché. Estos archivos de SPSS Modeler ahora se guardan en formato IBM® SPSS® Statistics *.sav*, que puede escribir utilizando un nodo Exportar Statistics. Si desea obtener más información, consulte el tema [Nodo Exportación de Statistics](#) en el capítulo 8 el p. 501.

### ***Pestaña Exportar del nodo Archivo plano***

Figura 7-7  
Nodo Archivo plano, pestaña Exportar



**Exportar archivo.** Especifica el nombre del archivo. Introduzca un nombre de archivo o pulse en el botón Buscar el archivo para examinar la ubicación del archivo.

**Modo escritura.** Si selecciona Sobrescribir, se sobrescribirá cualquier dato existente en el archivo especificado. Si selecciona Añadir, el resultado se añadirá al final del archivo existente, conservando cualquier dato que contenga.

- **Incluir nombres de campo.** Si selecciona esta opción, los nombres de los campos se escribirán en la primera línea del archivo de resultados. Esta opción sólo está disponible para el modo escritura Sobrescribir.

**Nueva línea después de cada registro.** Si selecciona esta opción, cada registro se escribirá en una línea nueva del archivo de resultados.

**Separador de campos.** Especifica el carácter que desee insertar entre los valores de campo del archivo de texto generado. Las opciones son Coma, Tabulador, Espacio y Otro. Si selecciona Otro, introduzca los caracteres delimitadores que desee en el cuadro de texto.

**Comillas para los símbolos.** Especifica el tipo de comillas que vaya a utilizar para los valores de campos simbólicos. Las opciones son Ninguna (los valores no irán entre comillas), Simple ('), Doble (") y Otro. Si selecciona Otro, introduzca los caracteres de entrecomillado que desee en el cuadro de texto.

**Codificación.** Especifica el método de codificación de textos utilizado. Puede elegir entre valor por defecto del sistema, el valor por defecto de la ruta o UTF-8.

- El valor por defecto del sistema se especifica en el Panel de control de Windows o, si lo ejecuta en modo distribuido, en el equipo servidor.
- La ruta por defecto se especifica en el cuadro de diálogo Propiedades de ruta.

**Símbolo decimal.** Especifica cómo deben representarse los decimales en los datos.

- **Valor por defecto de la ruta.** Se utilizará el separador decimal definido por la configuración del valor por defecto de la ruta actual. Este será normalmente el separador decimal definido por la configuración regional del ordenador.
- **Punto (.).** Se utilizará el punto como separador decimal.
- **Coma (,).** Se utilizará la coma como separador decimal.

**Generar un nodo de importación para este dato.** Seleccione esta opción para generar automáticamente un nodo de fuente Archivo var. que leerá el archivo de datos exportado. Si desea obtener más información, consulte el tema [Nodo Archivo variable](#) en el capítulo 2 el p. 25.

## ***Nodos de exportación IBM SPSS Data Collection***

El nodo de exportación IBM® SPSS® Data Collection guarda datos en el formato utilizado por el software de investigación de mercados Data Collection, en función del modelo de datos de Data Collection. Este formato distingue los datos de casos, las respuestas reales a las preguntas recogidas durante una encuesta, de los metadatos que describen cómo se recopilan y organizan los datos de casos. Los metadatos constan de información como textos de preguntas, nombres de variables y descripciones, conjuntos de respuestas múltiples, traducciones de los diversos textos y las definiciones de la estructura de los datos de casos. Si desea obtener más información, consulte el tema [Nodo Data Collection](#) en el capítulo 2 el p. 35.

Figura 7-8  
 Nodo Exportar IBM SPSS Data Collection, pestaña Exportar



*Nota:* este nodo requiere tener instalado Data Collection Data Model versión 4.0 o posterior, que se distribuye con los productos de software de Data Collection. Si desea obtener más información, consulte la página Web de Data Collection en <http://www.ibm.com/software/analytics/spss/products/data-collection/>. Además de la instalación de Modelo de datos, no se requieren configuraciones adicionales.

**Archivo de metadatos.** Especifica el nombre del archivo de definición de cuestionario (*.mdd*) donde se guardarán los metadatos exportados. Se crea un cuestionario por defecto basándose en la información sobre el tipo de campo. Por ejemplo, un campo nominal (conjunto) podría representarse como una única pregunta con la descripción de campo utilizada como el texto de la pregunta y una casilla de verificación independiente para cada valor definido.

**Fundir metadatos.** Especifica si los metadatos sobrescribirán las versiones existentes o se combinarán con los metadatos existentes. Si se selecciona la opción de combinación, se creará una nueva versión cada vez que se ejecute la ruta. Esto permite realizar un seguimiento de las versiones de un cuestionario a medida que se va modificando. Cada versión puede considerarse como una instantánea de los metadatos que se utilizan para recopilar un conjunto de datos de casos determinado.

**Activar variables de sistema.** Especifica si las variables de sistema se incluyen en el archivo *.mdd* exportado. Entre ellas se incluyen variables como *Respondent.Serial*, *Respondent.Origin* y *DataCollection.StartTime*.

**Configuración de datos de casos.** Especifica el archivo de datos IBM® SPSS® Statistics (*.sav*) donde se exportan los datos de casos. Tenga en cuenta que todas las restricciones sobre los nombres de variables y valores son aplicables, así que, por ejemplo, puede que necesite cambiar

a la pestaña Filtro y utilizar la opción “Cambiar nombre para SPSS Statistics” en el menú de opciones Filtro para corregir los caracteres que no son válidos en los nombres de campo.

**Generar un nodo de importación para este dato.** Seleccione esta opción para generar automáticamente un nodo de fuente Data Collection que leerá el archivo de datos exportado.

**Conjuntos de respuestas múltiples.** Cualquier conjunto de respuestas múltiples definido en la ruta se conservará automáticamente cuando se exporte el archivo. Puede definir o editar conjuntos de respuestas múltiples en un nodo con una pestaña Filtro. Si desea obtener más información, consulte el tema [Edición de conjuntos de respuestas múltiples](#) en el capítulo 4 el p. 158.

## ***Nodo Exportar de IBM Cognos BI***

El nodo Exportar de IBM Cognos BI le permite exportar datos desde una ruta de IBM® SPSS® Modeler a Cognos BI, en formato UTF-8. De este modo, Cognos BI puede utilizar los datos transformados o puntuados desde SPSS Modeler. Por ejemplo, podría utilizar Cognos BI Report Studio para crear un informe basado en los datos exportados, incluidas las predicciones y los valores de confianza. El informe luego puede guardarse en el servidor de Cognos BI y distribuirse a usuarios de Cognos BI.

*Nota:* Solamente puede exportar datos relacionales, no datos OLAP.

Para exportar datos a Cognos BI, debe especificar lo siguiente:

- Conexión de Cognos: la conexión al servidor de Cognos BI
- Conexión ODBC: la conexión al servidor de datos de Cognos que utiliza el servidor de Cognos BI

En la conexión Cognos especifica una fuente de datos de Cognos que se utilizará. Esta fuente de datos debe utilizar el mismo inicio de sesión que la fuente de datos ODBC.

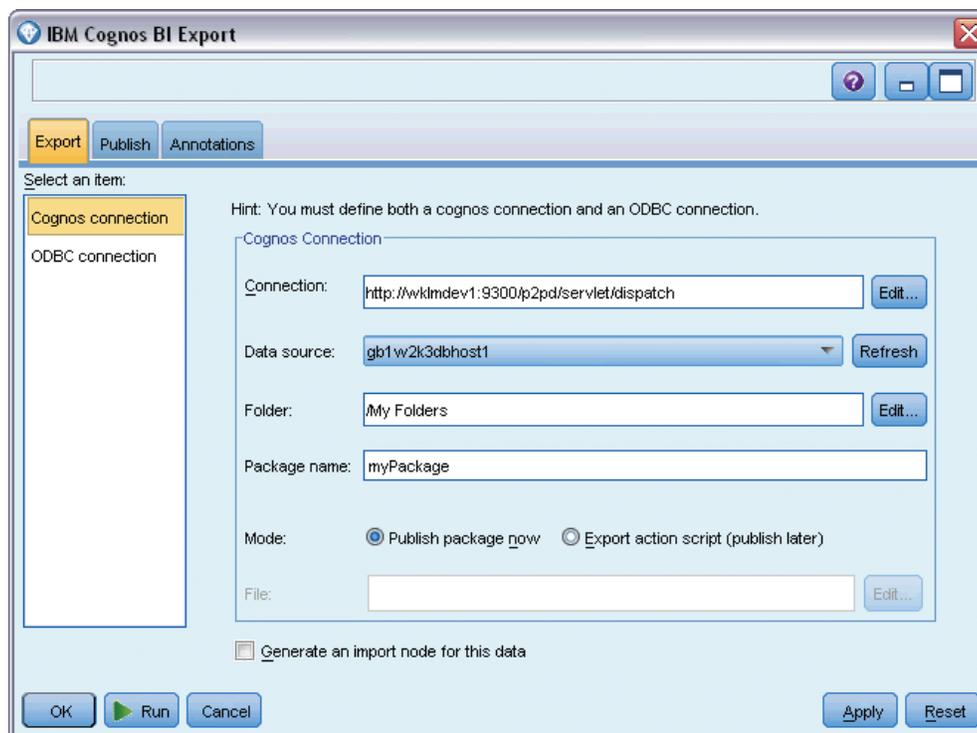
Exporta los datos de ruta reales al servidor de datos y los metadatos de paquete al servidor de Cognos BI.

Al igual que con cualquier otro nodo de exportación, también puede utilizar la pestaña Publicar del cuadro de diálogo del nodo para publicar la ruta para su distribución mediante IBM® SPSS® Modeler Solution Publisher.

### ***Conexión de Cognos***

Aquí es donde se especifica la conexión al servidor de Cognos BI que quiera utilizar para la exportación. El procedimiento implica la exportación de los metadatos a un nuevo paquete del servidor de Cognos BI, mientras que los datos de ruta se exportan al servidor de datos de Cognos.

Figura 7-9  
Exportación de datos Cognos



**Conexión.** Pulse en el botón Editar para mostrar un cuadro de diálogo donde puede definir la URL y otros detalles del servidor de Cognos BI al que quiere exportar los datos. Si ya ha iniciado sesión en un servidor de Cognos BI mediante IBM® SPSS® Modeler, también puede editar los detalles de la conexión actual. Si desea obtener más información, consulte el tema [Conexiones de Cognos](#) en el capítulo 2 el p. 49.

**Fuente de datos.** El nombre de la fuente de datos de Cognos (normalmente una base de datos) a la que está exportando los datos. La lista desplegable muestra todas las fuentes de datos de Cognos a los que puede acceder desde la conexión actual. Pulse en el botón Actualizar para actualizar la lista.

**Carpeta.** La ruta y el nombre de la carpeta del servidor de Cognos BI donde se creará el paquete de exportación.

**Nombre del paquete.** El nombre del paquete de la carpeta especificada que debe contener los metadatos exportados. Éste debe ser un nuevo paquete con un único asunto de consulta; no puede exportar a un paquete existente.

**Moda.** Especifica cómo quiere realizar la exportación:

- **Publicar paquete ahora.** (valor por defecto) Realiza la operación de exportación tan pronto como pulsa en Ejecutar.
- **Exportar proceso de acción.** Crea un proceso XML que puede ejecutar más tarde (por ejemplo, utilizando Framework Manager) para realizar la exportación. Escriba la ruta y el nombre de archivo del proceso en el campo Archivo o utilice el botón Editar para especificar el nombre y la ubicación del archivo de proceso.

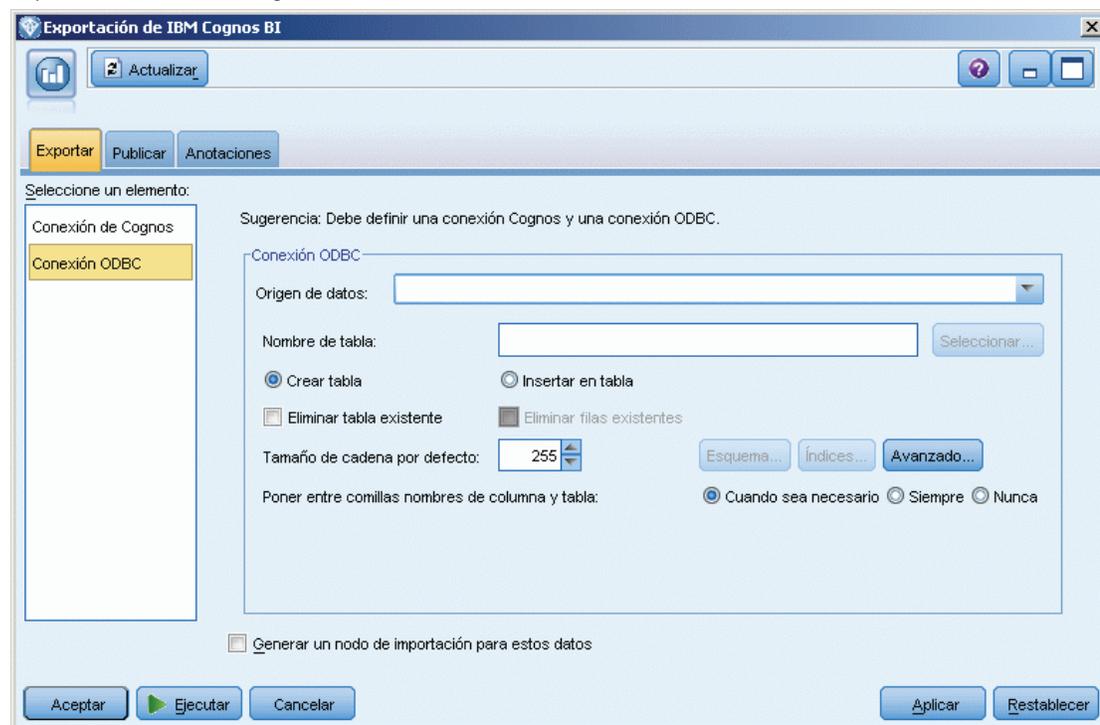
**Generar un nodo de importación para este dato.** Seleccione esta opción para generar un nodo de fuente para los datos tal y como se exportan en la tabla y la fuente de datos especificadas. Cuando pulsa en Ejecutar, este nodo se añade al lienzo de rutas.

## conexión ODBC

Aquí especifica la conexión al servidor de datos de Cognos (es decir, la base de datos) al que deben exportarse los datos de ruta.

*Nota:* Debe asegurarse de que la fuente de datos que especifique aquí apunta a la misma que se ha especificado en el panel Conexiones de Cognos. También debe asegurarse de que la fuente de datos de conexión Cognos utiliza el mismo inicio de sesión que la fuente de datos ODBC.

Figura 7-10  
Exportación de datos Cognos



**Origen de datos.** Muestra el origen de datos seleccionado. Introduzca el nombre o selecciónelo de la lista desplegable. Si no ve en la lista la base de datos que desea, seleccione Añadir nueva conexión de base de datos y busque la base de datos desde el cuadro de diálogo Conexiones a la base de datos. Si desea obtener más información, consulte el tema [Adición de conexión a la base de datos](#) en el capítulo 2 el p. 18.

**Nombre de tabla.** Introduzca el nombre de la tabla a la que desee enviar los datos. Si selecciona la opción Insertar en tabla, puede seleccionar una tabla existente de la base de datos pulsando en el botón Seleccionar.

**Crear tabla.** Seleccione esta opción para crear una nueva tabla de base de datos o para sobrescribir una tabla de base de datos existente.

**Insertar en tabla.** Seleccione esta opción para insertar los datos como nuevas filas en una tabla de base de datos existente.

**Tabla Fusión.** (Donde esté disponible) Seleccione esta opción para actualizar columnas de base de datos seleccionadas con valores de los campos de datos de origen correspondiente. Si selecciona esta opción activará el botón Fusión, que muestra un cuadro de diálogo donde podrá asignar campos de orígenes de datos a columnas de base de datos.

**Eliminar tabla existente.** Seleccione esta opción para eliminar cualquier tabla existente con el mismo nombre al crear una tabla nueva.

**Eliminar filas existentes.** Seleccione esta opción para eliminar filas existentes de la tabla antes de exportar cuando se inserte en una tabla.

*Nota:* si está seleccionada alguna de las dos opciones anteriores, recibirá el mensaje Advertencia de sobrescritura cuando ejecute el nodo. Para suprimir las advertencias, anule la selección de Avisar cuando un nodo sobrescriba una tabla de la base de datos en la pestaña Notificaciones del cuadro de diálogo Opciones de usuario.

**Tamaño de cadena por defecto.** Los campos marcados como sin tipo en un nodo Tipo anterior de la ruta se escriben en la base de datos como campos de cadena. Especifique el tamaño de las cadenas que se van a utilizar para campos sin tipo.

Pulse en Esquema para abrir un cuadro de diálogo donde podrá establecer diversas opciones de exportación (para bases de datos que admitan esta función), establecer los tipos de datos de SQL para los campos y especificar la clave primaria para la indización de la base de datos. Si desea obtener más información, consulte el tema [Opciones de Esquema de exportación de base de datos](#) el p. 455.

Pulse en Índices para especificar las opciones de indización de la tabla exportada para mejorar el rendimiento de la base de datos. Si desea obtener más información, consulte el tema [Opciones de índice de exportación de base de datos](#) el p. 459.

Pulse en Avanzado para especificar opciones de carga masiva y de confirmación de base de datos. Si desea obtener más información, consulte el tema [Opciones avanzadas de exportación de base de datos](#) el p. 461.

**Poner entre comillas nombres de columna y tabla.** Seleccione opciones utilizadas al enviar una instrucción CREATE TABLE a la base de datos. Las tablas o columnas con espacios o caracteres que no sean estándar deben ponerse entre comillas.

- **Cuando sea necesario.** Selecciónelo para permitir que IBM® SPSS® Modeler determine automáticamente cuándo son necesarias las comillas de forma individual.
- **Siempre.** Para poner siempre entre comillas los nombres de tabla y columna.
- **Nunca.** Para desactivar la utilización de comillas.

**Generar un nodo de importación para este dato.** Seleccione esta opción para generar un nodo de fuente para los datos tal y como se exportan en la tabla y la fuente de datos especificadas. Cuando pulsa en Ejecutar, este nodo se añade al lienzo de rutas.

## Nodo Exportar SAS

*Nota:* Esta función está disponible en SPSS Modeler Professional y SPSS Modeler Premium.

El nodo Exportar SAS permite escribir datos en formato SAS para leerlos en SAS o en un paquete de software compatible con SAS. Puede exportar en tres formatos de archivo SAS: SAS para Windows/OS2, SAS para UNIX o SAS versión 7/8.

### Pestaña Exportar del nodo Exportar SAS

Figura 7-11  
Nodo Exportar SAS, pestaña Exportar



**Exportar archivo.** Especifique el nombre del archivo. Introduzca un nombre de archivo o pulse en el botón Buscar el archivo para examinar la ubicación del archivo.

**Exportar.** Especifique el formato del archivo de exportación. Las opciones son SAS para Windows/OS2, SAS para UNIX o SAS versión 7/8.

**Exportar nombres de campo como variables.** Seleccione opciones para exportar nombres de campos y etiquetas desde IBM® SPSS® Modeler con objeto de utilizarlos con SAS.

- **nombres.** Para exportar los nombres y las etiquetas de campo de SPSS Modeler. Los nombres se exportan como nombres de variables de SAS, mientras que las etiquetas se exportan como etiquetas de variables de SAS.
- **etiquetas.** Seleccione esta opción para utilizar los nombres de campo de SPSS Modeler como etiquetas de variable en SAS. SPSS Modeler admite los caracteres de nombres de campo que no son válidos en nombres de variables de SAS. Para evitar la posibilidad de crear nombres SAS no válidos, seleccione Nombres en su lugar.

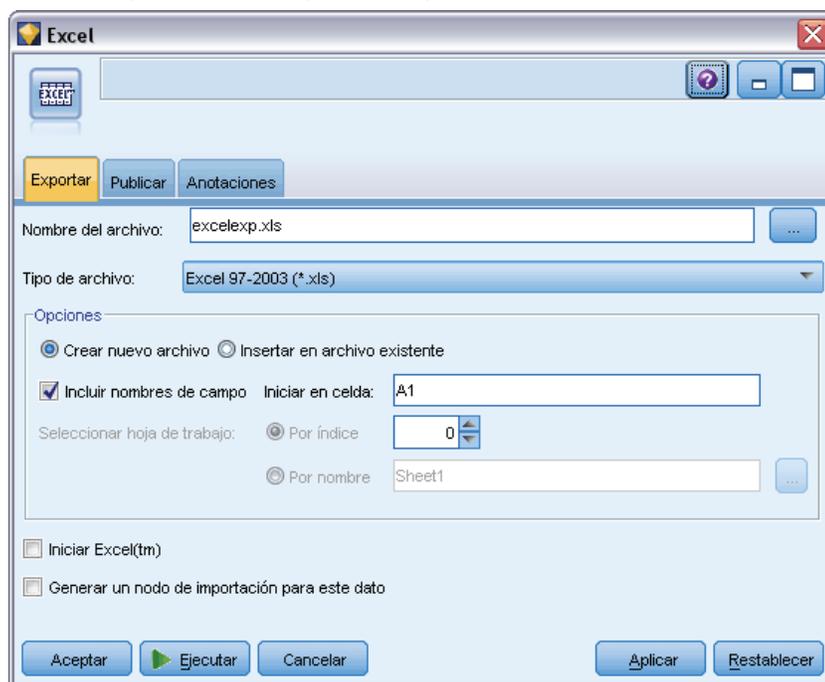
**Generar un nodo de importación para este dato.** Seleccione esta opción para generar automáticamente un nodo de fuente SAS que leerá el archivo de datos exportado. Si desea obtener más información, consulte el tema [Nodo de fuente SAS](#) en el capítulo 2 el p. 51.

## Nodo de exportación Excel

El nodo de exportación Excel ofrece los datos resultantes en formato de Microsoft Excel (.xls). Si lo desea, puede elegir iniciar automáticamente Excel y abrir el archivo exportado cuando se ejecute el nodo.

### Pestaña Exportar del nodo Excel

Figura 7-12  
Nodo de exportación Excel, pestaña Exportar



**Nombre del archivo.** Escriba el nombre del archivo o pulse en el botón de selección de archivos para acceder a la ubicación del archivo. El nombre de archivo por defecto es *excelexp.xls*.

**Tipo de archivo.** Seleccione el tipo de archivo de Excel que está exportando.

**Crear archivo nuevo.** Crea un nuevo archivo de Excel.

**Insertar en el archivo existente.** El contenido se sustituye comenzando en la casilla que designa el campo Iniciar en celda. El resto de casillas de la hoja de cálculo se quedan con su contenido original.

**Incluir nombres de campo.** Especifica si los nombres de campos deben incluirse en la primera fila de la hoja de trabajo.

**Iniciar en celda.** La ubicación de la casilla se utiliza para el primer registro de exportación (o el nombre del primer campo si Incluir nombres de campo está seleccionado). Los datos se complimentan a la derecha y abajo desde la casilla inicial.

**Seleccionar hoja de trabajo.** Especifica la hoja de cálculo en la que desea exportar los datos. Puede identificar la hoja de datos por índice o por nombre:

- **Por índice.** Si crea un nuevo archivo, especifique un número entre 0 y 9 para identificar la hoja de cálculo a la que desea exportar, comenzando por 0 para la primera hoja de cálculo, 1 para la segunda, etcétera. Puede utilizar los valores de 10 o más si ya existe una única hoja de cálculo en esta posición.
- **Por nombre.** Si crea un nuevo archivo, especifique el nombre que se utiliza para la hoja de cálculo. Si lo inserta en un archivo existente, los datos se insertan en esta hoja de cálculo si existe, de lo contrario se crea una nueva hoja de cálculo con este nombre.

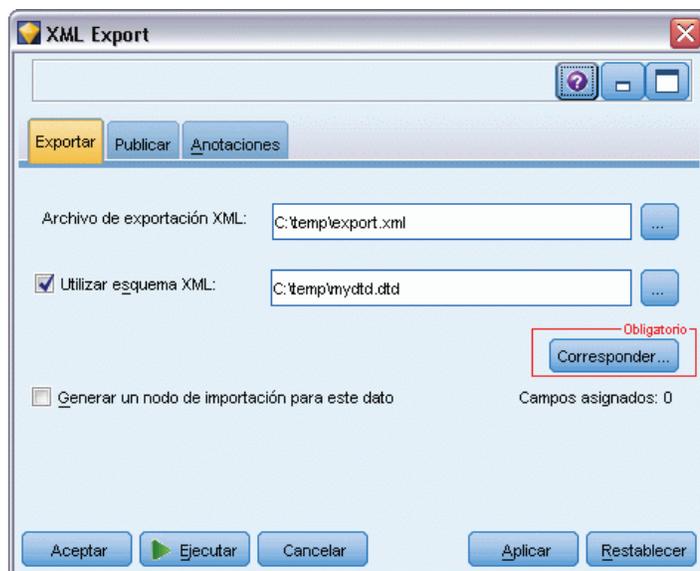
**Iniciar Excel.** Especifica si Excel se inicia automáticamente en el archivo exportado cuando se ejecuta el nodo. Tenga en cuenta que al trabajar en modo distribuido respecto a IBM® SPSS® Modeler Server, el resultado se guarda en el sistema de archivos del servidor y Excel se inicia en el cliente con una copia del archivo exportado.

**Generar un nodo de importación para este dato.** Seleccione esta opción para generar automáticamente un nodo de fuente Excel que leerá el archivo de datos exportado. Si desea obtener más información, consulte el tema [Nodo de fuente de Excel](#) en el capítulo 2 el p. 53.

## ***Nodo de exportación XML***

El nodo Exportar XML permite obtener datos de resultados en formato XML, utilizando codificación UTF-8. También puede crear un nodo de fuente XML para leer los datos exportados a la ruta.

Figura 7-13  
Exportación de datos XML



**Archivo de exportación XML.** La ruta completa y el nombre del archivo XML al que desea exportar los datos.

**Utilizar esquema XML.** Seleccione esta casilla de verificación si desea utilizar un esquema o DTD para controlar la estructura de los datos exportados. Al hacerlo se activa el botón Mapa que se describe a continuación.

Si no utiliza un esquema o DTD, se utiliza la siguiente estructura para los datos exportados:

```
<registros>
  <registro>
    <nombrecampo1>value</nombrecampo1>
    <nombrecampo2>value</nombrecampo2>
    :
    <nombrecampoN>value</nombrecampoN>
  </registro>
  <registro>
    :
    :
  </registro>
  :
  :
</registros>
```

Los espacios en un nombre de campo se sustituyen por guiones bajos; por ejemplo, “Mi campo” es <My\_Field>.

**Corresponder.** Si ha optado por utilizar un esquema XML, este botón abre un cuadro de diálogo donde puede especificar la parte de la estructura XML que se debe utilizar para iniciar cada nuevo registro. Si desea obtener más información, consulte el tema [Opciones de registros de correspondencia XML](#) el p. 483.

**Campos de correspondencia.** Indica el número de campos que se han asignado.

**Generar un nodo de importación para este dato.** Seleccione esta opción para generar automáticamente un nodo de fuente de XML que leerá el archivo de datos exportado en la ruta. Si desea obtener más información, consulte el tema [Nodo Fuente XML](#) en el capítulo 2 el p. 54.

## ***Escritura de datos XML***

Si se especifica un elemento XML, el valor del campo se coloca en la pestaña del elemento:

```
<element>value</element>
```

Cuando se corresponde un atributo, el valor del campo se coloca como el valor del atributo:

```
<element atributo="value">
```

Si un campo se corresponde con un elemento por encima del elemento <records> el campo solamente se asigna una vez y será una constante para todos los registros. El valor de este elemento proviene del primer registro.

Si se va a escribir un valor nulo, se realiza especificando contenido vacío. En elementos es:

```
<element></element>
```

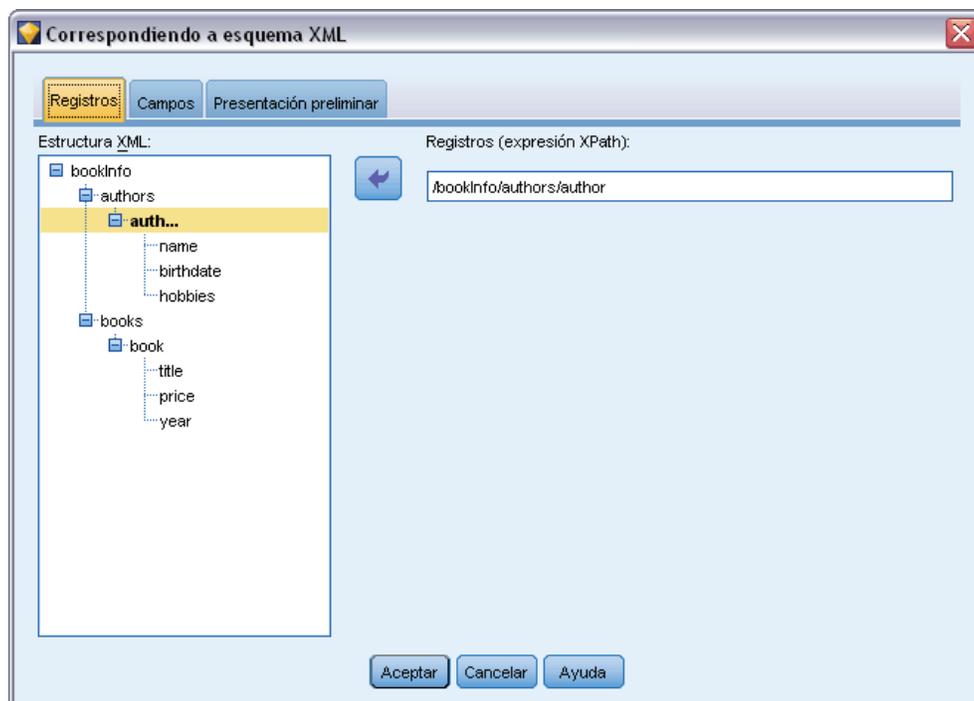
En atributos, es:

```
<element atributo="">
```

## ***Opciones de registros de correspondencia XML***

La pestaña Registros permite especificar la parte de la estructura XML que se utilizará para iniciar cada nuevo registro. Para corresponder correctamente un esquema, necesita especificar el delimitador del registro.

Figura 7-14  
Registros de correspondencia XML



**Estructura XML.** Un árbol jerárquico que muestra la estructura del esquema XML especificado en la pantalla anterior.

**Registros (expresión XPath).** Para definir el delimitador del registro, seleccione un elemento de la estructura XML y pulse en el botón de flecha derecha. Cada vez que este elemento se encuentra en los datos de fuente, se crea un nuevo registro en el archivo de resultados.

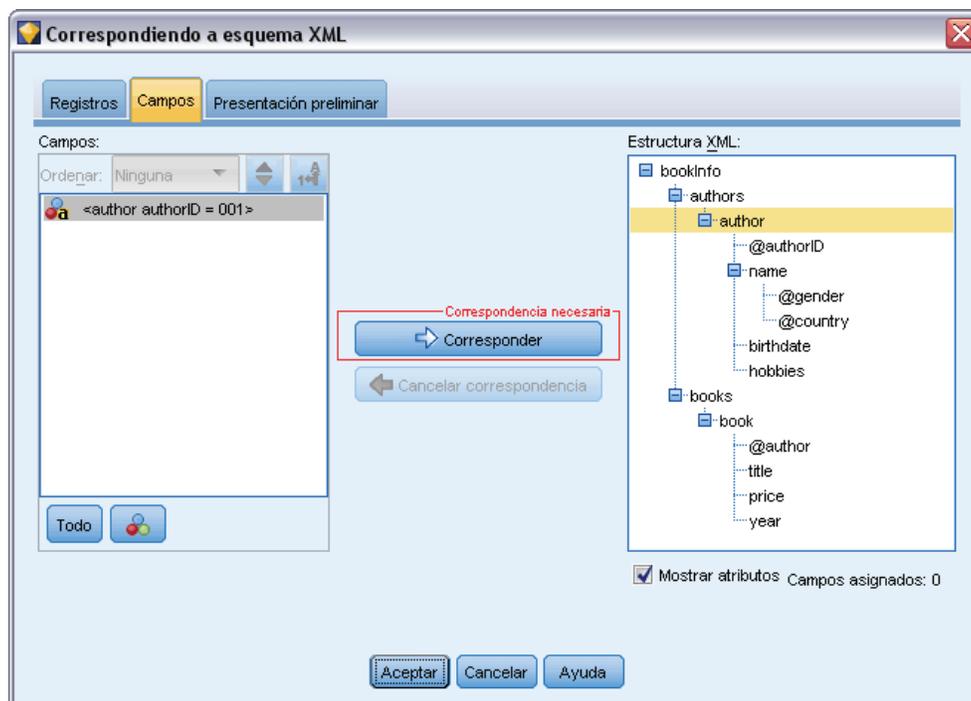
*Nota:* Si selecciona el elemento de raíz en la estructura XML, solamente se puede escribir un registro y el resto se omiten.

### ***Opciones de campos de correspondencia XML***

La pestaña Campos se utiliza para corresponder campos en el conjunto de datos con elementos o atributos en la estructura XML cuando se utiliza un archivo de esquema.

Los nombres de campos que coinciden con un elemento o atributo se corresponden automáticamente mientras el nombre del elemento o atributo sea único. Además, si hay un elemento y un atributo con el nombre field1, no se produce la correspondencia automática. Si solamente hay un elemento en la estructura denominado field1, se corresponde automáticamente en la ruta un campo con ese nombre.

Figura 7-15  
Campos de correspondencia XML



**Campos.** La lista de campos del modelo. Seleccione uno o más campos como parte de la fuente de la asignación. Puede utilizar estos botones en la parte inferior de la lista para seleccionar todos los campos o todos los campos con un nivel de medición concreto.

**Estructura XML.** Seleccione un elemento en la estructura XML como el destino de la correspondencia. Para crear la asignación, pulse en **Corresponder**. Se muestra la correspondencia. El número de campos que se han asignado de esta forma se muestran bajo la lista.

Para eliminar una correspondencia, seleccione el elemento en la lista de estructuras XML y pulse en **Cancelar correspondencia**.

**Mostrar atributos.** Muestra u oculta los atributos, si los tiene, de los elementos en la estructura.

### ***Presentación preliminar de correspondencia XML***

En la pestaña **Presentación preliminar**, pulse en **Actualizar** para ver una vista previa del XML que se escribirá.

Si la correspondencia es incorrecta, vuelva a la pestaña **Registros** o **Campos** para corregir los errores y vuelva a pulsar en **Actualizar** para ver el resultado.

# Nodos de IBM SPSS Statistics

## Conceptos básicos de nodos de IBM SPSS Statistics

Para complementar IBM® SPSS® Modeler y su capacidad de minería de datos, IBM® SPSS® Statistics le ofrece la capacidad de realizar análisis estadístico y gestión de datos más profundos.

Si tiene una copia compatible con licencia de SPSS Statistics instalada, puede conectarla desde SPSS Modeler y realizar una manipulación y análisis complejos y en pasos múltiples de datos que de otro modo no sería posible con SPSS Modeler. Para usuarios avanzados también existe la opción de modificar más los análisis utilizando sintaxis de comandos. Consulte las notas de publicación para obtener información sobre compatibilidad de la versión.

Si están disponibles, los nodos de SPSS Statistics se muestran en una parte exclusiva de la paleta de nodos.

*Nota:* Recomendamos que instancie sus datos en un nodo Tipo antes de utilizar los nodos Transformar, Modelo o Resultado de SPSS Statistics. Esto también es un requisito cuando se utiliza el comando de sintaxis AUTORECODE.

La paleta de SPSS Statistics contiene los siguientes nodos:



El nodo Archivo Statistics lee los datos desde un formato de archivo *.sav* que utiliza SPSS Statistics y archivos caché guardados en SPSS Modeler, que también puede utilizar el mismo formato. Si desea obtener más información, consulte el tema [Nodo de archivos estadísticos](#) el p. 487.



El nodo Transformación Statistics ejecuta una selección de comandos de sintaxis de SPSS Statistics en los orígenes de datos de SPSS Modeler. Este nodo requiere una copia de SPSS Statistics con licencia. Si desea obtener más información, consulte el tema [Nodo Transformación de Statistics](#) el p. 489.



El nodo Modelo Statistics permite analizar y trabajar con sus datos ejecutando los procedimientos de SPSS Statistics que producen PMML. Este nodo requiere una copia de SPSS Statistics con licencia. Si desea obtener más información, consulte el tema [Nodo Modelo de Statistics](#) el p. 493.



El nodo Resultados de Statistics le permite llamar a un procedimiento de SPSS Statistics para analizar los datos de SPSS Modeler. A wide variety of SPSS Statistics analytical procedures is available. This node requires a licensed copy of SPSS Statistics. Si desea obtener más información, consulte el tema [Nodo Resultados de Statistics](#) el p. 497.



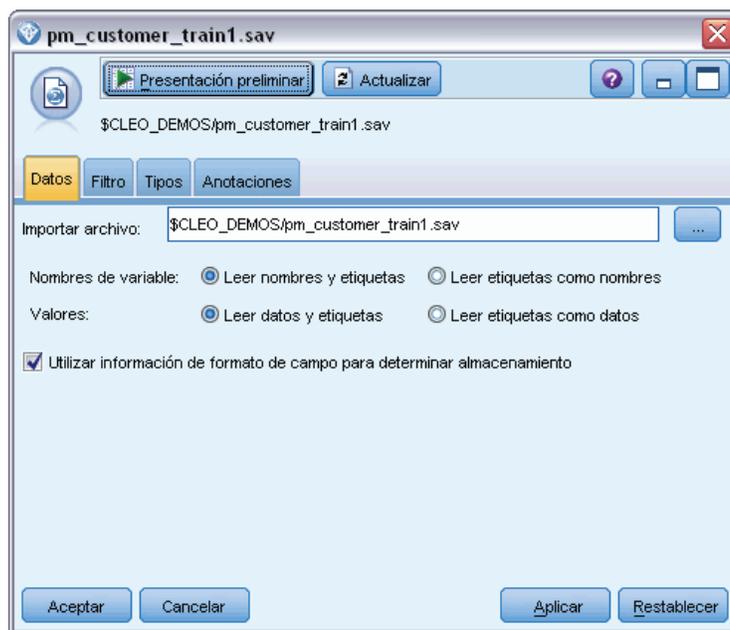
El nodo Exportar Statistics ofrece los resultados en formato SPSS Statistics.sav *de*. Los archivos .sav se pueden leer con SPSS Statistics Base y otros productos. Este es también el formato utilizado para los archivos caché de SPSS Modeler. Si desea obtener más información, consulte el tema [Nodo Exportación de Statistics](#) el p. 501.

*Nota:* Si su copia de SPSS Statistics solamente tiene licencia para un único usuario y ejecuta una ruta con dos o más ramas, las cuales contienen un nodo de SPSS Statistics cada una, puede que reciba un error de licencia. Esto se produce cuando la sesión de SPSS Statistics para una rama no se ha terminado antes de que la sesión de otra rama intenta comenzar. Si es posible, rediseñe la ruta para que varias ramas con nodos de SPSS Statistics no se ejecuten en paralelo.

## Nodo de archivos estadísticos

Puede utilizar el nodo Archivo Statistics para leer los datos directamente desde un archivo de IBM® SPSS® Statistics guardado (.sav). Este formato se utiliza ahora para reemplazar el archivo de caché de las versiones anteriores de IBM® SPSS® Modeler. Si desea importar un archivo de caché guardado, debe utilizar el nodo Archivo de SPSS Statistics.

Figura 8-1  
Importación de un archivo .sav



**Importar archivo.** Especifique el nombre del archivo. Puede introducir un nombre de archivo o pulsar en el botón de puntos suspensivos (...) para seleccionar un archivo. La ruta de archivo se mostrará una vez haya seleccionado un archivo.

**Nombres de variable.** Seleccione un método para gestionar etiquetas y nombres de variables en la importación desde un archivo de SPSS Statistics.sav. Los metadatos seleccionados para incluirlos aquí permanecen durante el trabajo en SPSS Modeler y se pueden volver a exportar para su uso en SPSS Statistics.

- **Leer nombres y etiquetas.** Seleccione para leer las etiquetas y nombres de variables en SPSS Modeler. Esta opción está seleccionada por defecto y los nombres de variables se muestran en el nodo Tipo. Las etiquetas se pueden mostrar en gráficos, exploradores de modelos y otros tipos de resultados, dependiendo de las opciones especificadas en el cuadro de diálogo de propiedades de la ruta. La representación de etiquetas en los resultados está desactivada por defecto.
- **Leer etiquetas como nombres.** Seleccione para leer las etiquetas descriptivas de variables desde el archivo de SPSS Statistics *.sav* en lugar de los nombres abreviados de campos y utilice estas etiquetas como nombres de variables en SPSS Modeler.

**Valores.** Seleccione un método para gestionar etiquetas y variables en la importación desde un archivo de SPSS Statistics *.sav*. Los metadatos seleccionados para incluirlos aquí permanecen durante el trabajo en SPSS Modeler y se pueden volver a exportar para su uso en SPSS Statistics.

- **Leer datos y etiquetas.** Seleccione para leer valores reales y etiquetas de valor en SPSS Modeler. Esta opción está seleccionada por defecto y los propios valores se muestran en el nodo Tipo. Las etiquetas de valor se pueden mostrar en el generador de expresiones, gráficos, exploradores de modelos y otros tipos de resultados, dependiendo de las opciones especificadas en el cuadro de diálogo de propiedades de la ruta.
- **Leer etiquetas como datos.** Indique si desea utilizar las etiquetas de valor del archivo *.sav* en lugar de los códigos numéricos o simbólicos utilizados para representar los valores. Por ejemplo, si selecciona esta opción para datos con un campo de sexo cuyos valores 1 y 2 representan realmente *masculino* y *femenino* respectivamente, convertirá el campo en una cadena e importará *masculino* y *femenino* como los valores reales.

Es importante tener en cuenta los valores perdidos en los datos de SPSS Statistics antes de seleccionar esta opción. Por ejemplo, si un campo numérico utiliza etiquetas sólo para valores perdidos (0 = *Sin respuesta*, -99 = *Desconocido*), al seleccionar la opción anterior se importarán sólo las etiquetas de valor *Sin respuesta* y *Desconocido* y el campo se convertirá en una cadena. En estos casos, debe importar los propios valores y establecer los valores perdidos en un nodo Tipo.

**Utilice la información de formato de campo para determinar el almacenamiento.** Si esta casilla está marcada, los valores de campo que tienen el formato en el archivo *.sav* como enteros (por ej. los campos especificados como *F<sub>n</sub>.0* en la Vista de variables en SPSS Statistics) se importan utilizando el almacenamiento de enteros. El resto de los valores de campos excepto las cadenas se importan como números reales.

Si esta casilla no está marcada (por defecto), todos los valores de campos excepto las cadenas se importan como números reales, ya tengan formato en el archivo *.sav* como enteros o no.

**Conjuntos de respuestas múltiples.** Cualquier conjunto de respuestas múltiples definido en el archivo SPSS Statistics se conservará cuando se importe el archivo. Puede definir o editar conjuntos de respuestas múltiples en un nodo con una pestaña Filtro. Si desea obtener más información, consulte el tema [Edición de conjuntos de respuestas múltiples](#) en el capítulo 4 el p. 158.

## **Nodo Transformación de Statistics**

El nodo Transformación de Statistics permite llevar a cabo transformaciones de datos utilizando la sintaxis de comandos de IBM® SPSS® Statistics, lo que permite llevar a cabo las transformaciones que no admite IBM® SPSS® Modeler y automatizar transformaciones complejas que requieren varios pasos, incluida la creación de varios campos desde un único nodo. Es similar al nodo Resultados de Statistics, la diferencia es que los datos se devuelven a SPSS Modeler para su posterior análisis, mientras que en el nodo Resultados los datos se devuelven como objetos de resultados solicitados (por ejemplo, gráficos o tablas).

Debe tener una versión compatible de SPSS Statistics instalada en su equipo y con licencia para poder usar este nodo. Si desea obtener más información, consulte el tema [Complementos de IBM SPSS Statistics](#) en el capítulo 6 el p. 448. Consulte las notas de publicación para obtener información sobre compatibilidad.

Si es necesario, puede utilizar la ficha Filtro para filtrar los campos o cambiar su nombre para que cumplan los estándares de nomenclatura de SPSS Statistics. Si desea obtener más información, consulte el tema [Cambio de nombre o filtrado de campos para IBM SPSS Statistics](#) el p. 503.

**Referencia de sintaxis.** Para obtener detalles sobre procedimientos SPSS Statistics específicos, consulte la guía *Referencia de sintaxis de comandos de SPSS Statistics*, que se incluye con su copia del software SPSS Statistics. Para ver la guía de la pestaña Sintaxis, seleccione la opción Editor de sintaxis y pulse en el botón Iniciar ayuda para la sintaxis de SPSS Statistics.

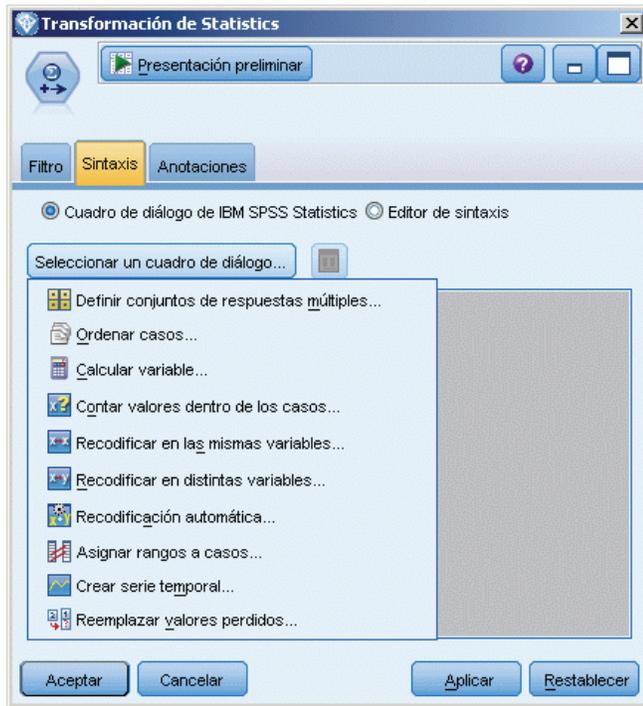
*Nota:* este nodo no admite toda la sintaxis de SPSS Statistics. Si desea obtener más información, consulte el tema [Sintaxis permitida](#) el p. 491.

## **Pestaña Sintaxis del nodo Transformación de Statistics**

### **Opción de cuadro de diálogo de IBM SPSS Statistics**

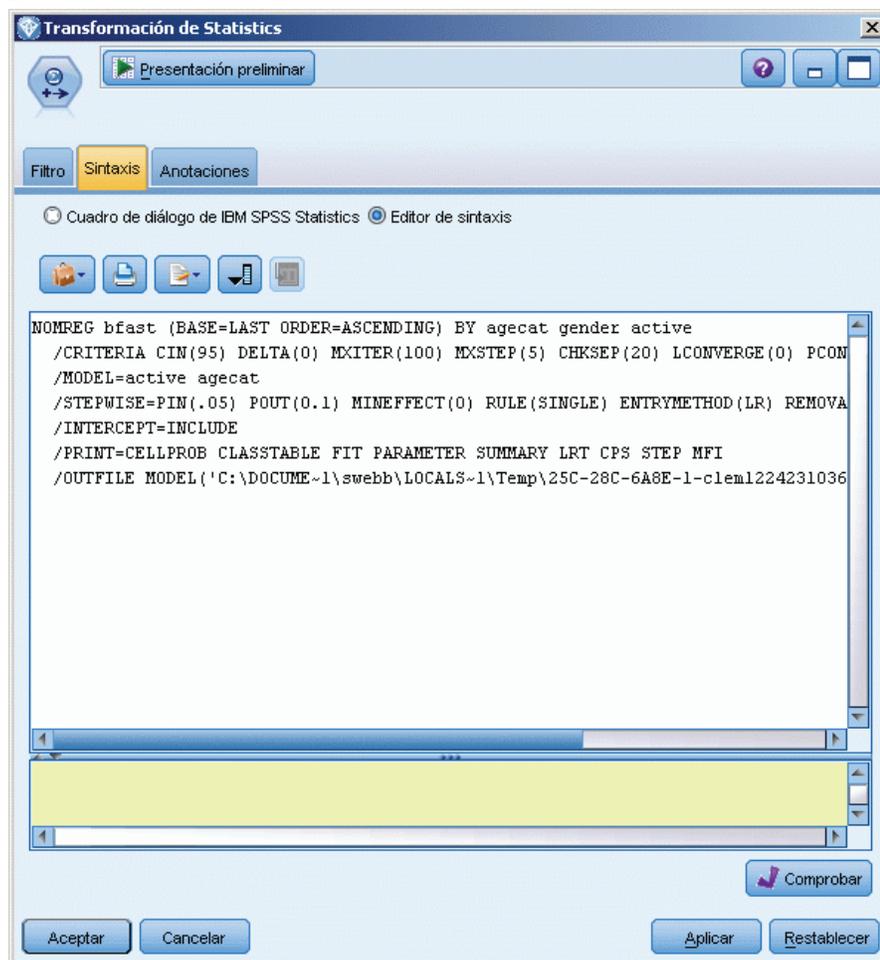
Si no está familiarizado con la sintaxis de IBM® SPSS® Statistics para un procedimiento, el modo más sencillo de crear sintaxis en IBM® SPSS® Modeler es seleccionar la opción Cuadro de diálogo IBM SPSS Statistics, seleccionar el cuadro de diálogo del procedimiento, cumplimentar el cuadro de diálogo y pulsar en Aceptar. Al realizar estas acciones, colocará la sintaxis en la pestaña Sintaxis del nodo de SPSS Statistics que está utilizando en SPSS Modeler. A continuación, podrá ejecutar la ruta para obtener los resultados del procedimiento.

Figura 8-2  
Cuadro de diálogo del nodo Transformación de Statistics



### Opción del editor Sintaxis de IBM SPSS Statistics

Figura 8-3  
Editor Sintaxis del nodo Transformación de Statistics



**Comprobar.** Tras introducir los comandos de sintaxis en la parte superior del cuadro de diálogo, utilice este botón para validar las entradas. Todos los errores de sintaxis aparecen identificados en la parte inferior del cuadro de diálogo.

Para garantizar que el proceso de comprobación no tarda demasiado, al validar la sintaxis, se comprueba una muestra representativa de los datos para comprobar que las entradas son válidas, en lugar de comprobar el conjunto de datos completo.

### Sintaxis permitida

Si tiene una gran cantidad de sintaxis heredada de IBM® SPSS® Statistics o está familiarizado con las funciones de preparación de datos de SPSS Statistics, puede utilizar el nodo Transformación de Statistics para ejecutar muchas de las transformaciones existentes. A modo de orientación, el nodo permite transformar los datos de maneras predecibles (por ejemplo, ejecutando bucles de comandos o cambiando, añadiendo, ordenando, filtrando o seleccionando datos).

Estos son algunos ejemplos de comandos que se pueden llevar a cabo:

- Calcular números aleatorios según una distribución binomial:

```
COMPUTE nuevavar = RV.BINOM(10000,0.1)
```

- Recodificar una variable en una nueva variable:

```
RECODE Edad (Lowest thru 30=1) (30 thru 50=2) (50 thru Highest=3) INTO EdadRecodificada
```

- Sustituir valores perdidos:

```
RMV Edad_1=SMEAN(Edad)
```

En la siguiente tabla se enumera la sintaxis de SPSS Statistics admitida por el nodo Transformación de Statistics:

**Nombre de comando**

```
ADD VALUE LABELS
APPLY DICTIONARY
AUTORECODE
BREAK
CD
CLEAR MODEL PROGRAMS
CLEAR TIME PROGRAM
CLEAR TRANSFORMATIONS
COMPUTE
COUNT
CREATE
DATE
DEFINE-!ENDDEFINE
DELETE VARIABLES
DO IF
DO REPEAT
ELSE
ELSE IF
END CASE
END FILE
END IF
END INPUT PROGRAM
END LOOP
END REPEAT
EXECUTE
FILE HANDLE
FILE LABEL
FILE TYPE-END FILE TYPE
FILTER
FORMATS
IF
INCLUDE
INPUT PROGRAM-END INPUT PROGRAM
INSERT
LEAVE
LOOP-END LOOP
```

**Nombre de comando**

MATRIX-END MATRIX  
MISSING VALUES  
N OF CASES  
NUMERIC  
PERMISSIONS  
PRESERVE  
RANK  
RECODE  
RENAME VARIABLES  
RESTORE  
RMV  
SAMPLE  
SELECT IF  
SET  
SORT CASES  
SORT CASES  
STRING  
SUBTITLE  
TEMPORARY  
TITLE  
UPDATE  
V2C  
VALIDATEDATA  
VALUE LABELS  
VARIABLE ATTRIBUTE  
VARSTOCASES  
VECTOR

## ***Nodo Modelo de Statistics***

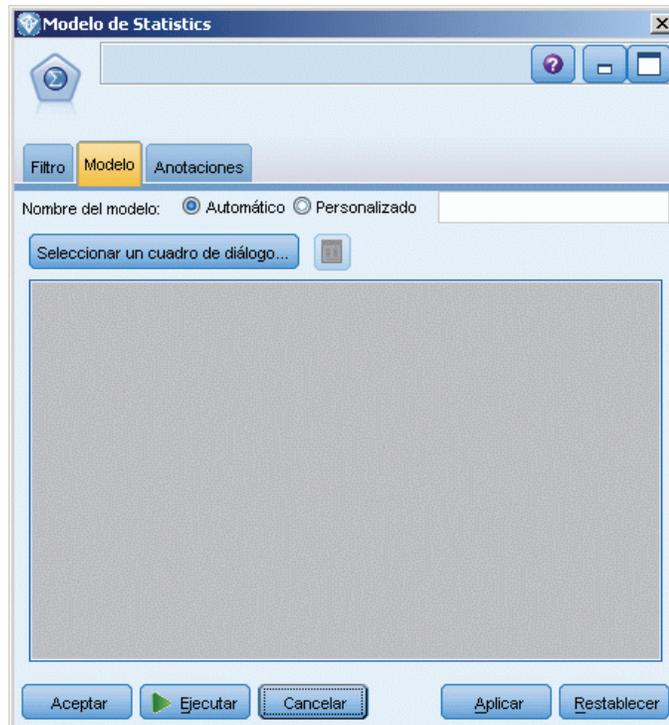
El nodo Modelos de Statistics permite analizar y trabajar con sus datos ejecutando los procedimientos de IBM® SPSS® Statistics que producen PMML. Los nuggets de modelo que cree pueden utilizarse luego de la manera habitual en las rutas de IBM® SPSS® Modeler para su puntuación y otras operaciones.

Debe tener una versión compatible de SPSS Statistics instalada en su equipo y con licencia para poder usar este nodo. Si desea obtener más información, consulte el tema [Complementos de IBM SPSS Statistics](#) en el capítulo 6 el p. 448. Consulte las notas de publicación para obtener información sobre compatibilidad.

Los procedimientos analíticos de SPSS Statistics disponibles dependen del tipo de licencia que disponga.

## Pestaña Modelo del nodo Modelos de Statistics

Figura 8-4  
Pestaña Modelo del nodo Modelos de Statistics

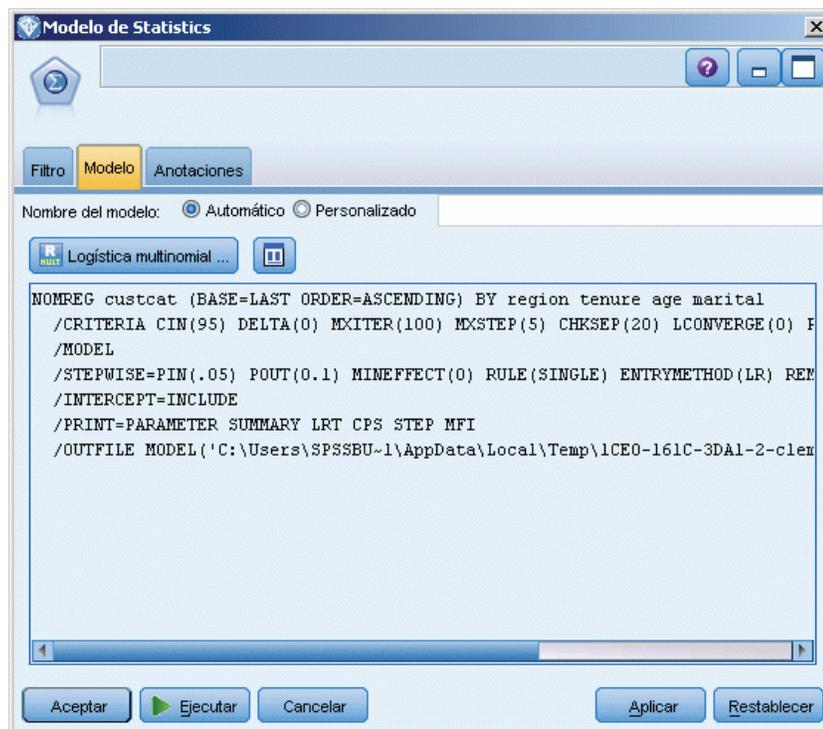


**Nombre del modelo.** Puede generar el nombre del modelo de forma automática basándose en el campo objetivo o de ID (o en el nombre del tipo de modelo si se especifica ningún campo objetivo), o bien especificar un nombre personalizado.

**Seleccione un cuadro de diálogo.** Pulse para mostrar una lista de procedimientos de IBM® SPSS® Statistics disponibles que puede seleccionar y ejecutar. La lista sólo muestra aquellos procedimientos que producen PMML y para los que tiene licencia, y no incluye los procedimientos escritos por el usuario.

- ▶ Pulse en el procedimiento necesario; aparecerá el cuadro de diálogo de SPSS Statistics correspondiente.
- ▶ En el cuadro de diálogo de SPSS Statistics, introduzca los detalles para el procedimiento.
- ▶ Pulse en Aceptar para volver al nodo Modelos de Statistics; la sintaxis de SPSS Statistics se muestra en la pestaña Modelo.

Figura 8-5  
Sintaxis mostrada en la pestaña Modelo

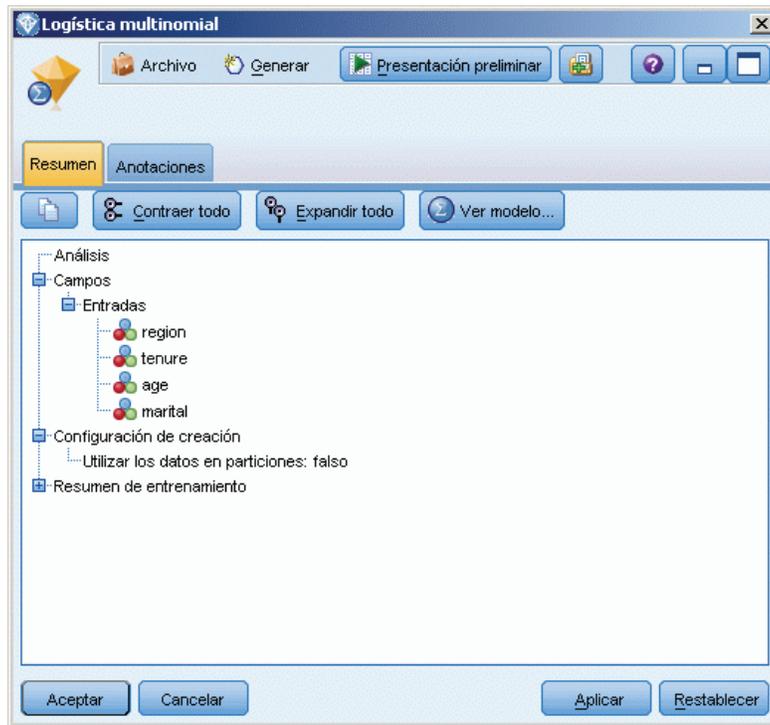


- Para volver al cuadro de diálogo de SPSS Statistics en cualquier momento, como por ejemplo para modificar su consulta, pulse en el cuadro de diálogo de SPSS Statistics a la derecha del botón de selección de procedimiento.

### ***Nodo Modelo de Statistics - Resumen del nugget de modelo***

Cuando ejecuta el nodo Modelo de Statistics, se ejecuta el procedimiento asociado de IBM® SPSS® Statistics y se crea un nugget de modelo que puede utilizar en las rutas de IBM® SPSS® Modeler para su puntuación.

Figura 8-6  
Nugget Modelo de Statistics, pestaña Resumen



La pestaña Resumen de un nugget de modelo muestra información sobre los campos, la configuración de creación y el proceso de estimación del modelo. Los resultados se muestran en una vista de árbol que se puede expandir o contraer pulsando los elementos específicos.

El botón Ver modelo muestra los resultados en un formato modificado del Visor de resultados de SPSS Statistics. Para obtener más información sobre este visor, consulte la documentación de SPSS Statistics.

Las opciones normales de exportación e impresión están disponibles en el menú Archivo. Si desea obtener más información, consulte el tema [Visualización de resultados](#) en el capítulo 6 el p. 393.

Figura 8-7  
Pestaña Avanzada del nugget Modelo de Statistics

The screenshot shows the 'output21115.spv - PASW Statistics Output Viewer' window. The left pane shows a tree view with 'Resultado' expanded to 'Regresión nominal'. The main area displays three tables:

**Información del ajuste del modelo**

Modelo	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
		-2 log verosimilitud	Chi-cuadrado	gl
Sólo la intersección	2.737E3			
Final	2.142E3	594.986	399	.000

**Pseudo R-cuadrado**

Cox y Snell	.448
Nagelkerke	.479
McFadden	.215

**Contrastes de la razón de verosimilitud**

Efecto	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
		-2 log verosimilitud del modelo reducido	Chi-cuadrado	gl

## Nodo Resultados de Statistics

El nodo Resultados de Statistics le permite llamar a un procedimiento de IBM® SPSS® Statistics para analizar los datos de IBM® SPSS® Modeler. Puede ver los resultados en una ventana del explorador o guardarlos en el formato de archivo de resultados de SPSS Statistics. Se puede acceder a una gran variedad de procedimientos analíticos de SPSS Statistics desde SPSS Modeler.

Debe tener una versión compatible de SPSS Statistics instalada en su equipo y con licencia para poder usar este nodo. Si desea obtener más información, consulte el tema [Complementos de IBM SPSS Statistics](#) en el capítulo 6 el p. 448. Consulte las notas de publicación para obtener información sobre compatibilidad.

Si es necesario, puede utilizar la ficha Filtro para filtrar los campos o cambiar su nombre para que cumplan los estándares de nomenclatura de SPSS Statistics. Si desea obtener más información, consulte el tema [Cambio de nombre o filtrado de campos para IBM SPSS Statistics](#) el p. 503.

**Referencia de sintaxis.** Para obtener detalles sobre procedimientos SPSS Statistics específicos, consulte la guía *Referencia de sintaxis de comandos de SPSS Statistics*, que se incluye con su copia del software SPSS Statistics. Para ver la guía de la pestaña Sintaxis, seleccione la opción Editor de sintaxis y pulse en el botón Iniciar ayuda para la sintaxis de SPSS Statistics.

## Pestaña Sintaxis del nodo Resultados de Statistics

Use esta pestaña para crear sintaxis para el procedimiento de IBM® SPSS® Statistics que desea utilizar para analizar los datos. La sintaxis consta de dos partes: una **instrucción** y **opciones** asociadas. La instrucción especifica el análisis o la operación que se va a realizar y los campos que se van a utilizar. Las opciones especifican todo lo demás, incluyendo los estadísticos que se van a mostrar, los campos derivados que se van a guardar, etc.

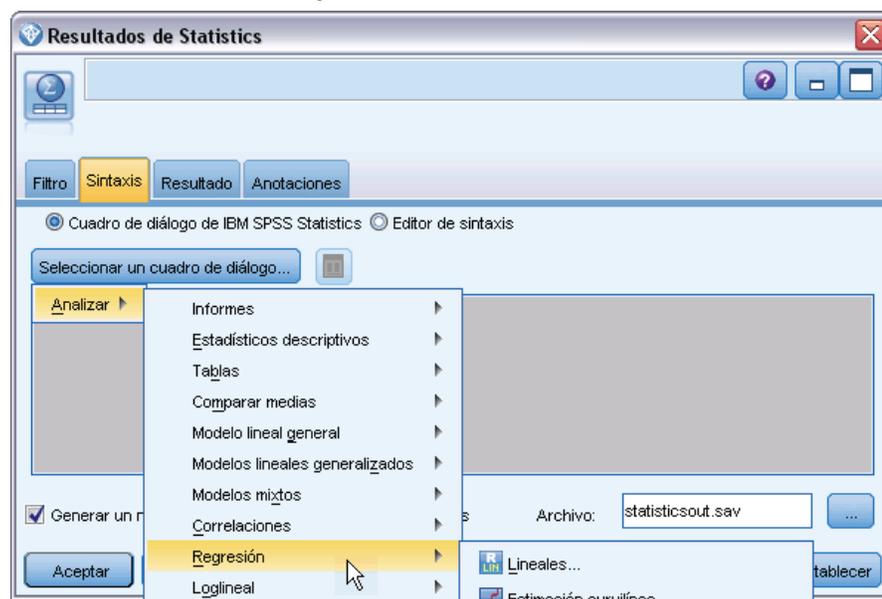
### Opción de cuadro de diálogo de IBM SPSS Statistics

Si no está familiarizado con la sintaxis de SPSS Statistics para un procedimiento, el modo más sencillo de crear sintaxis en IBM® SPSS® Modeler es seleccionar la opción Cuadro de diálogo IBM SPSS Statistics, seleccionar el cuadro de diálogo del procedimiento, cumplimentar el cuadro de diálogo y pulsar en Aceptar. Al realizar estas acciones, colocará la sintaxis en la pestaña Sintaxis del nodo de SPSS Statistics que está utilizando en SPSS Modeler. A continuación, podrá ejecutar la ruta para obtener los resultados del procedimiento.

También puede generar un nodo de origen Archivo de Statistics para importar los datos resultantes. Esto es de utilidad, por ejemplo, si un procedimiento escribe campos como puntuaciones para el conjunto de datos activo además de mostrar los resultados.

Figura 8-8

Selección de cuadro de diálogo del nodo Resultados de Statistics



Para crear la sintaxis:

- ▶ Pulse en el botón Seleccionar un cuadro de diálogo....
- ▶ Seleccione una de las opciones:
  - **Analizar.** Enumera el contenido del menú Analizar de SPSS Statistics; seleccione el procedimiento que desee utilizar.
  - **Otro.** Si aparece, enumera los cuadros de diálogo creados por el generador de cuadros de diálogo personalizados en SPSS Statistics, así como otros cuadros de diálogo de SPSS Statistics que no aparecen en el menú Analizar y para los que tiene licencia. Si no se especifica ningún cuadro de diálogo aplicable, esta opción no aparece.

*Nota:* Los cuadros de diálogo Preparación de datos automática no se muestran.

Si tiene un cuadro de diálogo personalizado de SPSS Statistics que crea nuevos campos, estos campos no se podrán utilizar en SPSS Modeler porque el nodo Resultados de Statistics es un nodo terminal.

- ▶ También puede marcar la casilla Generar un nodo de importación para los datos resultantes para crear un nodo de origen Archivo de Statistics que puede utilizarse para importar los datos resultantes en otra ruta. El nodo se coloca en el lienzo de pantalla, con los datos incluidos en el archivo *.sav* especificados en el campo Archivo (la ubicación por defecto es el directorio de instalación de SPSS Modeler).

### ***Opción del editor Sintaxis***

Para guardar la sintaxis que se ha creado para un procedimiento de uso frecuente:

- ▶ Pulse en el botón Opciones de archivo (el primero de la barra de herramientas).
- ▶ En el menú seleccione Guardar o Guardar como.
- ▶ Guarde el archivo como un archivo *sps*.

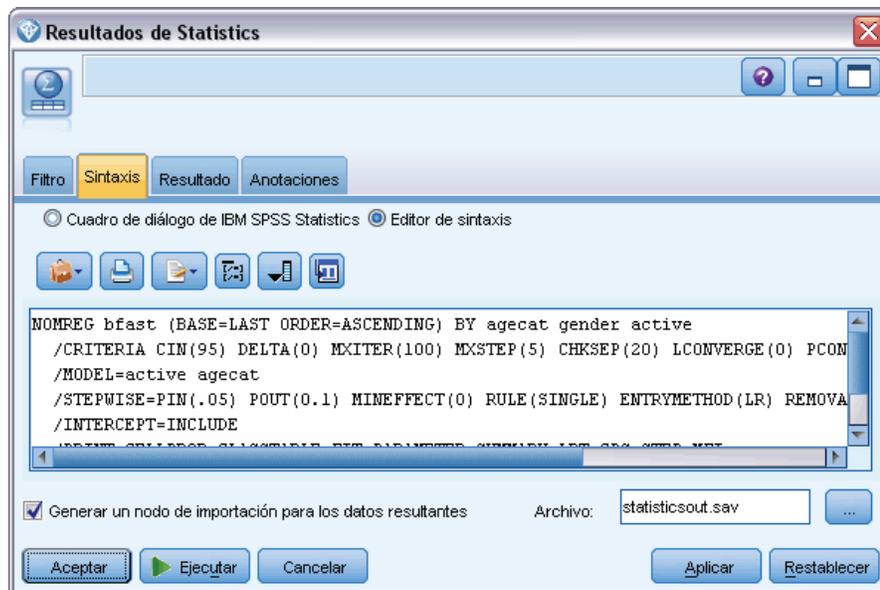
Para usar archivos de sintaxis creados previamente, sustituya el contenido actual, si los hay, del editor Sintaxis:

- ▶ Pulse en el botón Opciones de archivo (el primero de la barra de herramientas).
- ▶ Seleccione Abrir en el menú.
- ▶ Al seleccionar un archivo *.sps* se pegará el contenido en la pestaña Sintaxis del nodo Resultados.

Para insertar una sintaxis guardada previamente sin sustituir el contenido actual:

- ▶ Pulse en el botón Opciones de archivo (el primero de la barra de herramientas).
- ▶ Seleccione Insertar en el menú.
- ▶ Al seleccionar un archivo *.sps* se pegará el contenido en el nodo Resultados en el punto especificado por el cursor.

Figura 8-9  
Editor Sintaxis del nodo Resultados de Statistics



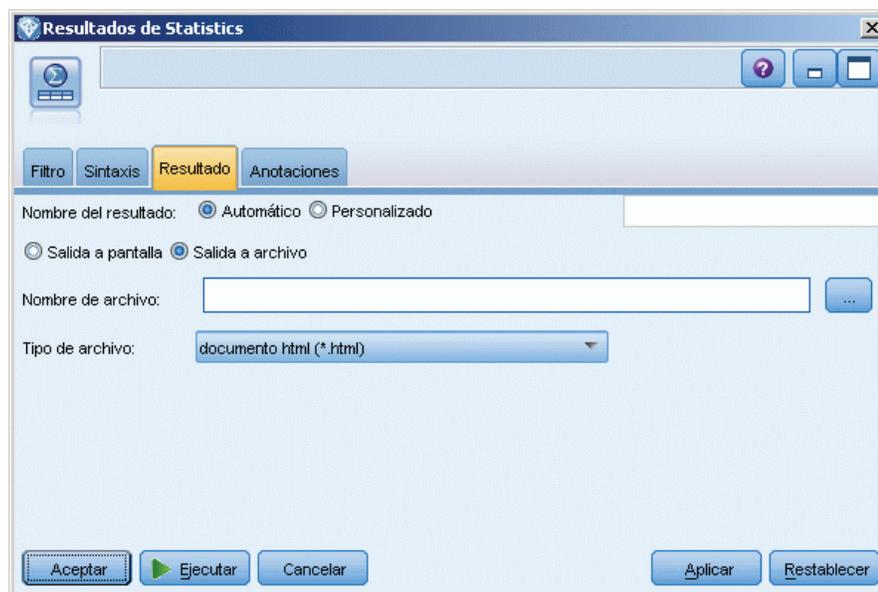
- También puede marcar la casilla Generar un nodo de importación para los datos resultantes para crear un nodo de origen Archivo de Statistics que puede utilizarse para importar los datos resultantes en otra ruta. El nodo se coloca en el lienzo de pantalla, con los datos incluidos en el archivo *.sav* especificados en el campo Archivo (la ubicación por defecto es el directorio de instalación de SPSS Modeler).

Al pulsar en Ejecutar, se muestran los resultados en el Visor de resultados de SPSS Statistics. Para obtener más información sobre el visor, consulte la documentación de SPSS Statistics.

### ***Pestaña Resultados del nodo Resultados de Statistics***

La pestaña Resultado permite especificar el formato y la ubicación del resultado. Puede elegir entre mostrar los resultados en pantalla o enviarlos a uno de los tipos de archivo disponibles.

Figura 8-10  
Pestaña Resultados del nodo Resultados de Statistics



**Nombre del resultado.** Especifica el nombre del resultado generado cuando se ejecuta el nodo. Automático selecciona un nombre según el nodo que genera el resultado. De forma opcional, puede seleccionar Personalizado para especificar un nombre diferente.

**Salida a pantalla** (por defecto). Crea un objeto de resultados para ver en línea. El objeto de resultados aparecerá en la pestaña Resultados de la ventana del administrador al ejecutar el nodo de resultados.

**Salida a archivo.** Almacena el resultado en un archivo cuando se ejecuta el nodo. Si selecciona esta opción, introduzca un nombre de archivo en el campo Nombre de archivo (o navegue a un directorio y especifique un nombre de archivo mediante el botón Buscar el archivo) y seleccione un tipo de archivo.

**Tipo de archivo.** Seleccione el tipo de archivo al que desee enviar los resultados.

- **Documento HTML (\*.html)** Escribe los resultados en un formato HTML.
- **SPSS Statistics Archivo de visor (\*.spv).** Escribe los resultados en un formato que puede leer el Visor de resultados de IBM® SPSS® Statistics.
- **SPSS Statistics Archivo de informes web (\*.spw).** Escribe los resultados en formato de informes web de SPSS Statistics, que pueden publicarse en un repositorio de un IBM SPSS Collaboration and Deployment Services y verse posteriormente en un navegador web. Si desea obtener más información, consulte el tema [Publicar en Web](#) en el capítulo 6 el p. 393.

## Nodo Exportación de Statistics

El nodo Exportar de Statistics permite exportar datos en formato *.sav* de IBM® SPSS® Statistics. SPSS Statistics Los archivos *.sav* se pueden leer con SPSS Statistics Base y otros módulos. Este es también el formato utilizado para los archivos caché de IBM® SPSS® Modeler.

Establecer una correspondencia entre los nombres de campos de SPSS Modeler y los nombres de variables de SPSS Statistics puede provocar errores ya que los nombres de variables de SPSS Statistics están limitados a 64 caracteres y no pueden incluir determinados caracteres, como espacios, el signo del dólar (\$) y guiones (-). Existen dos modos de ajustar estas restricciones:

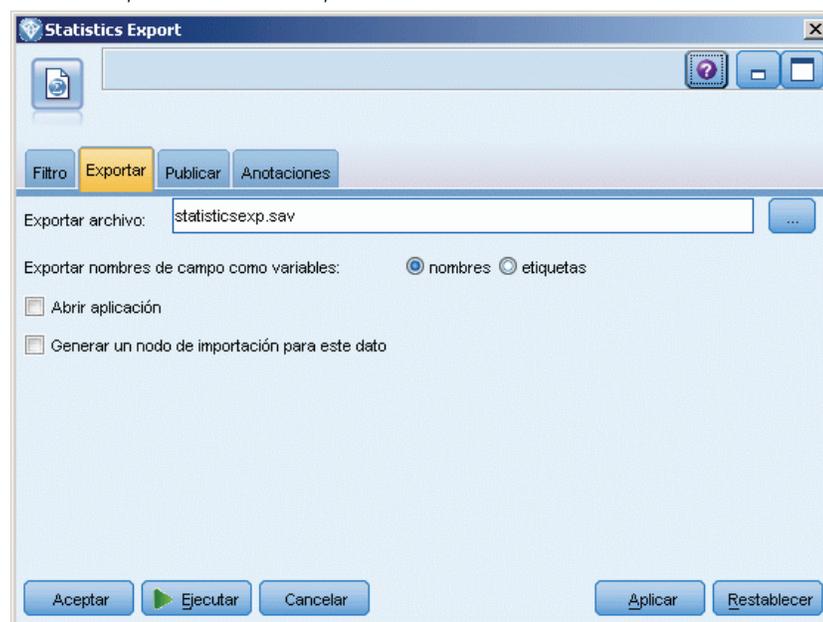
- Puede cambiar el nombre de los campos ajustándose a los requisitos del nombre de las variables de SPSS Statistics pulsando en la pestaña Filtro. Si desea obtener más información, consulte el tema [Cambio de nombre o filtrado de campos para IBM SPSS Statistics](#) el p. 503.
- Elija exportar las etiquetas y los nombres de campos desde SPSS Modeler.

**Nota:** SPSS Modeler describe archivos .sav en formato Unicode UTF-8. SPSS Statistics sólo admite archivos en formato Unicode UTF-8 de la versión 16.0 en adelante. Para evitar la posibilidad de corrupción de datos de archivos .sav guardados con codificación Unicode, no se deberían utilizar en versiones de SPSS Statistics previas a 16.0. Si desea más información, consulte la ayuda de SPSS Statistics.

**Conjuntos de respuestas múltiples.** Cualquier conjunto de respuestas múltiples definido en la ruta se conservará automáticamente cuando se exporte el archivo. Puede definir o editar conjuntos de respuestas múltiples en un nodo con una pestaña Filtro. Si desea obtener más información, consulte el tema [Edición de conjuntos de respuestas múltiples](#) en el capítulo 4 el p. 158.

## ***Pestaña Exportar del nodo Exportar Statistics***

Figura 8-11  
Pestaña Exportar del nodo Exportar Statistics



**Exportar archivo.** Especifica el nombre del archivo. Introduzca un nombre de archivo o pulse en el botón Buscar el archivo para examinar la ubicación del archivo.

**Exportar nombres de campo como variables.** Especifica un método para la gestión de etiquetas y nombres de variables tras la exportación desde IBM® SPSS® Modeler a un archivo *.sav* de IBM® SPSS® Statistics.

- **nombres.** Para exportar los nombres y las etiquetas de campo de SPSS Modeler. Los nombres se exportan como nombres de variables de SPSS Statistics, mientras que las etiquetas se exportan como etiquetas de variables de SPSS Statistics.
- **etiquetas.** Seleccione esta opción para utilizar los nombres de campo de SPSS Modeler como etiquetas de variable en SPSS Statistics. SPSS Modeler admite los caracteres de nombres de campo que no son válidos en nombres de variables de SPSS Statistics. Para evitar crear posibles nombres de SPSS Statistics no válidos, seleccione nombres en Exportar nombres de campo como variables en su lugar o utilice la pestaña Filtro para ajustar los nombres de campos.

**Abrir aplicación.** Si SPSS Statistics está instalado en su ordenador, puede seleccionar esta opción para activarlos directamente en el archivo de datos almacenado. Las opciones de ejecución de cada una de las aplicaciones se deben especificar en el cuadro de diálogo Complementos. Si desea obtener más información, consulte el tema [Complementos de IBM SPSS Statistics](#) en el capítulo 6 el p. 448. Para crear simplemente un archivo *.sav* de SPSS Statistics sin abrir un programa externo, anule la selección de esta opción.

**Generar un nodo de importación para este dato.** Seleccione esta opción para generar automáticamente un nodo de origen Archivo de Statistics que leerá el archivo de datos exportado. Si desea obtener más información, consulte el tema [Nodo de archivos estadísticos](#) el p. 487.

## ***Cambio de nombre o filtrado de campos para IBM SPSS Statistics***

Antes de exportar o distribuir los datos de IBM® SPSS® Modeler a aplicaciones externas como IBM® SPSS® Statistics, puede que sea necesario cambiar o ajustar los nombres de los campos. Los cuadros de diálogo Transformación de Statistics, Resultados de Statistics y Exportar Statistics contienen una pestaña Filtro para facilitar este proceso.

En otras secciones se ofrece una descripción básica de la funcionalidad de la pestaña Filtro. Si desea obtener más información, consulte el tema [Opciones de configuración de filtrado](#) en el capítulo 4 el p. 155. Este tema ofrece sugerencias para leer datos en SPSS Statistics.

Figura 8-12

Cambio de nombre de los campos para IBM SPSS Statistics en la pestaña Filtro del nodo Archivo de Statistics.

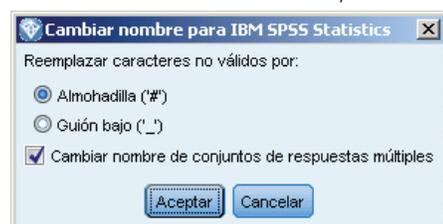


Para ajustar los nombres de campo para conformar a las convenciones de nomenclatura de SPSS Statistics:

- ▶ En la pestaña Filtro, pulse en el botón de la barra de herramientas del menú Opciones de filtro (el primero de la barra de herramientas).
- ▶ Seleccione Cambiar nombre para SPSS Statistics.

Figura 8-13

Cambio de nombre de los campos



- ▶ En el cuadro de diálogo Cambiar nombre para SPSS Statistics puede seleccionar sustituir los caracteres no válidos de los nombres de archivo con un carácter de almohadilla (#) o un guión bajo (\_).

**Cambiar nombre de conjuntos de respuestas múltiples.** Seleccione esta opción si desea ajustar los nombres de conjuntos de respuestas múltiples, que pueden importarse a SPSS Modeler mediante un nodo de origen Archivo de Statistics. Se usan para registrar datos que pueden tener más de un valor para cada caso, como en respuestas de encuestas.

# Supernodos

## Conceptos básicos de supernodos

El que cada nodo tenga una función claramente definida es uno de los motivos por los que resulta tan fácil aprender la interfaz de programación visual de IBM® SPSS® Modeler. Sin embargo, para procesamientos complejos puede ser necesaria una larga secuencia. Con el tiempo, esto puede recargar el lienzo de rutas y puede resultar difícil seguir los diagramas de rutas. Hay dos formas de evitar que una ruta larga y compleja se recargue:

- Puede dividir una secuencia de procesamiento en varias rutas introduciéndose una en la otra. La primera ruta, por ejemplo, crea un archivo de datos que la segunda utiliza como entrada. La segunda crea un archivo que la tercera utiliza como entrada y así sucesivamente. Puede administrar varias rutas guardándolas en un **proyecto**. Un proyecto proporciona organización para varias rutas y sus salidas. Sin embargo, un archivo de proyecto sólo contiene una referencia a los objetos que incluye y aún tendrá que administrar varios archivos de rutas.
- Puede crear un **Supernodo** como alternativa más simplificada al trabajar con procesos de ruta complejos.

Los supernodos agrupan varios nodos en uno solo encapsulando las secciones de una ruta de datos. Esto proporciona numerosas ventajas al minero de datos:

- Las rutas estarán más ordenadas y serán más manejables.
- Los nodos se podrán unir en un supernodo específico de negocio.
- Los supernodos se podrán exportar a bibliotecas para reutilizarlos en varios proyectos de minería de datos.

## Tipos de supernodos

Los supernodos se representan en la ruta de datos mediante un icono de estrella. El icono aparece sombreado para representar el tipo de supernodo y la dirección en la que debe fluir la ruta o desde la que debe fluir.

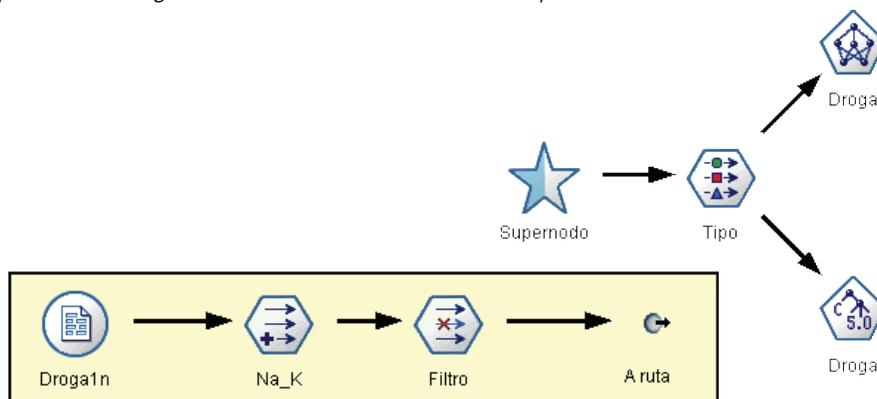
Hay tres tipos de supernodos:

- Supernodos de origen
- Supernodos de proceso
- Supernodos terminales

## Supernodos de origen

Los supernodos de origen contienen un origen de datos como un nodo de origen normal y se pueden utilizar en cualquier lugar al igual que un nodo de origen normal. La parte izquierda del supernodo de origen está sombreada para indicar que está “cerrado” en esa parte y que los datos deben fluir hacia abajo *desde* el supernodo.

Figura 9-1  
Supernodo de origen con versión de acercamiento impuesta sobre la ruta

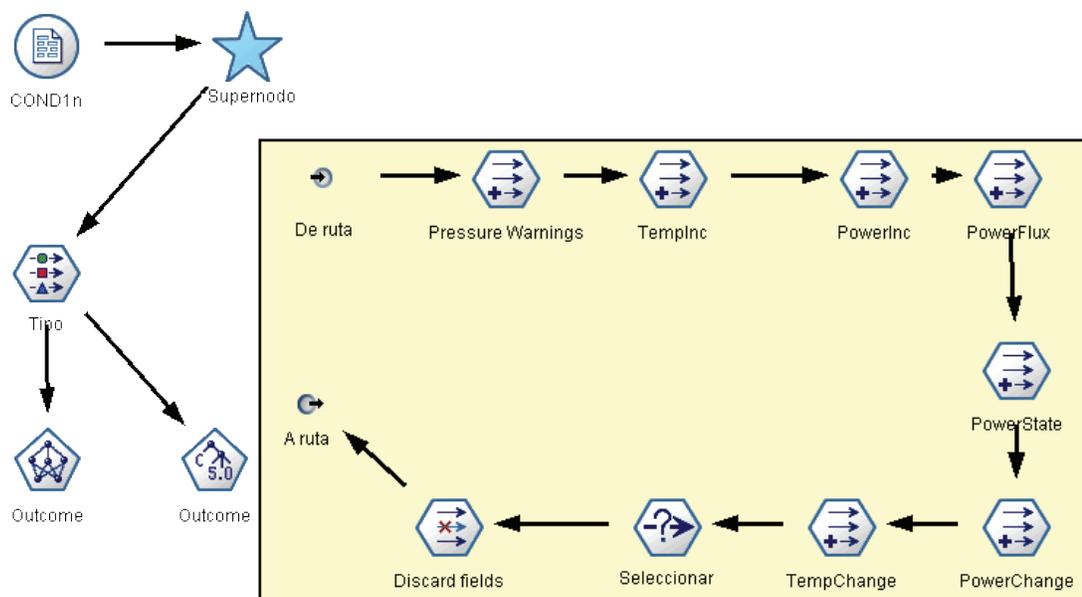


Los supernodos de origen tienen sólo un punto de conexión a la derecha, mostrando que los datos dejan el supernodo y fluyen hacia la ruta.

## Supernodos de proceso

Los supernodos de proceso sólo contienen nodos de proceso y no están sombreados para mostrar que los datos pueden fluir *dentro* y *fuera* de este tipo de supernodo.

Figura 9-2  
Supernodo de proceso con versión de acercamiento impuesta sobre la ruta



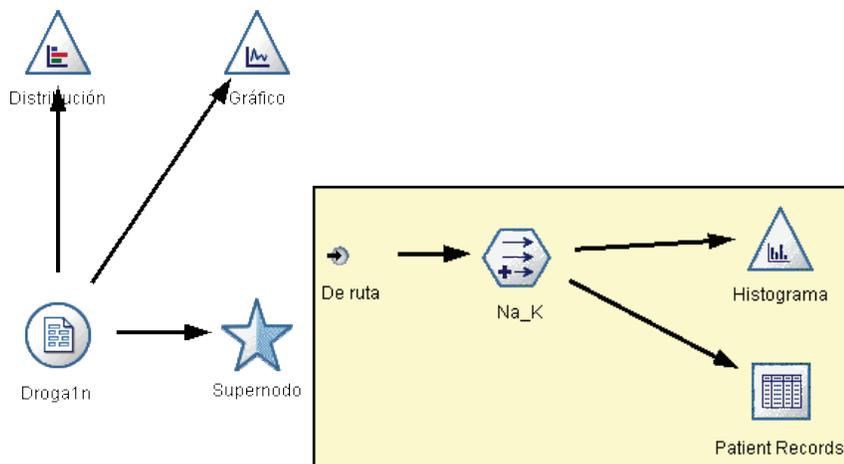
Los supernodos de proceso tienen puntos de conexión a la derecha y a la izquierda, mostrando que los datos entran en el supernodo y lo dejan para volver a la ruta. Aunque los supernodos pueden contener fragmentos de rutas adicionales e incluso rutas extra, los dos puntos de conexión deben fluir a través de una sola ruta de acceso conectando los puntos *De ruta* y *A ruta*.

*Nota:* los supernodos de proceso a veces se denominan *Supernodos de manipulación*.

### **Supernodos terminales**

Los supernodos terminales contienen uno o varios nodos terminales (gráfico, tabla, etc.) y se pueden utilizar de la misma forma que un nodo terminal. Un supernodo terminal aparece sombreado en la parte derecha para indicar que está “cerrado” en esa parte y que los datos pueden fluir sólo *en* un supernodo terminal.

**Figura 9-3**  
Supernodo terminal con versión de acercamiento impuesta sobre la ruta



Los supernodos terminales sólo tienen un punto de conexión a la derecha, mostrando que los datos entran en el supernodo desde la ruta y finalizan dentro del supernodo.

Los supernodos terminales también pueden contener procesos que se utilizan para especificar el orden de ejecución para todos los nodos terminales dentro del supernodo. Si desea obtener más información, consulte el tema [Supernodos y procesos](#) el p. 523.

## Creación de supernodos

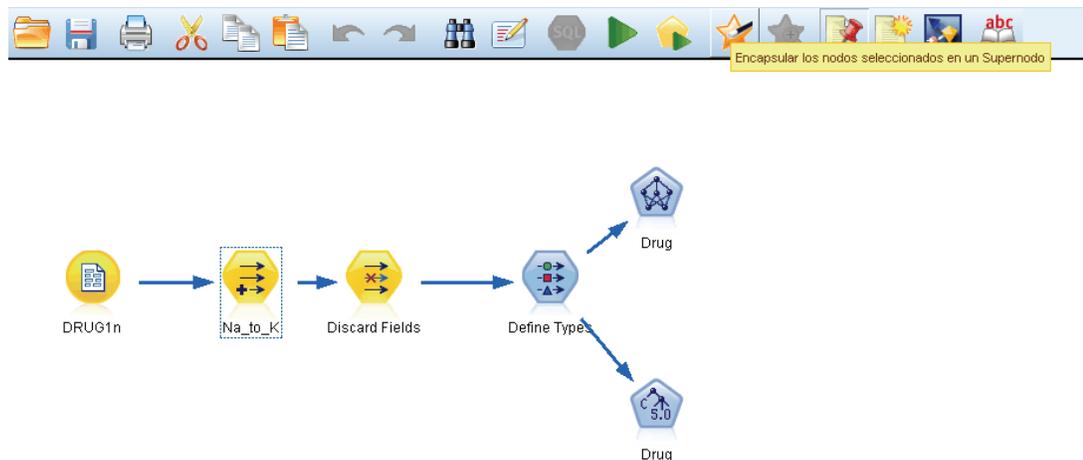
La creación de un supernodo “reduce” la ruta de datos encapsulando varios nodos en uno. Una vez se ha creado o cargado una ruta en el lienzo, hay varias formas de crear un supernodo.

### Selección múltiple

La forma más sencilla de crear un supernodo es seleccionando todos los nodos que desee encapsular:

- ▶ Utilice el ratón para seleccionar varios nodos en el lienzo de rutas. También puede utilizar Mayús-pulsar para seleccionar una ruta o la sección de una ruta. *Nota:* los nodos que seleccione deben ser de una ruta continua o bifida. No puede seleccionar nodos que no sean adyacentes o no estén conectados de alguna forma.
- ▶ A continuación, utilizando uno de los tres métodos siguientes, encapsule los nodos seleccionados:
  - Pulse en el icono Supernodo (sombreado con forma de estrella) de la barra de herramientas.
  - Pulse con el botón derecho en el supernodo y en el menú contextual seleccione lo siguiente:  
Crear Supernodo > Desde selección
  - En el menú de Supernodo, seleccione:  
Crear Supernodo > Desde selección

**Figura 9-4**  
Creación de un supernodo mediante la selección múltiple



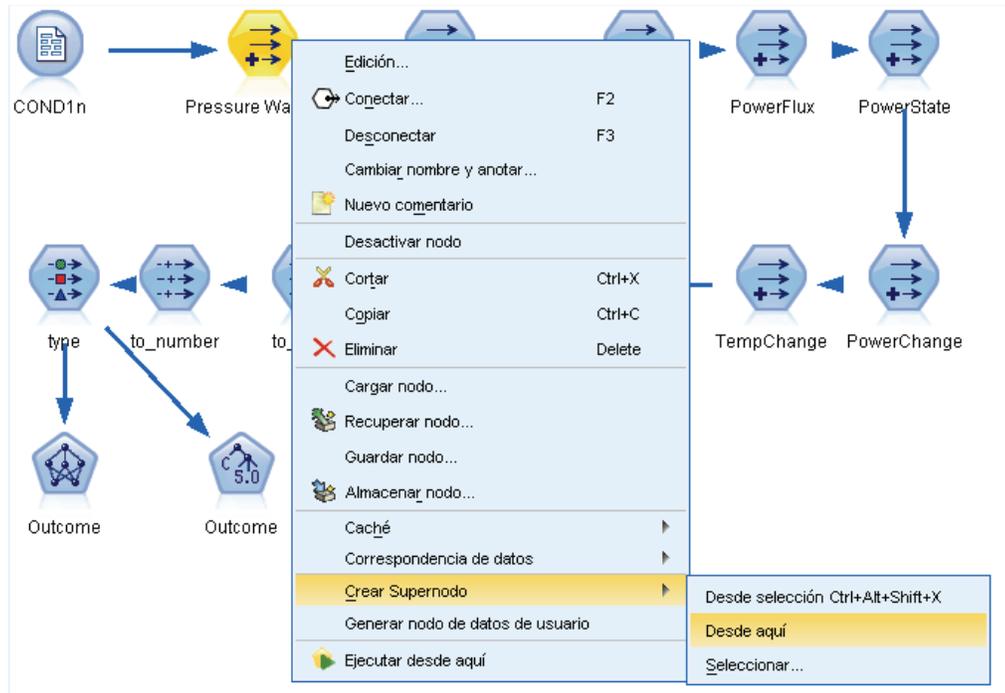
Estas tres opciones encapsulan los nodos en un supernodo sombreado para reflejar el tipo (de origen, de proceso o terminal) según el contenido.

### ***Selección única***

También puede crear un supernodo seleccionando un solo nodo y utilizando las opciones de menú para determinar el inicio y finalización del supernodo o encapsulando todo lo que hay debajo del nodo seleccionado.

- ▶ Pulse en el nodo que determine el inicio de la encapsulación.
- ▶ En el menú de Supernodo, seleccione:  
Crear Supernodo > Desde aquí

**Figura 9-5**  
Creación de un supernodo mediante el menú contextual para una selección única



Los supernodos también se pueden crear de forma más interactiva seleccionando el inicio y el final de la sección de la ruta para encapsular los nodos:

- ▶ Pulse en el primer o último nodo que desee incluir en el supernodo.
- ▶ En el menú de Supernodo, seleccione:  
Crear Supernodo > Seleccionar...
- ▶ Si lo prefiere, puede utilizar las opciones del menú contextual pulsando con el botón derecho en el nodo deseado.
- ▶ El cursor se convierte en un icono de Supernodo, que indica que debe seleccionar otro punto de la ruta. Desplácese hacia arriba o hacia abajo al “otro extremo” del fragmento del supernodo y pulse en un nodo. Esta acción reemplazará todos los nodos entre ellos con el icono de estrella del supernodo.

*Nota:* los nodos que seleccione deben ser de una ruta continua o bífida. No puede seleccionar nodos que no sean adyacentes o no estén conectados de alguna forma.

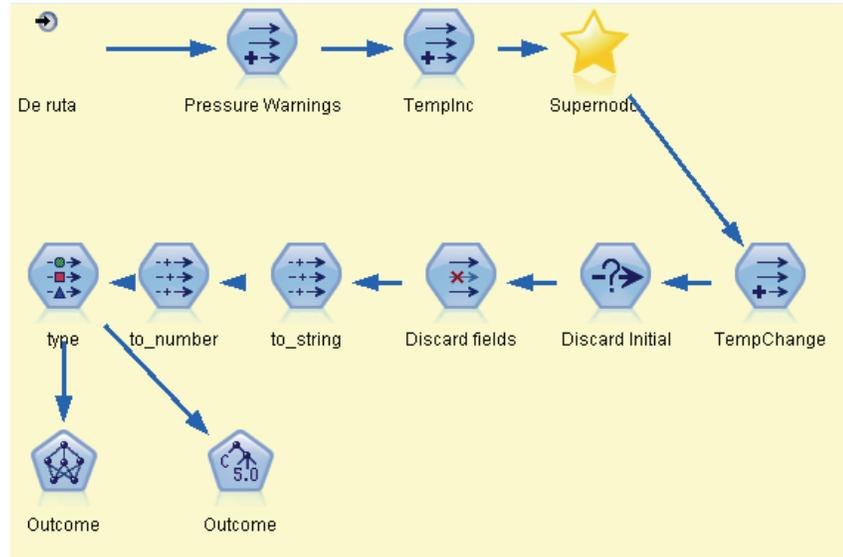
## **Anidación de supernodos**

Los supernodos se pueden anidar en otros supernodos. Las mismas reglas para cada tipo de supernodo (de origen, de proceso y terminal) se aplican a los supernodos anidados. Por ejemplo, un supernodo de proceso con anidación debe tener un flujo continuo de datos a través de todos los

supernodos anidados para que siga siendo un supernodo de proceso. Si uno de los supernodos anidados es terminal, los datos no seguirán fluyendo a través de la jerarquía.

Figura 9-6

Supernodo de proceso anidado en otro supernodo de proceso



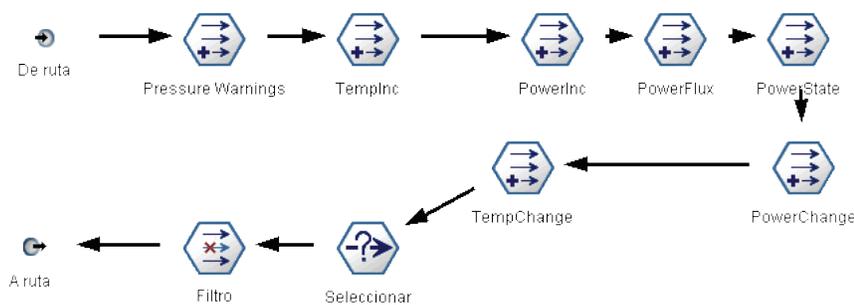
Los supernodos terminales y de origen pueden contener otros tipos de supernodos anidados pero se aplican las mismas reglas básicas para la creación de supernodos.

### Ejemplos de supernodos válidos

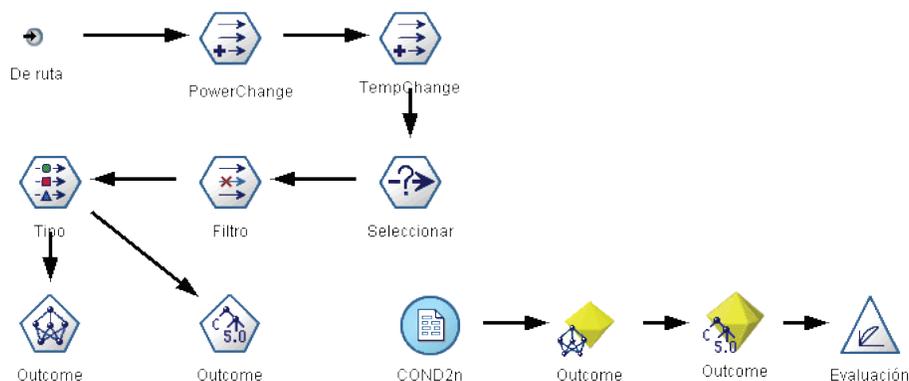
Casi todo lo que pueda crear en IBM® SPSS® Modeler se puede encapsular en un supernodo. A continuación se muestran ejemplos de supernodos válidos:

Figura 9-7

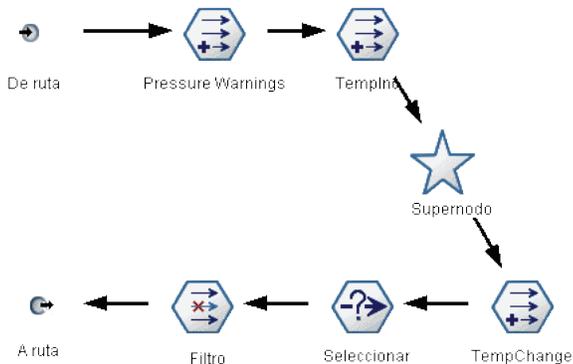
Supernodo de proceso válido con dos conexiones en un flujo de ruta válido



**Figura 9-8**  
Supernodo terminal válido que incluye una ruta independiente utilizada para comprobar los modelos generados



**Figura 9-9**  
Supernodo de proceso válido que contiene un supernodo anidado



### **Ejemplos de supernodos no válidos**

El aspecto más importante a la hora de crear supernodos válidos es asegurar que los datos fluyen de forma lineal a través de las conexiones del supernodo. Si hay dos conexiones (supernodo de proceso), los datos deberán fluir en una ruta desde el conector de inicio al de finalización. De igual modo, un supernodo de origen debe permitir que los datos fluyan del nodo de origen al conector único que devuelve los datos a la ruta de alejamiento.

Figura 9-10

Supernodo de origen no válido: Nodo de origen no conectado a la ruta de flujo de datos

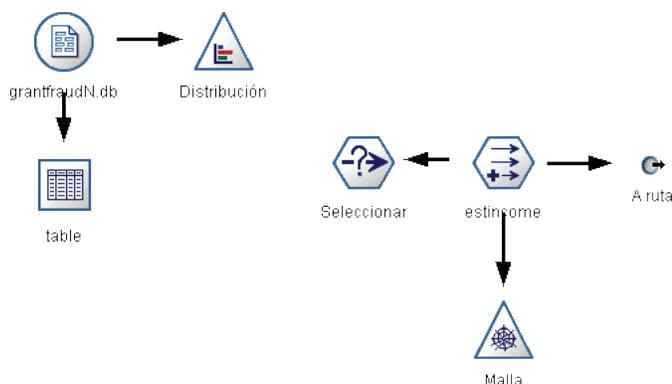
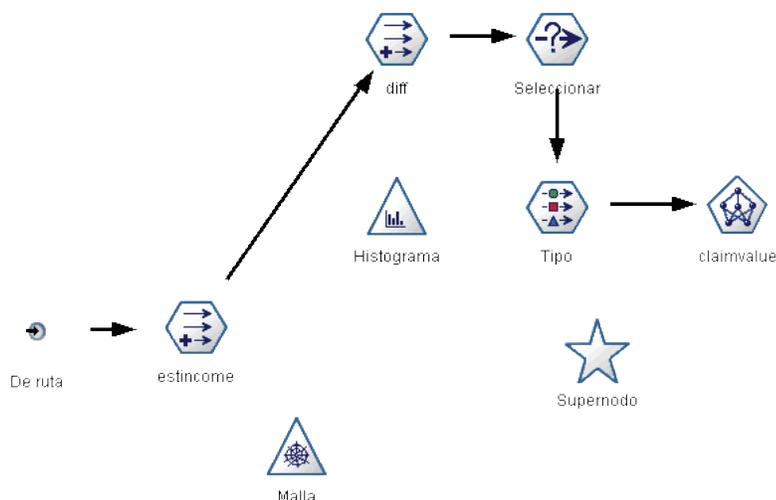


Figura 9-11

Supernodo terminal no válido: Supernodo anidado no conectado a la ruta de flujo de datos



## Bloqueo de Supernodos

Una vez haya creado un supernodo, puede bloquearlo con una contraseña para evitar que se modifique. Por ejemplo, puede hacer esto si crea rutas o partes de rutas, como plantillas de valor fijo para el uso de otros usuarios en su organización que tengan menos experiencia con consultas de configuración de IBM® SPSS® Modeler.

Cuando se bloquea un supernodo, los usuarios podrán introducir valores en la pestaña Parámetros que se hayan definido y se puede ejecutar un supernodo sin introducir la contraseña.

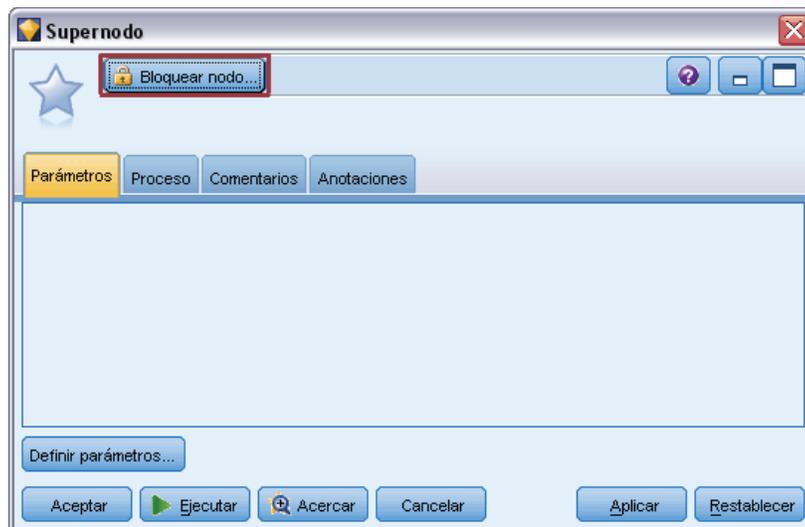
**Nota:** Las operaciones de bloqueo y desbloqueo no se pueden ejecutar utilizando procesos.

## Bloqueo y desbloqueo de un supernodo

**Advertencia:** Las contraseñas perdidas no se pueden recuperar.

Puede bloquear o desbloquear un supernodo desde cualquiera de las tres pestañas.

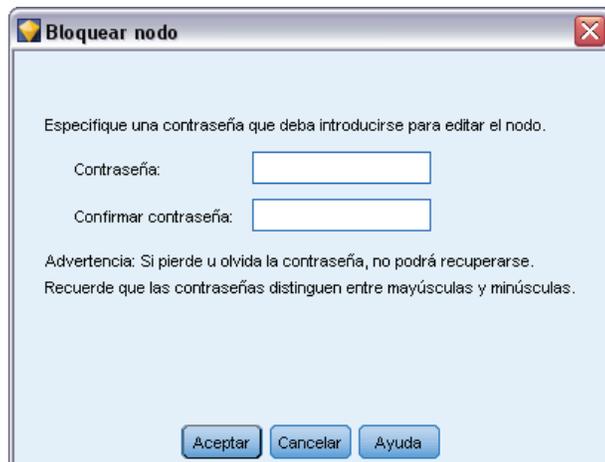
Figura 9-12  
*Bloqueo de un supernodo*



Pulse en Bloquear nodo.

Introduzca y confirme la contraseña.

Figura 9-13  
*Introduzca y confirme la contraseña del supernodo.*



- Pulse en Aceptar.

Un supernodo protegido por contraseña se identifica en el lienzo de rutas por un pequeño icono de candado en la parte superior izquierda del icono del supernodo.

Figura 9-14

*Supernodo de origen bloqueado como parte de una ruta*

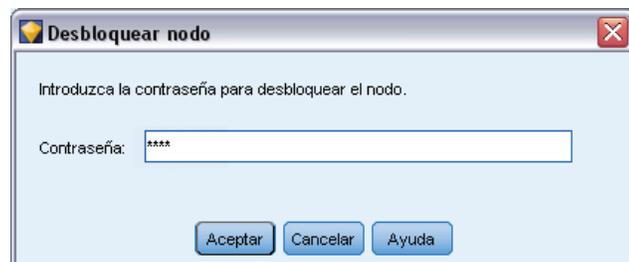


### **Desbloquear un supernodo**

- Para eliminar permanentemente la protección por contraseña, pulse en Desbloquear nodo. Se le solicitará la contraseña.

Figura 9-15

*Introducción de la contraseña para desbloquear un supernodo*



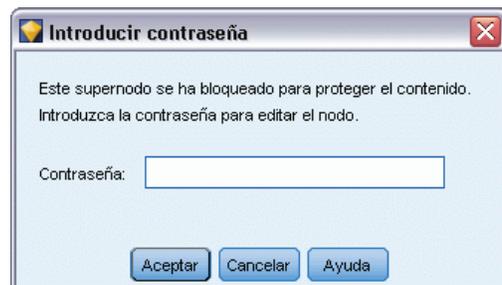
- Introduzca la contraseña y pulse en Aceptar. El supernodo ya no estará protegido por contraseña y el icono de candado ya no aparecerá junto al icono de la ruta.

### **Edición de un supernodo bloqueado**

Si intenta definir los parámetros o ampliar el zoom para ver un supernodo bloqueado, se le solicitará la contraseña.

Figura 9-16

*Introducción de la contraseña de zoom o de edición de un supernodo*



- Introduzca la contraseña y pulse en Aceptar.

Ahora podrá editar las definiciones de los parámetros y ampliar y reducir el zoom cuando lo necesite hasta que cierre la ruta en la que se encuentra el supernodo.

Tenga en cuenta que no se elimina la protección por contraseña, sólo permite acceder a trabajar con el supernodo. Si desea obtener más información, consulte el tema [Bloqueo y desbloqueo de un supernodo](#) el p. 514.

## **Edición de Supernodos**

Una vez haya creado un supernodo, puede examinarlo más detenidamente ampliando el zoom; si el supernodo está bloqueado, se le pedirá que introduzca la contraseña. Si desea obtener más información, consulte el tema [Edición de un supernodo bloqueado](#) el p. 515.

Para ver el contenido de un supernodo, puede utilizar el icono acercar de la barra de herramientas de IBM® SPSS® Modeler o el método siguiente:

- ▶ Pulse con el botón derecho en un supernodo.
- ▶ En el menú contextual, seleccione Acercar.

El contenido del supernodo seleccionado se mostrará en un entorno de SPSS Modeler algo diferente, con conectores que muestran el flujo de datos a través de la ruta o del fragmento de ruta. En este nivel del lienzo de rutas, hay varias tareas que puede realizar:

- Modificar el tipo de supernodo: de origen, de proceso o terminal.
- Crear parámetros o editar los valores de un parámetro. Los parámetros se utilizan en procesos y expresiones CLEM.
- Especificar las opciones de almacenamiento en caché para el supernodo y los subnodos.
- Crear o modificar un proceso del supernodo (sólo para supernodos terminales).

## **Modificación de los tipos de Supernodos**

En algunas circunstancias, conviene modificar el tipo de un supernodo. Esta opción está disponible sólo cuando se acerque a un supernodo y sólo se aplica al supernodo en este nivel. Hay tres tipos de supernodos:

<b>Supernodo de origen</b>	Una conexión saliente
<b>Supernodo de proceso</b>	Dos conexiones: una entrante y otra saliente
<b>Supernodo terminal</b>	Una conexión entrante

### **Para cambiar el tipo de un supernodo**

- ▶ Asegúrese de que se ha acercado al supernodo.
- ▶ En el menú Supernodo seleccione Tipo de Supernodo y elija el tipo.

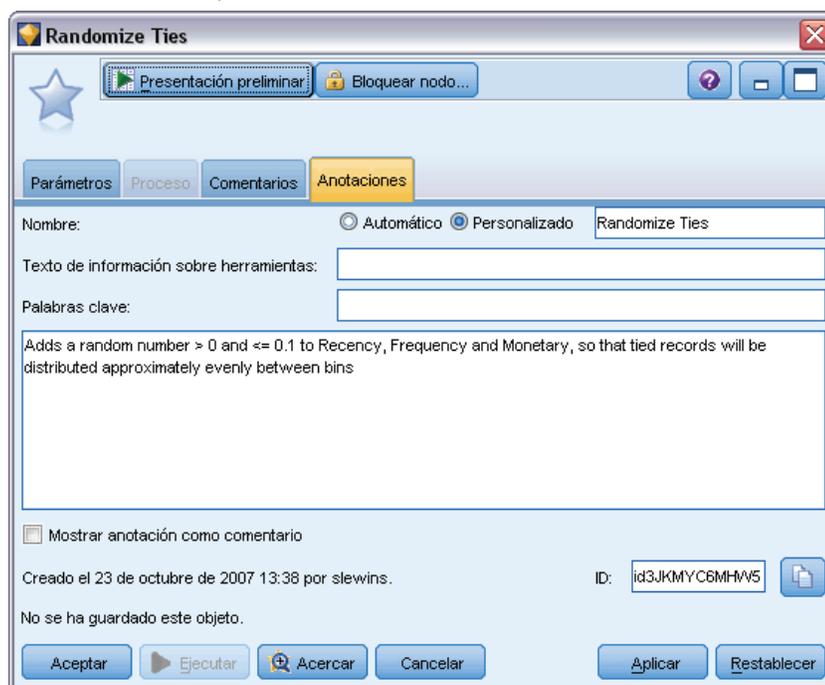
## Anotación y cambio de nombre de Supernodos

Puede cambiar el nombre de un supernodo como aparece en la ruta y realizar anotaciones utilizadas en un proyecto o informe. Para acceder a estas propiedades:

- ▶ Pulse con el botón derecho en un supernodo (alejamiento) y seleccione Cambiar nombre y anotar.
- ▶ Si lo prefiere, en el menú Supernodo seleccione Cambiar nombre y anotar. Esta opción está disponible tanto en el modo de acercamiento como en el de alejamiento.

En ambos casos, se abrirá un cuadro de diálogo con la pestaña Anotaciones seleccionada. Utilice estas opciones para personalizar el nombre mostrado en el lienzo de rutas y proporcionar documentación relacionada con las operaciones del supernodo.

Figura 9-17  
Anotación de un Supernodo



### Uso de comentarios con supernodos

Si crea un supernodo de un nodo o nugget de comentario, debe incluir el comentario en la selección para crear el nodo si desea que aparezca en el supernodo. Si omite el comentario de la selección, el comentario permanecerá desconectado en la ruta cuando cree el supernodo.

Si expande el nodo que incluye los comentarios, los comentarios se quedarán en su lugar antes de que se cree el supernodo.

Si expande un supernodo que incluye objetos comentados, pero los comentarios no se incluyen en el supernodo, los objetos se reinstancian a su ubicación original, sin sus comentarios

## **Parámetros de Supernodos**

En IBM® SPSS® Modeler, puede establecer variables definidas por el usuario, como *Minvalue*, cuyos valores se pueden especificar cuando se utilicen en procesos o en expresiones CLEM. Estas variables se denominan **parámetros**. Puede establecer parámetros para las rutas, las sesiones y los supernodos. Los parámetros establecidos para un supernodo estarán disponibles al generar expresiones CLEM en dicho supernodo o en los nodos anidados. Los parámetros establecidos para los supernodos anidados no están disponibles en el supernodo parental.

Hay dos pasos para crear y configurar los parámetros para los supernodos:

- Defina los parámetros para el supernodo.
- A continuación, especifique el valor para cada parámetro del supernodo.

Estos parámetros se podrán utilizar en expresiones CLEM para los nodos encapsulados.

### **Definición de los parámetros de Supernodos**

Los parámetros de un supernodo se pueden definir tanto en un modo de alejamiento como de acercamiento. Los parámetros definidos se aplican a todos los nodos encapsulados. Para definir los parámetros de un supernodo, primero deberá acceder a la pestaña Parámetros del cuadro de diálogo Supernodo. Utilice uno de los siguientes métodos para abrir el cuadro de diálogo:

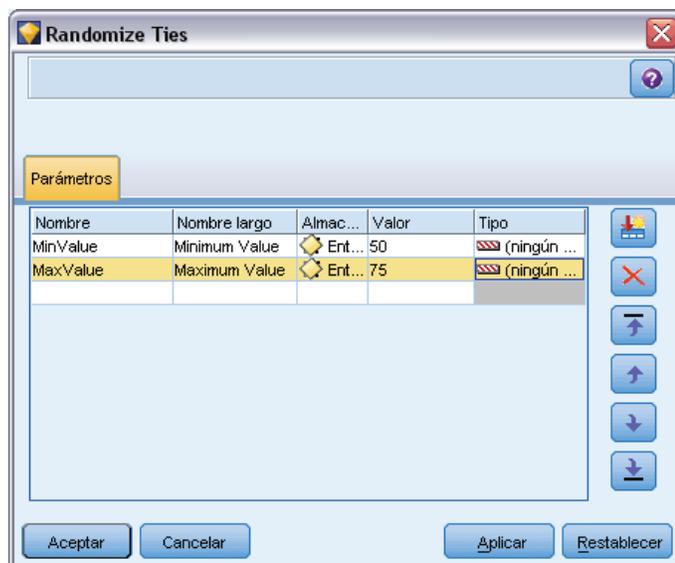
- Pulse dos veces en un supernodo de la ruta.
- En el menú Supernodo, seleccione Definir parámetros.
- Si lo prefiere, cuando se acerque a un supernodo, seleccione Definir parámetros en el menú contextual.

Una vez abierto el cuadro de diálogo, la pestaña Parámetros estará visible con los parámetros definidos previamente.

#### **Para definir un nuevo parámetro**

- ▶ Pulse en el botón Definir parámetros para abrir el cuadro de diálogo.

Figura 9-18  
Definición de parámetros para un supernodo



**Nombre.** Los nombres de los parámetros se enumeran aquí. Se puede crear un parámetro nuevo escribiendo un nombre en este campo. Por ejemplo, si desea crear un parámetro para la temperatura mínima, puede escribir minvalue. No incluya el prefijo \$P- que denota un parámetro en las expresiones CLEM. Este nombre se utiliza para su representación en el generador de expresiones de CLEM.

**Nombre largo.** Indica el nombre descriptivo de cada parámetro creado.

**Almacenamiento.** Seleccione un tipo de almacenamiento de la lista. Indica cómo se almacenan los valores de datos en el parámetro. Por ejemplo, cuando trabaje con valores con ceros al principio que desee conservar (como 008), debe seleccionar Cadena como el tipo de almacenamiento. De lo contrario, los ceros se eliminarán del valor. Los tipos de almacenamiento disponibles son cadena, entero, real, hora, fecha y marca de tiempo. Tenga en cuenta que para los parámetros de fecha, los valores se deben especificar utilizando la notación estándar ISO tal y como se detalla en el siguiente párrafo.

**Valor.** Indica el valor actual para cada parámetro. Ajuste el parámetro como desee. Tenga en cuenta que para los parámetros de fecha, los valores se deben especificar en la notación estándar ISO (que es YYYY-MM-DD). No se aceptan fechas especificadas en otros formatos.

**Tipo (opcional).** Si desea distribuir la ruta en una aplicación externa, seleccione un nivel de medición de la lista. En caso contrario, se recomienda dejar la columna *Tipo* tal como está. Si desea especificar restricciones de valores para el parámetro, como límites superiores e inferiores para un rango numérico, seleccione Especificar en la lista.

Tenga en cuenta que las opciones de tipo, almacenamiento y nombre largo se pueden establecer para los parámetros sólo a través de la interfaz de usuario. Estas opciones no se pueden establecer utilizando procesos.

Pulse en las flechas de la derecha para mover el parámetro seleccionado hacia arriba o hacia abajo en la lista de parámetros disponibles. Utilice el botón de eliminación (marcado con una X) para eliminar el parámetro seleccionado.

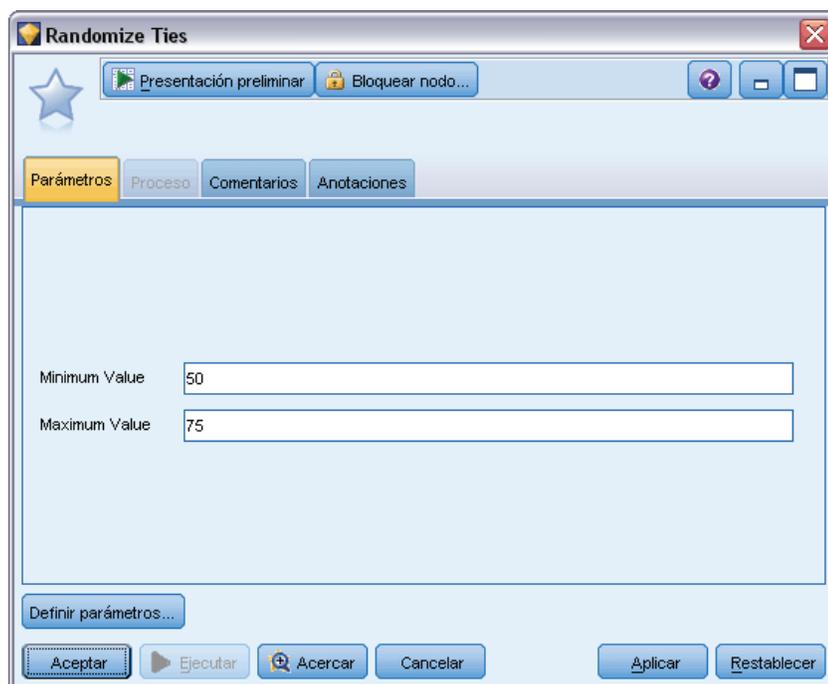
### **Configuración de valores para los parámetros de Supernodo**

Una vez haya definido los parámetros para un supernodo, podrá especificar los valores mediante los parámetros en un proceso o expresión CLEM.

#### **Para especificar los parámetros de un supernodo**

- ▶ Pulse dos veces en el icono Supernodo para abrir el cuadro de diálogo Supernodo.
- ▶ Si lo prefiere, en el menú Supernodo, seleccione Definir parámetros.
- ▶ Pulse en la pestaña Parámetros. *Nota:* los campos de este cuadro de diálogo se definen pulsando en el botón Definir parámetros de esta pestaña.
- ▶ Introduzca un valor en el cuadro de texto para cada parámetro creado. Por ejemplo, puede definir el valor *minvalue* para un umbral determinado de interés. Este parámetro se podrá utilizar en numerosas operaciones, como la selección de registros por encima o debajo de este umbral para realizar una exploración más a fondo.

Figura 9-19  
Especificación de parámetros para un supernodo



### Uso de parámetros de Supernodo para acceder a las propiedades de los nodos

Los parámetros de supernodo también se pueden utilizar para definir las propiedades de los nodos (asimismo denominadas **parámetros de propiedades**) para los nodos encapsulados. Por ejemplo, imagine que desea especificar que un supernodo entrena un nodo Red neuronal encapsulado para un período de tiempo determinado mediante una muestra aleatoria de los datos disponibles. Con los parámetros, podrá especificar los valores para el período de tiempo y la muestra de porcentaje.

Figura 9-20  
Fragmento de ruta encapsulado en un supernodo



Este Supernodo de ejemplo contiene un nodo Muestrear denominado *Sample* y un nodo Red neuronal denominado *Train*. Puede utilizar los cuadros de diálogo del nodo para especificar la configuración **Muestrear** del nodo Muestrear como Aleatorio % y la configuración **Detener cuando** del nodo Red neuronal como Tiempo. Una vez se hayan especificado estas opciones, podrá acceder a las propiedades de los nodos y determinar los valores específicos del supernodo. En el cuadro de diálogo Supernodo, pulse en Definir parámetros y cree los siguientes parámetros:

Figura 9-21  
Definición de los parámetros para acceder a las propiedades de los nodos



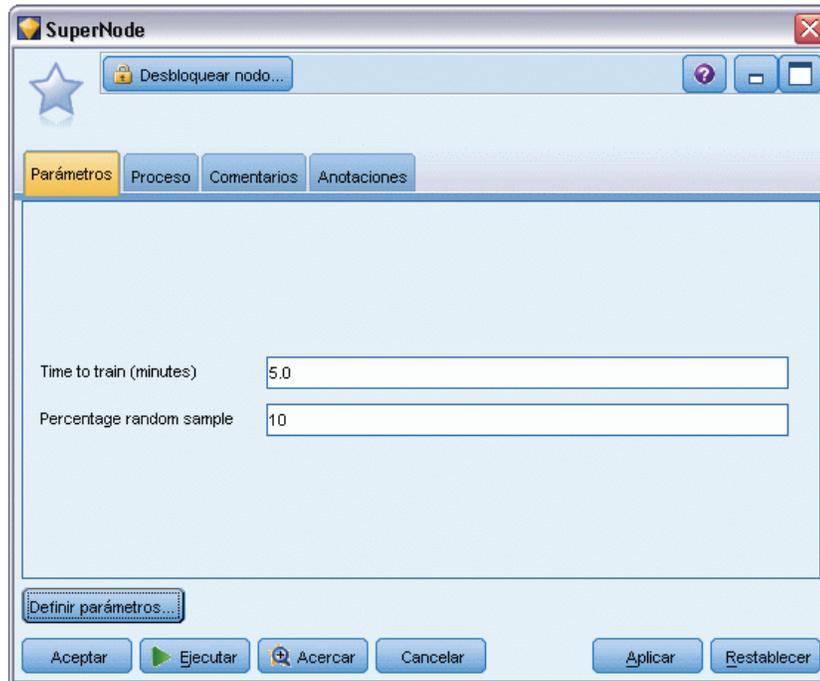
*Nota:* los nombres de los parámetros, como *Sample.random*, utilizan una sintaxis correcta para referirse a las propiedades de los nodos, donde *Sample* representa el nombre del nodo y *random* es la propiedad.

Una vez haya definido estos parámetros, podrá modificar fácilmente los valores para las propiedades de los dos nodos Muestrear y Red neuronal sin necesidad de volver a abrir los cuadros de diálogo. En su lugar, sólo tiene que seleccionar Definir parámetros en el menú Supernodo para acceder a la pestaña Parámetros del cuadro de diálogo Supernodo, donde puede especificar

valores nuevos para Aleatorio % y Tiempo. Esto resulta muy útil a la hora de explorar los datos durante las numerosas iteraciones de la generación del modelo.

Figura 9-22

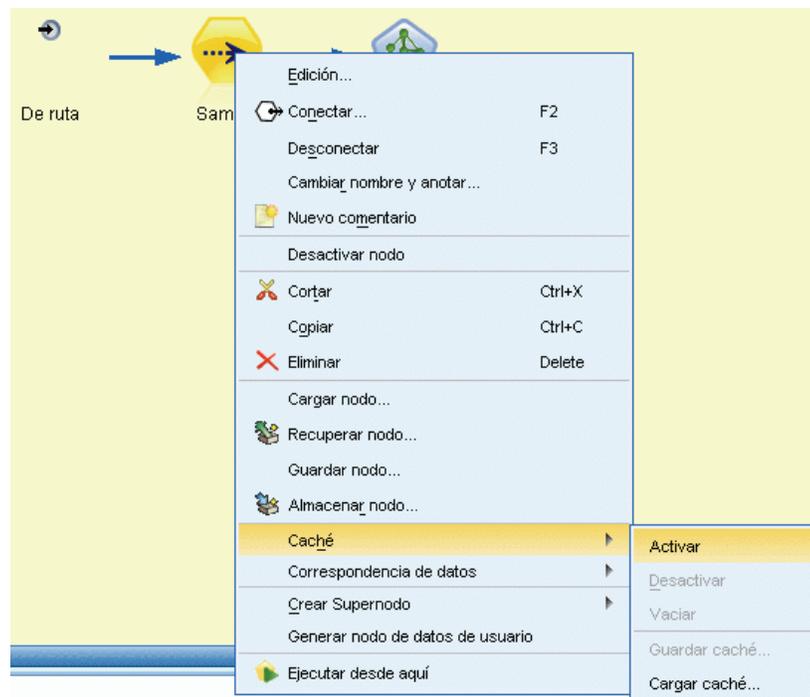
Especificación de valores para las propiedades de los nodos en la pestaña *Parámetros* del cuadro de diálogo *Supernodo*



### ***Supernodos y almacenamiento en caché***

Desde un supernodo, todos los nodos excepto los terminales se pueden almacenar en caché. El almacenamiento en caché se controla pulsando con el botón derecho en un nodo y seleccionando una de las diversas opciones del menú contextual Caché. Esta opción de menú está disponible desde el exterior de un supernodo y para los nodos encapsulados en un supernodo.

Figura 9-23  
Selección de las opciones de almacenamiento en caché para un supernodo



Hay varias directrices para las cachés del supernodo:

- Si hay algún nodo encapsulado en un supernodo que tenga el almacenamiento en caché activado, el supernodo también lo tendrá.
- La desactivación de la caché en un supernodo desactiva la caché de *todos* los nodos encapsulados.
- La activación del almacenamiento en caché en un supernodo, activa realmente la caché en el último subnodo en el que se puede realizar dicho almacenamiento. Es decir, si el último subnodo es un nodo Seleccionar, la caché se activará para dicho nodo. Si el último subnodo es un nodo terminal (que no permite el almacenamiento en caché), el siguiente nodo por encima que admita el almacenamiento en caché, estará activado.
- Una vez haya definido las cachés para los subnodos de un supernodo, las actividades por encima del nodo de cachés, como añadir o editar nodos, vaciarán las cachés.

## Supernodos y procesos

Puede usar el lenguaje de procesamiento de IBM® SPSS® Modeler para escribir programas simples que manipulen y ejecuten el contenido de un supernodo terminal. Por ejemplo, puede que le resulte útil especificar el orden de ejecución de una ruta compleja. A modo de ejemplo, si un supernodo contiene un nodo Val. globales que debe ejecutarse antes que un nodo Gráfico, podrá crear un proceso que ejecute primero el nodo Val. globales. Los valores calculados por este nodo, como la media o la desviación estándar, se pueden usar posteriormente cuando se ejecute el nodo Gráfico.

La pestaña Proceso del cuadro de diálogo Supernodo sólo estará disponible para los supernodos terminales.

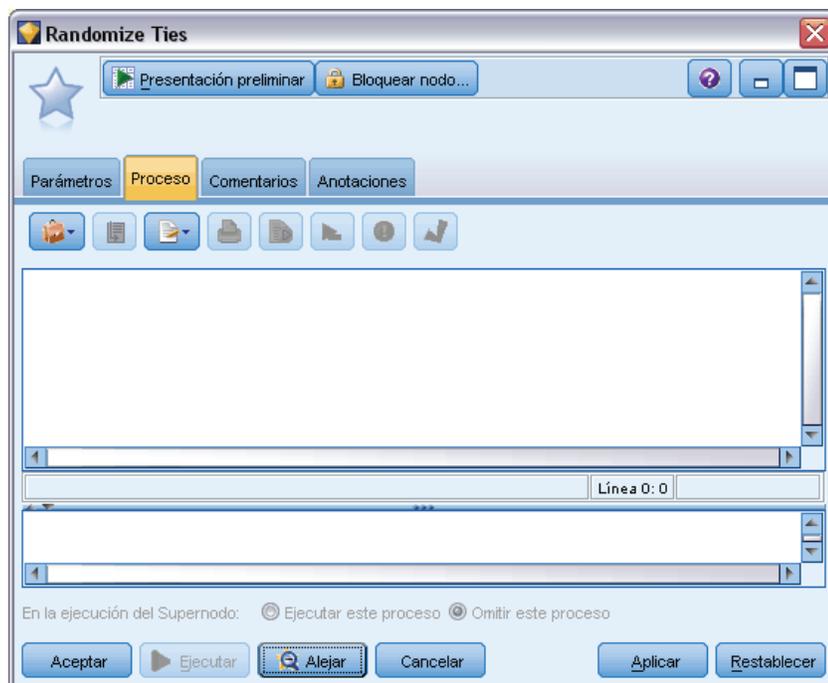
**Para abrir el cuadro de diálogo de procesos para un supernodo terminal**

- ▶ Pulse con el botón derecho en el lienzo del supernodo y seleccione Proceso de Supernodo.
- ▶ También puede seleccionar, tanto en el modo de acercamiento como en el de alejamiento, la opción Proceso de Supernodo del menú Supernodo.

*Nota:* los procesos de supernodo sólo se ejecutan con la ruta y el supernodo cuando haya seleccionado Ejecutar este proceso en el cuadro de diálogo.

Figura 9-24

Creación de un proceso para un supernodo



Las opciones específicas de los procesos y su uso en SPSS Modeler se describen en la *Guía de procesos y automatización*, disponible en SPSS ModelerDVD.

## Almacenamiento y carga de supernodos

Una de las ventajas de los supernodos es que se pueden guardar y reutilizar en otras rutas. Al guardar y cargar supernodos, tenga en cuenta que usan una extensión *.slb*.

**Para guardar un supernodo**

- ▶ Acérquese al supernodo.
- ▶ En el menú Supernodo, seleccione Guardar Supernodo.

- ▶ Especifique un nombre de archivo y directorio en el cuadro de diálogo.
- ▶ Seleccione si desea añadir el supernodo guardado al proyecto actual.
- ▶ Pulse en Guardar.

***Para cargar un supernodo***

- ▶ En el menú Insertar de la ventana de IBM® SPSS® Modeler, seleccione Supernodo.
- ▶ Seleccione un archivo de Supernodo (.slb) en el directorio actual o examine uno diferente.
- ▶ Pulse en Cargar.

*Nota:* los supernodos importados tiene los valores por defecto para todos los parámetros. Para cambiar los parámetros, pulse dos veces en un supernodo del lienzo de rutas.

# Avisos

Esta información se ha desarrollado para los productos y servicios ofrecidos en todo el mundo.

Puede que IBM no ofrezca los productos, los servicios o las características de los que se habla en este documento en otros países. Consulte a su representante local de IBM para obtener información acerca de los productos y servicios que está disponibles actualmente en su zona. Toda referencia que se haga de un producto, programa o servicio de IBM no implica que sólo se deba utilizar ese producto, programa o servicio de IBM. En su lugar, puede utilizarse todo producto, programa o servicio con funcionalidades equivalentes que no infrinjan los derechos de propiedad intelectual de IBM. Sin embargo, es responsabilidad del usuario evaluar y comprobar el funcionamiento de todo producto, programa o servicio que no sea de IBM.

IBM puede tener patentes o aplicaciones de patentes pendientes que cubren el asunto descrito en este documento. Este documento no le otorga ninguna licencia para estas patentes. Puede enviar preguntas acerca de las licencias, por escrito, a:

*IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, Estados Unidos*

Si tiene alguna pregunta sobre la licencia relacionada con la información del juego de caracteres de doble byte (DBCS), póngase en contacto con el departamento de propiedad intelectual de IBM de su país o envíe sus preguntas por escrito a:

*Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Japan.*

**El párrafo siguiente no se aplica a los Reino Unido o cualquier otro país donde tales disposiciones son incompatibles con la legislación local:** INTERNATIONAL BUSINESS MACHINES PROPORCIONA ESTA PUBLICACIÓN “TAL CUAL” SIN GARANTÍA DE NINGÚN TIPO, YA SEA EXPRESA O IMPLÍCITA, INCLUYENDO, PERO NO LIMITADA A, LAS GARANTÍAS IMPLÍCITAS DE NO INFRACCIÓN, COMERCIALIZACIÓN O IDONEIDAD PARA UN PROPÓSITO PARTICULAR. Algunos estados no permiten la renuncia a expresar o a garantías implícitas en determinadas transacciones, por lo tanto, esta declaración no se aplique a usted.

Esta información puede incluir imprecisiones técnicas o errores tipográficos. Periódicamente, se efectúan cambios en la información aquí y estos cambios se incorporarán en nuevas ediciones de la publicación. IBM puede realizar mejoras y/o cambios en los productos y/o los programas descritos en esta publicación en cualquier momento sin previo aviso.

Cualquier referencia a sitios Web que no sean de IBM en esta información sólo es ofrecida por comodidad y de ningún modo sirve como aprobación de esos sitios Web. Los materiales en esos sitios Web no forman parte del material de este producto de IBM y el uso de estos sitios Web es bajo su propio riesgo.

IBM puede utilizar cualquier información que le suministre en cualquier forma que considere adecuada, sin incurrir en ninguna obligación para usted.

Los licenciatarios de este programa que deseen tener información sobre el mismo con el objetivo de habilitar: (i) el intercambio de información entre programas creados independientemente y otros programas (incluyendo este) y (ii) el uso común de la información que se ha intercambiado, deben ponerse en contacto con:

*IBM Software Group, a la atención de: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.*

Esta información estará disponible, bajo las condiciones adecuadas, incluyendo en algunos casos el pago de una cuota.

IBM proporciona el programa bajo licencia que se describe en este documento y todo el material bajo licencia disponible para el mismo bajo los términos de IBM Customer Agreement (Acuerdo de cliente de IBM), IBM International Program License Agreement (Acuerdo de licencia de programa internacional de IBM) o cualquier acuerdo equivalente entre las partes.

Cualquier dato de rendimiento mencionado aquí ha sido determinado en un entorno controlado. Por lo tanto, los resultados obtenidos en otros entornos operativos pueden variar de forma significativa. Es posible que algunas medidas se hayan realizado en sistemas en desarrollo y no existe ninguna garantía de que estas medidas sean las mismas en los sistemas comerciales. Además, es posible que algunas medidas hayan sido estimadas a través de extrapolación. Los resultados reales pueden variar. Los usuarios de este documento deben consultar los datos que corresponden a su entorno específico.

Se ha obtenido información acerca de productos que no son de IBM de los proveedores de esos productos, de sus publicaciones anunciadas o de otras fuentes disponibles públicamente. IBM no ha probado estos productos y no puede confirmar la precisión de su rendimiento, su compatibilidad o cualquier otra reclamación relacionada con productos que no sean de IBM. Las preguntas acerca de las aptitudes de productos que no sean de IBM deben dirigirse a los proveedores de dichos productos.

Todas las declaraciones sobre el futuro del rumbo y la intención de IBM están sujetas a cambio o retirada sin previo aviso y representan únicamente metas y objetivos.

Esta información contiene ejemplos de datos e informes utilizados en operaciones comerciales diarias. Para ilustrarlos lo máximo posible, los ejemplos incluyen los nombres de las personas, empresas, marcas y productos. Todos esos nombres son ficticios y cualquier parecido con los nombres y direcciones utilizados por una empresa real es pura coincidencia.

Si está viendo esta información en copia electrónica, es posible que las fotografías y las ilustraciones en color no aparezcan.

### ***Marcas comerciales***

IBM, el logotipo de IBM, [ibm.com](http://ibm.com) y SPSS son marcas comerciales de IBM Corporation, registradas en muchas jurisdicciones de todo el mundo. Existe una lista actualizada de marcas comerciales de IBM en Internet en <http://www.ibm.com/legal/copytrade.shtml>.

Intel, el logotipo de Intel, Intel Inside, el logotipo de Intel Inside, Intel Centrino, el logotipo de Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium y Pentium son marcas comerciales o marcas registradas de Intel Corporation o sus filiales en Estados Unidos y otros países.

Linux es una marca registrada de Linus Torvalds en Estados Unidos, otros países o ambos.

Microsoft, Windows, Windows NT, y el logotipo de Windows son marcas comerciales de Microsoft Corporation en Estados Unidos, otros países o ambos.

UNIX es una marca registrada de The Open Group en Estados Unidos y otros países.

Java y todas las marcas comerciales y los logotipos basados en Java son marcas comerciales de Sun Microsystems, Inc. en Estados Unidos, otros países o ambos.

Otros productos y nombres de servicio pueden ser marcas comerciales de IBM u otras empresas.



---

# Índice

- aciertos
  - opciones del diagrama de evaluación, 348
- actualidad
  - configuración de fecha relativa, 86
- acumulación de datos de series temporales, 219
- adición
  - registros, 82
- adición de paneles, 244, 246
- administrador de resultados, 392
- administradores
  - pestaña Resultados, 392
- agregación de datos de series temporales, 219
- agregación de registros, 208
- agrupación supervisada, 197
- ajustes -automáticos-, 369
- almacenamiento, 143
  - conversión, 177, 179
  - objetos de resultados, 392, 399
  - resultados, 393
- almacenamiento de campos
  - conversión, 177
- ancho de columna
  - para campos, 151
- animación
  - en visualizaciones, 247
- animación en gráficos, 244, 246
- año fiscal
  - nodo Intervalos de tiempo, 227
- anonimización nombres de campos, 158
- ANOVA
  - nodo Medias, 437
- ANOVA unidireccional
  - nodo Medias, 437
- anti-uniión, 90
- apertura
  - objetos de resultados, 392
- archivo de datos *employee\_data.sav*, 489
- archivos (SAS) *.sd2*, 51
- archivos (SAS) *.ssd*, 51
- archivos (SAS) *.tpt*, 51
- archivos *.sav*, 487
- archivos *.slb*, 524
- archivos DAT
  - almacenamiento, 399
  - exportación, 396, 480
- Archivos de datos de IBM SPSS Statistics
  - importación de datos de encuesta, 37
- archivos de Excel
  - exportación, 480
- archivos de formato, 52
- archivos de mapas
  - cambio del nombre, 289
  - eliminación, 289
  - exportación, 289
  - importación, 289
  - seleccionar en el selector de plantillas de tablero, 258
  - ubicación, 287
- archivos de resultados
  - almacenamiento, 399
- archivos de texto, 25
  - exportación, 480
- archivos de transporte
  - Nodo de fuente SAS, 51
- archivos delimitados por comas
  - almacenamiento, 399
  - exportación, 396, 480
- archivos ESRI, 290
- archivos planos, 25
- archivos SMZ
  - cambio del nombre, 289
  - conceptos básicos, 290
  - creación, 290
  - edición de archivos SMZ preinstalados, 290
  - eliminación, 289
  - exportación, 289
  - importación, 289
  - preinstalados, 290
- archivos XLS
  - exportación, 480
- asignación de tipos de datos, 65, 106, 135
- aspectos estéticos
  - en visualizaciones, 244
- atributos
  - en mapas, 291
- atributos de campo, 150
- atributos de tipo, 150
- auditoría
  - auditoría de datos iniciales, 412
  - Nodo Auditar datos, 412
- avisos legales, 526
  
- base de datos
  - carga masiva, 461, 464
  - niveles de soporte, 15
- bases de datos de ADO
  - importación, 37
- bases de datos de gran tamaño, 69
  - realización de una auditoría de datos, 412
- bases de datos de In2data
  - importación, 37
- bases de datos de Quanvert
  - importación, 37
- bloqueo de supernodos, 513–514
- borrar valores, 65
- búsqueda
  - explorador de tablas, 401

- caché
  - Supernodos, 522
- calcular duraciones
  - preparación automática de datos, 113
- cálculo de duraciones
  - preparación automática de datos, 113
- calidad de datos
  - explorador de auditoría de datos, 421
- cambio del nombre
  - archivos de mapas, 289
  - campos para exportar, 503
  - hojas de estilo de visualización, 289
  - plantillas de visualización, 289
- cambio del nombre de los objetos de resultados., 392
- Campo EtiquetaTiempo
  - nodo Intervalos de tiempo, 219
- Campo ÍndiceTiempo
  - nodo Intervalos de tiempo, 219
- campo Recuento
  - nodo Intervalos de tiempo, 221
  - series temporales de agregación o relleno, 221
- campos
  - anonimización datos, 180
  - delimitadores, 28
  - derivación de múltiples campos, 167
  - etiquetas de valor y de campo, 65, 135, 145
  - reordenación, 239
  - selección múltiple, 169
  - transposición, 211–212
- campos clave, 82, 208
- campos claves primarios
  - nodo Exportar base de datos, 457
- campos de correspondencia, 453
- campos de partición, 65, 135, 149, 205–206
- caracteres de comentario
  - en archivos variables, 27
- caracteres de EOL, 27
- carga masiva, 461, 464
- casos de rangos, 195
- casos reservados
  - modelado de series temporales, 221
- chi-cuadrado
  - nodo Matriz, 406
- chi-cuadrado de Pearson
  - nodo Matriz, 406
- cifras decimales
  - formatos de presentación, 153
- clave de recuento para la agregación, 82
- claves contiguas, 82
- codificación, 28, 31, 473
- codificación de variables dummy, 207
- codificación UTF-8, 28, 31, 473
- Cognos, consulte IBM Cognos BI, 49
- colores
  - en visualizaciones, 244
- coma, 28, 151
- comando CREATE INDEX, 459
- combinación de datos, 100
  - de varios archivos, 89
- comentarios
  - uso con supernodos, 517
- comillas
  - importación de archivos de texto, 29
  - para exportar base de datos, 451
- complementos, 448
- comprobación de tipos, 148
- condiciones
  - especificación de una serie, 175
  - especificaciones para una fusión, 94
- conexiones
  - al IBM SPSS Collaboration and Deployment Services Repository, 9
- conexiones a la base de datos
  - definición, 18
  - valores predefinidos, 19
- conglomerado, 383
- conjunto de datos principal, 101
- conjunto de respuestas múltiples;respuestas múltiples
  - conjuntos de categorías múltiples, 158
  - conjuntos de dicotomía múltiple, 158
  - definición, 158
  - eliminación, 158
  - en visualizaciones, 254
  - nodo de fuente Data Collection, 35–36
  - nodo de fuente IBM SPSS Data Collection, 42–43
  - nodo de origen IBM SPSS Statistics, 488
- conjuntos
  - conversión a marcas, 207, 209
  - transformación, 185, 188
- conjuntos de categorías múltiples, 158
- conjuntos de dicotomía múltiple, 158
- consultas
  - Nodo de fuente Base de datos, 15–16
- Consultas SQL
  - Nodo de fuente Base de datos, 15–16, 23–24
- conversión de conjuntos a marcas, 207, 209
- convertir niveles de medición, 140
- coordenadas polares, 379
- copia de atributos de tipo, 150
- copia de visualizaciones, 383
- coropleta
  - ejemplo, 281
- correlaciones, 433
  - etiquetas descriptivas, 433
  - nodo Medias, 442
  - probabilidad, 433
  - resultado de estadísticos, 434
  - significación, 433
  - valor absoluto, 433
- correlaciones de Pearson
  - nodo Medias, 442
  - resultado de estadísticos, 434
- costes
  - diagramas de evaluación, 348

- creación
  - campos nuevos, 165–166
- creación de facetas
  - en visualizaciones, 246
- creación de una instancia, 65, 135, 137, 141–142
  - nodo de fuente, 67
- CRISP-DM
  - comprensión de los datos, 8
- cuantiles
  - nodo Intervalos, 192
  
- datos
  - agregación, 82
  - almacenamiento, 177, 179
  - anonimización, 180
  - auditoría, 412
  - comprensión, 69
  - exploración, 412
  - preparación, 69
  - storage, 33, 63
  - tipo de almacenamiento, 143
- datos anuales
  - nodo Intervalos de tiempo, 226
- datos categóricos, 138, 140
- datos continuos, 138, 140, 145
- Datos CSV
  - importación, 37
- datos de casos
  - nodo de fuente Data Collection, 35–36
- datos de encuesta
  - importación, 36, 42–43
  - nodo de fuente Data Collection, 35
- datos de encuesta Data Collection
  - importación, 35–36
- datos de investigación de mercados
  - importación, 36, 43
  - nodo de fuente Data Collection, 35
  - nodo de fuente IBM SPSS Data Collection, 42
- Datos de marca, 138
- datos de muestreo, 78
- datos de Quancept
  - importación, 37
- datos de Quantum
  - importación, 37
- datos de resumen, 82
- datos de serie temporal
  - agregación, 216, 219
  - casos reservados, 221
  - definición, 216–217, 219, 221
  - etiquetado, 216–217, 219, 221
  - generación a partir de datos, 219
  - intervalos, 217
  - período de estimación, 221
  - relleno, 216, 219
- datos de Surveycraft
  - importación, 37
- datos de texto de campo fijo, 29
- datos de texto de campo libre, 25
- datos de texto delimitado, 25
- datos decrecientes, 70, 72
- datos desequilibrados, 80
- datos insesgados, 80
- datos mensuales
  - nodo Intervalos de tiempo, 228
- datos nominales, 138, 146
- datos ordinales, 138, 146
- datos retardados, 237
- datos semanales
  - nodo Intervalos de tiempo, 229
- datos sesgados, 80
- datos sin tipo, 139
- datos sintéticos
  - nodo de datos de usuario, 58
- datos trimestrales
  - nodo Intervalos de tiempo, 227
- Datos triple-S
  - importación, 37
- definición automática de tipos, 139, 142
- Definiciones de proveedor de datos, 9
- delimitadores, 27–28, 461
- densidad
  - 3-D, 263
- Densidad 3-D, 263
- derivación múltiple, 167
- desbloqueo de supernodos, 514
- descarte
  - campos, 154
- desviación típica
  - nodo Intervalos, 196
  - nodo Val. globales, 447
  - resultado de estadísticos, 434
- desviación típica para la agregación, 82
- diagrama de dispersión, 261
  - 3-D, 262
  - en intervalos, 262
  - intervalos hexagonales, 262
- diagrama de dispersión en intervalos, 262
  - intervalos hexagonales, 262
- diagrama de dispersión en intervalos hexagonales, 262
- diagramas de dispersión, 299, 322
- Diagramas de dispersión 3-D, 262
- dibujos de asociaciones, 326
- dirección de campos, 65, 135, 149
- diseño de red para gráficos Malla, 331
- diseño direccional para gráficos Malla , 331
- distribución, 313
- documentación, 4
- documentos MDD
  - importación, 37
- DPD, 9
- duplicados
  - campos, 89, 155
  - registros, 102

- edición de gráficos
  - tamaño de los elementos gráficos, 372
- edición de visualizaciones, 367
  - adición de efectos 3-D, 379
  - categorías, 376
  - colores y tramas, 370
  - combinación de categorías, 376
  - configuración automática, 368
  - contracción de categorías, 376
  - ejes, 374
  - escalas, 374
  - estilo de guión, 370
  - exclusión de categorías, 376
  - forma de puntos, 371
  - formatos de numeración, 373
  - márgenes, 373
  - ordenación de categorías, 376
  - paneles, 378
  - posición de la leyenda, 383
  - reglas, 368
  - relación de aspecto de puntos, 371
  - relleno, 373
  - rotación de puntos, 371
  - texto, 369
  - transformación de los sistemas de coordenadas, 379
  - transparencia, 370
  - transponer, 378–379
  - única, 368
- editor de consultas
  - Nodo de fuente Base de datos, 23–24
- ejecución
  - especificación del orden de, 523
- ejemplos
  - conceptos básicos, 6
  - Manual de aplicaciones, 4
- ejemplos de aplicaciones, 4
- elementos de hora cíclicos
  - preparación automática de datos, 113
- elementos gráficos
  - cambio, 380
  - conversión, 380
  - modificadores de colisión, 382
  - tipos, 380
- eliminación
  - archivos de mapas, 289
  - hojas de estilo de visualización, 289
  - objetos de resultados, 392
  - plantillas de visualización, 289
- empates
  - nodo Intervalos, 192
- encapsulación de nodos, 508
- enlace a lo largo de las columnas, 461
- enlace a lo largo de las filas, 461
- enlaces
  - nodo Malla, 329
- entradas múltiples, 89
- error típico de la media
  - resultado de estadísticos, 434
- escenario, 9
- esquema
  - nodo Exportar base de datos, 455
- esquivar, 383
- establecer semilla aleatoria
  - registros de muestreo, 76, 207
- estadístico de la evaluación del rendimiento, 408
- estadístico *F*
  - nodo Medias, 441
- estadísticos
  - descripciones, 253, 381
  - edición de visualizaciones, 380
  - Nodo Auditar datos, 412
  - nodo Matriz, 403
- estadísticos de resumen
  - Nodo Auditar datos, 412
- etiquetas, 89, 97, 147
  - en visualizaciones, 245
  - especificación, 65, 135, 143, 145–147
  - exportación, 479, 503
  - importación, 53, 487
- etiquetas de campo
  - etiquetado de registros en el resultado, 149
- etiquetas de datos
  - en visualizaciones, 245
- etiquetas de valor
  - Nodo Archivo Statistics, 487
- etiquetas de variable
  - Nodo Archivo Statistics, 487
  - Nodo Exportación de Statistics, 501
- evaluación de modelos, 407
- evaluación del modelo, 341
- eventos
  - creación, 338
- Excel
  - inicio desde IBM SPSS Modeler, 480
- exclusión de campos no utilizados
  - preparación automática de datos, 112
- exploración de datos
  - Nodo Auditar datos, 412
- exploración de gráficos, 355
  - marcación de elementos, 363
  - regiones, 360
  - tramos de gráficos, 356
  - varita mágica, 363
- explorador de análisis
  - interpretación, 410
- explorador de auditoría de datos
  - generación de gráficos, 426
  - generación de nodos, 426
  - Menú Archivo, 417
  - Menú Edición, 417
- explorador de calidad
  - generación de nodos Filtro, 424
  - generación de nodos Seleccionar, 425

- explorador de estadísticos
  - generación de nodos Filtro, 436
  - interpretación, 434
  - Menú Generar, 434
- explorador de informes, 445
- explorador de matrices
  - Menú Generar, 406
- explorador de tablas
  - búsqueda, 401
  - Menú Generar, 401
  - reordenación de columnas, 397, 401
  - selección de casillas, 397, 401
- exportación
  - archivos de mapas, 289
  - hojas de estilo de visualización, 289
  - plantillas de visualización, 289
  - resultados, 396
  - Supernodos, 524
- exportación de cifras decimales, 153
- exportación de datos
  - a Excel, 480
  - a una base de datos, 451
  - al IBM SPSS Statistics, 501
  - archivos DAT, 480
  - formato de archivo plano, 472
  - Formato SAS, 479
  - Formato XML, 481
  - nodo Exportar de IBM Cognos BI, 49, 475, 477
  - texto, 480
- expresiones CLEM, 69
- extensión
  - campo derivado, 167
  
- factores de equilibrado, 81
- factores de escala, 81
- fecha/hora, 137
- fechas
  - formatos de configuración, 151, 153
- filas en blanco
  - archivos de Excel, 53
- filtro de campos, 95, 154
  - de IBM SPSS Statistics, 503
- forma
  - en visualizaciones, 245
- formato de almacenamiento de cadenas, 33, 63
- formato de almacenamiento de enteros, 33, 63
- formato de almacenamiento de fechas, 33, 63
- formato de almacenamiento de horas, 33, 63
- formato de almacenamiento de marcas de tiempo, 33, 63
- formato de almacenamiento real, 33, 63
- formato de presentación científica, 153
- formato de presentación de las monedas, 153
- Formato HDATA
  - nodo de fuente Data Collection, 35
- Formato VDATA
  - nodo de fuente Data Collection, 35
- formatos
  - datos, 32, 151
- formatos de almacenamiento, 32
- formatos de hora, 153
- formatos de presentación
  - científica, 153
  - cifras decimales, 153
  - moneda, 153
  - números, 153
  - símbolo de agrupación, 153
- formatos de presentación de los números, 153
- formatos de resultados, 399
- formatos shapefile, 290
- fórmula de derivación de campo, 169
- forzado de valores, 148
- frecuencias
  - nodo Intervalos, 192
- fuentes de datos
  - conexiones a la base de datos, 18
- función de los más recientes
  - series temporales de relleno, 221
- función de media de los más recientes
  - series temporales de relleno, 221
- función hassubstring, 172
- función Máx
  - agregación de series temporales, 220
- función Media
  - agregación de series temporales, 220
- función Mín
  - agregación de series temporales, 220
- función Moda
  - agregación de series temporales, 220
- función Primeros
  - agregación de series temporales, 220
- función Suma
  - agregación de series temporales, 220
- función Último
  - agregación de series temporales, 220
- función Vacío
  - series temporales de relleno, 221
- función Verdadero si alguno es verdadero
  - agregación de series temporales, 220
- funciones
  - en mapas, 291
- fusión de orden, 89
  
- generación de marcas, 208, 210
- generación de nodos a partir de gráficos, 364
  - nodos Derivar, 366
  - Nodos Equilibrar, 366
  - Nodos Filtro, 366
  - Nodos Reclasificar, 366
  - nodos Seleccionar, 365
- Generador de expresiones, 69
- gestión de valores perdidos, 106
- grados de libertad
  - nodo Matriz, 406

- nodo Medias, 441–442
- gráfico de áreas, 261
  - 3-D, 261
- Gráfico de áreas 3D
  - descripción, 261
- gráfico de barras, 260
  - 3-D, 260
  - de recuentos, 260, 265
  - de un mapa, 265
  - ejemplo, 267, 269
- Gráfico de barras 3-D, 260
- gráfico de barras acumuladas, 261
- gráfico de barras agrupadas
  - ejemplo, 269
- gráfico de burbujas, 261
- gráfico de caja, 263
  - ejemplo, 274
- gráfico de coordenadas paralelas, 264
- gráfico de líneas, 261
  - de un mapa, 265
- gráfico de puntos, 263
  - 2-D, 263
  - ejemplo, 272
- gráfico de puntos 2-D, 263
- gráfico de ruta, 261
- gráfico de sectores, 260
  - 3-D, 260
  - de un mapa, 265
  - ejemplo, 276
  - uso de recuentos, 260, 265
- Gráfico de sectores 3-D, 260
- gráfico de superficie, 261
- gráficos
  - 3-D, 250
  - almacenamiento, 389
  - almacenamiento de diseños modificados, 386
  - almacenamiento de modificaciones de diseño, 386
  - almacenamiento de resultados, 399
  - colecciones, 317
  - copia, 389
  - de tablero, 251
  - diagramas de evaluación, 341
  - distribuciones, 308
  - eliminación de regiones, 362
  - esquema de color por defecto, 386
  - etiquetas de eje, 384
  - exploración, 355
  - exportación, 389
  - generación de nodos, 364
  - generación desde una auditoría de datos, 426
  - gráfico múltiple, 322
  - gráficos, 299
  - histogramas, 313
  - hoja de estilo, 386
  - impresión, 389
  - mallas, 326
  - nota al pie, 384
  - pestaña anotaciones, 250
  - pestaña resultados, 249
  - regiones, 360
  - rotación de una imagen 3D, 250
  - tamaño de los elementos gráficos, 372
  - título, 384
  - tramos, 356
  - variables de series temporales, 338
- gráficos 3D, 250
- gráficos de asociación, 326
- gráficos de beneficios, 341, 352
- gráficos de elevación, 341, 352
- gráficos de ganancias, 341, 352
- gráficos de líneas, 299, 322
- gráficos de puntos, 299, 322
- gráficos de respuestas, 341, 352
- graphboard
  - tipos de gráficos, 259
- histograma, 262
  - 3-D, 263
  - ejemplo, 271
- histograma 3-D, 263
- hojas de cálculo
  - importación desde Excel, 53
- hojas de estilo
  - cambio del nombre, 289
  - eliminación, 289
  - exportación, 289
  - importación, 289
- hojas de estilo de visualización
  - aplicar, 387
  - cambio del nombre, 289
  - eliminación, 289
  - exportación, 289
  - importación, 289
  - ubicación, 287
- hora
  - formatos de configuración, 151
- HTML
  - almacenamiento de resultados, 400
- IBM SPSS Collaboration and Deployment Services Repository
  - conexión con, 9
  - uso como ubicación de plantillas de visualización, hojas de estilo y mapas, 288
- IBM SPSS Modeler, 1
  - documentación, 4
- IBM SPSS Statistics
  - inicio desde IBM SPSS Modeler, 448, 497, 502
  - nombres de campos válidos, 503
  - ubicación de la licencia, 448
- iconos, IBM Cognos BI, 45
- igual frecuencia
  - nodo Intervalos, 192

- importación
  - archivos de mapas, 289
  - datos de IBM Cognos BI, 45
  - hojas de estilo de visualización, 289
  - informes de IBM Cognos BI, 47
  - plantillas de visualización, 289
  - Supernodos, 524
- importancia
  - comparación de medias, 439
  - nodo Medias, 441–442
- impresión de resultados, 393
- incrementos por minuto
  - nodo Intervalos de tiempo, 233–234
- incrementos por segundo
  - nodo Intervalos de tiempo, 235–236
- índices BITMAP
  - tablas de bases de datos, 460
- indización de tablas de bases de datos, 459
- informe de calidad
  - explorador de auditoría de datos, 421
- informes
  - almacenamiento de resultados, 399
- ingresos
  - diagramas de evaluación, 348
- instrucciones si-entonces-en caso contrario, 176
- intervalos
  - datos de serie temporal, 216
- intervalos de confianza
  - nodo Medias, 441–442
- intervalos de cuartil, 192
- intervalos de decil, 192
- intervalos de percentil, 192
- intervalos de quintil, 192
- intervalos de veintil, 192
- intervalos óptimos, 197
- intervenciones
  - creación, 338
  
- jitter, 383
- justificación
  - para campos, 151
  
- lengua
  - nodo de fuente IBM SPSS Data Collection, 40
- leyenda
  - posición, 383
- línea base
  - opciones del diagrama de evaluación, 347
  
- mapa
  - colores, 264–265
  - con flechas, 266
  - con gráficos de barras, 265
  - con gráficos de líneas, 265
  - con gráficos de sectores, 265
  - con puntos, 265–266
  - superponer, 266
- mapa de calor, 263
  - ejemplo, 277
- mapa de color, 264–265
  - ejemplo, 281
- mapa de coordenadas, 265–266
- mapa de coropletas, 264–265
- mapa de flujos, 266
- mapa de superposición, 266
- mapas
  - combinación de funciones, 296
  - conversión de los formatos ESRI shapefile, 290
  - distribución, 299
  - eliminación de elementos individuales, 297
  - eliminación de funciones, 297
  - etiquetas de funciones, 295
  - mover funciones, 297
  - proyección, 298
  - reducir, 293–294
  - suavizar, 293–294
- mapas en formato shapefile
  - con la utilización del selector de plantillas de tablero, 290
  - conceptos, 291
  - edición de mapas SMZ preinstalados, 290
  - tipos, 291
- marca de tiempo, 137
- marcación de elementos, 360, 363
- marcas comerciales, 527
- marco de muestreo, 72
- matriz de coincidencias
  - Nodo Análisis, 408
- matriz de diagramas de dispersión
  - ejemplo, 279, 283
- matriz de diagramas de dispersión (SPLOM), 262
- máximo
  - nodo Val. globales, 447
  - resultado de estadísticos, 434
- media
  - nodo Intervalos, 196
  - nodo Val. globales, 447
  - resultado de estadísticos, 434
- media/desviación típica
  - uso para segmentar campos, 196
- mediana
  - resultado de estadísticos, 434
- medias
  - comparación, 437–438, 440
- medidas por días
  - nodo Intervalos de tiempo, 230–231
- medidas por horas
  - nodo Intervalos de tiempo, 231–232
- mejor línea
  - opciones del diagrama de evaluación, 347
- metadatos, 65, 135, 143
  - nodo de fuente Data Collection, 35–36
- método clave, 89

- miembro (importar SAS)
  - configuración, 52
- mínimo
  - nodo Val. globales, 447
  - resultado de estadísticos, 434
- mode
  - resultado de estadísticos, 434
- modelo del proceso CRISP-DM
  - preparación de datos, 106
- modelos
  - anonimizar datos para, 180
- Modelos de IBM SPSS Statistics, 493
  - acerca de, 493
  - detalles avanzados de nugget, 495
  - nugget de modelo, 495
  - opciones de modelo, 494
- modificación de valores de datos, 165
- modificadores de colisión, 380
- muestras conglomeradas, 72–73, 77
- muestras de comprobación
  - partición de datos, 205–206
- muestras de entrenamiento
  - equilibrado, 81
  - partición de datos, 205–206
- muestras de validación
  - partición de datos, 205–206
- muestras estratificadas, 72–73, 77–78
- muestras no aleatorias, 72–73
- muestras ponderadas, 77
- muestras sistemáticas, 72–73
- muestreo 1-de cada-*n*, 73
- muestreo de datos contiguos, 73
  
- nivel de medición
  - cambio de visualizaciones, 252
  - definido, 137
  - en visualizaciones, 254
- nivel de medida, 65, 135
- niveles, soporte de base de datos, 15
- Nodo Adición de RFM
  - conceptos básicos, 85
  - intervalos anidados, 85, 201
  - intervalos independientes, 85, 201
  - opciones de configuración, 86
- nodo Agregar
  - conceptos básicos, 82
  - opciones de configuración, 82
  - procesamiento paralelo, 85
  - rendimiento, 85
- nodo Añadir
  - campos de etiquetado, 97
  - coincidencia de campos, 101
  - conceptos básicos, 100
  - opciones de configuración, 101
- Nodo Análisis, 407
  - pestaña análisis, 408
  - pestaña Resultado, 399
- Nodo Análisis de RFM
  - conceptos básicos, 201
  - configuración, 202
  - intervalos anidados, 85, 201
  - intervalos independientes, 85, 201
  - valores de intervalos, 204
- nodo Anonimizar
  - conceptos básicos, 180
  - creación de valores anonimizados, 183
  - opciones de configuración, 181
- nodo Archivo fijo
  - conceptos básicos, 29
  - opciones de configuración, 29
  - reconocimiento automático de fecha, 31
- Nodo Archivo Statistics, 487
- nodo Archivo var., 25
- Nodo Archivo variable
  - opciones de configuración, 27
  - reconocimiento automático de fecha, 29
- Nodo Auditar datos, 412
  - pestaña Configuración, 414
  - pestaña Resultado, 399
- nodo Colección , 317
  - pestaña Aspecto, 320
  - pestaña Opciones, 318–319
  - utilización del gráfico, 321
- Nodo Conjunto
  - campos de salida, 161
  - combinación de puntuaciones, 161
- Nodo Datos Usuario
  - conceptos básicos, 58
  - opciones de configuración, 60
- nodo de archivo caché, 487
- nodo de exportación Archivo plano, 472
  - pestaña Exportar, 472
- nodo de exportación de ODBC. *Consulte* el nodo Exportar base de datos, 451
- Nodo de exportación Excel, 480
- nodo de exportación IBM SPSS Data Collection, 473
- nodo de exportación XML, 481
- Nodo de fuente Base de datos, 15
  - Consultas SQL, 16
  - editor de consultas, 23–24
  - selección de tablas y vistas, 22
- nodo de fuente Data Collection, 35–36
  - archivos de metadatos, 37
  - archivos de registro, 37
- Nodo de fuente de Excel, 53
- Nodo de fuente de IBM Cognos BI, 44, 49–50
  - iconos, 45
  - importación de datos, 45
  - importación de informes, 47
- Nodo de fuente de Microsoft Excel, 53
- nodo de fuente IBM SPSS Data Collection, 43
  - configuración de conexión de base de datos, 41–42
  - conjunto de respuestas múltiples;respuestas múltiples, 42

- lengua, 40
- tipos de etiquetas, 40
- Nodo de fuente SAS
  - archivos (SAS) *.sd2*, 51
  - archivos (SAS) *.ssd*, 51
  - archivos (SAS) *.ipt*, 51
  - archivos de transporte, 51
- Nodo de fuente XML, 54
- Nodo de preparación de datos automática, 108
- nodo de transformación, 427
- nodo Derivar
  - conceptos básicos, 165
  - condicional, 176
  - conversión de almacenamiento de campos, 177
  - derivación múltiple, 167
  - estado, 174
  - formula, 169
  - generación a partir de enlaces de gráficos Malla, 334
  - generación a partir de gráficos, 364
  - generación a partir de preparación automática de datos, 133
  - generación desde intervalos, 189
  - generación desde un nodo Intervalos, 199
  - marca, 170
  - opciones de configuración, 166
  - recodificación de los valores, 177
  - recuento, 175
  - set, 173
- nodo Distinguir
  - conceptos básicos, 102
  - configuración de optimización, 104
  - ordenación de registros, 103
- nodo Distribución , 308
  - pestaña Aspecto, 310
  - pestaña Gráfico, 309
  - utilización de la tabla, 311
  - utilización del gráfico, 311
- nodo Enterprise View, 9
- nodo Equilibrar
  - generación a partir de gráficos, 364
- Nodo Equilibrar
  - conceptos básicos, 80
  - opciones de configuración, 81
- nodo Estadísticos, 432
  - correlaciones, 433
  - estadísticos, 433
  - etiquetas de correlación, 433
  - pestaña Configuración, 433
  - pestaña Resultado, 399
- nodo Evaluación , 341
  - condición de acierto, 348
  - expresión de puntuación, 348
  - pestaña Aspecto, 350
  - pestaña Gráfico, 347
  - pestaña Opciones, 348
  - regla de negocios, 348
  - resultados de la lectura, 352
  - utilización del gráfico, 353
- Nodo Exportación de Statistics, 501
  - pestaña Exportar, 502
- nodo Exportar base de datos, 451
  - asignación de campos de datos de fuente a columnas de base de datos, 453
  - esquema, 455
  - fuentes de datos, 451
  - indización de tablas, 459
  - nombre de tabla, 451
  - opciones de fusión, 453
  - pestaña Exportar, 451
- nodo Exportar de IBM Cognos BI, 49, 475, 477
- Nodo Exportar SAS, 479
- nodo Filtro
  - conceptos básicos, 154
  - conjunto de respuestas múltiples;respuestas múltiples, 158
  - opciones de configuración, 155
- nodo Fundir, 90
  - campos de etiquetado, 97
  - conceptos básicos, 89
  - configuración de optimización, 99
  - filtrado de campos, 95
  - opciones de configuración, 92, 94
- nodo G. múltiple , 322
  - pestaña Aspecto, 325
  - pestaña Gráfico, 323
  - utilización del gráfico, 326
- nodo Gráfico, 299
  - pestaña Aspecto, 306
  - pestaña Gráfico, 302
  - pestaña Opciones, 305
  - utilización del gráfico, 307
- nodo Gráfico de tiempo, 338
  - pestaña Aspecto, 340
  - pestaña Gráfico, 339
  - utilización del gráfico, 341
- nodo Histograma , 313
  - pestaña Aspecto, 316
  - pestaña Gráfico, 314–315
  - utilización del gráfico, 316
- nodo Histórico, 238
  - conceptos básicos, 237
- nodo Importar a Excel
  - generación desde resultados, 480
- nodo Informe, 442
  - pestaña Plantilla, 443
  - pestaña Resultado, 399
- nodo Intervalos
  - conceptos básicos, 189
  - igual frecuencia, 192
  - igual suma, 192
  - intervalos de ancho fijo, 192
  - intervalos de desviación típica/media, 196
  - opciones de configuración, 190
  - óptimos, 197

- rangos, 195
  - vista preliminar de intervalos, 199
- nodo Intervalos de tiempo, 217, 219, 221
  - conceptos básicos, 216
- nodo Malla , 326
  - ajuste de puntos, 334
  - ajuste de umbrales, 335
  - cambiar diseño, 334
  - control deslizante, 334
  - definición de vínculos, 329
  - pestaña Aspecto, 332
  - pestaña Gráfico, 328
  - pestaña Opciones, 329
  - resumen malla, 337
  - utilización del gráfico, 333
- Nodo Marcas, 207–208
- nodo Matriz, 403
  - explorador de resultados, 406
  - ordenación de filas y columnas, 404
  - pestaña Aspecto, 404
  - pestaña Configuración, 403
  - pestaña Resultado, 399
  - porcentajes de fila, 404
  - porcentajes de la columna, 404
  - resaltado, 404
  - tabulación cruzada, 404
- nodo Medias, 437
  - campos relacionados, 438
  - explorador de resultados, 440–441
  - grupos independientes, 437
  - importancia, 439
  - pestaña Resultado, 399
- nodo Muestrear
  - marco de muestreo, 72
  - muestras aleatorias, 72–73
  - muestras conglomeradas, 72–73, 77
  - muestras estratificadas, 72–73, 77–78
  - muestras no aleatorias, 72–73
  - muestras ponderadas, 77
  - muestras sistemáticas, 72–73
  - Tamaños de muestra por estratos, 78
- nodo Ordenar
  - conceptos básicos, 87
  - configuración de optimización, 88
- nodo Partición, 205–206
- nodo Reclasificar, 185, 188
  - conceptos básicos, 185, 189
  - generación desde una distribución, 311
- nodo Reestructurar, 209–210
  - con nodo Agregar, 210
- nodo Rellenar
  - conceptos básicos, 177
- nodo Reorg. campos, 239
  - opciones de configuración, 239
  - orden personalizado, 239
  - ordenación automática, 241
- Nodo Resultados de IBM SPSS Statistics
  - Pestaña Resultados, 500
- Nodo Resultados de Statistics, 497
  - pestaña Sintaxis, 498
- nodo Seleccionar
  - conceptos básicos, 70
  - generación a partir de enlaces de gráficos Malla, 334
  - generación a partir de gráficos, 364
- nodo Tabla, 398
  - ancho de columna, 151
  - configuración de resultados, 398
  - justificación de columna, 151
  - pestaña Configuración, 398
  - pestaña formato, 151
  - pestaña Resultado, 399
- Nodo Tablero , 251
  - pestaña Aspecto, 285
- nodo Tipo
  - ancho de columna, 151
  - borrado de valores, 65
  - conceptos básicos, 135
  - configuración del papel de modelado, 149
  - copia de tipos, 150
  - datos continuos, 145
  - datos nominales, 146
  - datos ordinales, 146
  - justificación de columna, 151
  - opciones de configuración, 137, 140
  - pestaña formato, 151
  - tipo de campo de marcas, 147
  - tratamiento de los elementos vacíos, 143
- Nodo Transformación de Statistics, 489
  - opciones de configuración, 489
  - pestaña Sintaxis, 489
  - sintaxis permitida, 491
- nodo Transponer, 211
  - campos de cadena, 212
  - campos numéricos, 212
  - nombres de campos, 212
- nodo Val. globales, 446
  - pestaña Configuración, 447
- nodos de exportación, 450
- nodos de fuente
  - Nodo de fuente de Excel, 53
  - Nodo de fuente de IBM Cognos BI, 44, 49–50
  - Nodo de fuente XML, 54
- nodos de gráficos, 243
  - animación, 244, 246
  - Colección, 317
  - Distribución, 308
  - Evaluación, 341
  - G. múltiple, 322
  - Gráfico, 299
  - Gráfico de tiempo, 338
  - Histograma, 313
  - paneles, 244, 246
  - superposiciones, 244

- Tablero, 251
  - Web, 326
- Nodos de IBM SPSS Statistics, 486
- nodos de operaciones con campos, 106
  - generación desde una auditoría de datos, 426
- nodos de operaciones con registros, 69
  - nodo Intervalos de tiempo, 216
- nodos de origen
  - conceptos básicos, 8
  - nodo Archivo fijo, 29
  - Nodo Archivo Statistics, 487
  - nodo Archivo var., 25
  - nodo de datos de usuario, 58, 60
  - Nodo de fuente Base de datos, 15
  - Nodo de fuente SAS, 51
  - nodo Enterprise View, 9
  - tipos de creación de una instancia, 67
- nodos de resultados, 391, 398, 403, 407, 412, 432, 442, 446, 497
  - pestaña Resultado, 399
  - publicar en web, 393
- nombres de campos, 157
  - anonimización, 158
  - exportación de datos, 451, 472, 479, 502
- nombres de variable
  - exportación de datos, 451, 472, 479, 502
- normalizar destino continuo, 117, 133
- normalizar valores
  - nodos de gráficos, 323, 340
- nulos, 143, 422
  - en tablas Matriz, 404
- ocultación datos para utilizarlos en un modelo, 180
- ODBC
  - carga masiva a través de, 461, 464
  - conexión para el nodo Exportar de IBM Cognos BI, 477
  - Nodo de fuente Base de datos, 15
- opciones
  - IBM SPSS Statistics, 448
- opciones de fusión, exportación de base de datos, 453
- opciones de modelo
  - nodo Modelo de Statistics, 494
- Oracle, 15
- orden ascendente, 87
- orden de columna
  - explorador de tablas, 397, 401
- orden de datos de entrada, 97
- orden de ejecución
  - especificación, 523
- orden descendente, 87
- orden natural
  - alteración, 239
- ordenación
  - campos, 239
  - campos ordenados previamente, 88, 104
  - nodo Distinguir, 103
  - registros, 87
  - ordenación de datos, 87, 239
- Palabra clave FILLFACTOR
  - indización de tablas de bases de datos, 460
- Palabra clave UNIQUE
  - indización de tablas de bases de datos, 460
- paletas
  - desplazamiento, 369
  - ocultación, 369
  - presentación, 369
- paneles
  - en visualizaciones, 246
- papeles
  - especificación de campos, 65, 135, 149
- papeles de modelado
  - especificación de campos, 65, 135, 149
- parámetros
  - configuración de los supernodos, 518
  - en IBM Cognos BI, 50
  - propiedades de los nodos, 521
  - Supernodos, 518, 520
- parámetros de ruta, 23–24
- parámetros de Supernodos, 518, 520–521
- partición de datos, 205–206
  - diagramas de evaluación, 348
  - Nodo Análisis, 408
- periodicidad
  - datos de serie temporal, 216
- período de estimación, 221
- períodos
  - nodo Intervalos de tiempo, 225
- períodos cíclicos
  - nodo Intervalos de tiempo, 226
- perturbación, 305
- pestaña Sintaxis
  - Nodo Resultados de Statistics, 498
- pila, 383
- plantillas
  - cambio del nombre, 289
  - eliminación, 289
  - exportación, 289
  - importación, 289
  - nodo Informe, 443
- plantillas de visualización
  - cambio del nombre, 289
  - eliminación, 289
  - exportación, 289
  - importación, 289
  - ubicación, 287
- ponderaciones
  - diagramas de evaluación, 348
- preparación automática de datos
  - análisis de campos, 123
  - configuración de campos, 112
  - creación, 117
  - detalles de acción, 130
  - detalles de campo, 128

- enlaces entre vistas, 121
- excluir campos, 114
- exclusión de campos no utilizados, 112
- fields, 111
- generación del nodo derivar, 133
- nombrar campos, 119
- normalizar destino continuo, 117, 133
- objetivos, 108
- poder predictivo, 126
- preparación de destinos, 115
- preparación de entradas, 115
- preparar fechas y horas, 113
- restablecer vistas, 121
- resumen de acciones, 125
- resumen de procesamiento de campos, 122
- selección de características, 117
- tabla de campos, 127
- vista de modelo, 120
- procesamiento paralelo
  - fusión, 99
  - nodo Agregar, 85
  - ordenación, 89
- procesos
  - Supernodos, 523
- propiedades
  - nodo, 521
  - para campos, 151
- propiedades de los nodos, 521
- prueba *t*
  - muestras independientes, 437
  - muestras relacionadas, 438
  - nodo Medias, 437–438, 442
- publicar en web, 393
- punto, 151
- puntos de corte
  - nodo Intervalos, 189
- puntuación
  - opciones del diagrama de evaluación, 348
- puntuaciones ajustadas de propensión
  - equilibrado de datos, 81
- puntuaciones de propensión
  - equilibrado de datos, 81
- Python
  - procesos de carga masiva, 461, 464
- rango
  - resultado de estadísticos, 434
- rangos, 137
  - valores perdidos, 143
- rangos de casillas
  - archivos de Excel, 53
- rangos enteros, 145
- rangos fraccionales, 195
- rangos reales, 145
- recodificación, 185, 189
- recodificación automática, 185
- reconocimiento automático de fecha, 29, 31
- reconocimiento de fecha, 29, 31
- recuentos
  - nodo Intervalos, 192
  - resultado de estadísticos, 434
- reestructuración de los datos, 209
- regiones en gráficos, 360
- registro
  - etiquetas, 149
  - longitud, 29
  - recuentos, 82
- registros
  - fusión, 89
  - transposición, 211–212
- registros concatenados, 100
- registros incompletos, 92
- registros únicos, 102
- regla de negocios
  - opciones del diagrama de evaluación, 348
- regresión de mínimos cuadrados ponderados localmente
  - nodo Gráfico, 303
- relleno de datos de series temporales, 219
- rendimiento
  - datos de muestreo, 72
  - fusión, 99
  - nodo Agregar, 85
  - nodos Derivar, 199
  - nodos Intervalos, 199
  - ordenación, 89
- residuos
  - nodo Matriz, 404
- resultado de la matriz
  - almacenamiento como texto, 400
- resultado en formato HTML
  - nodo Informe, 444
  - ver en explorador, 396
- resultado en formato XML
  - nodo Informe, 444
- resultado tabular
  - reordenación de columnas, 397
  - selección de casillas, 397
- resultados
  - almacenamiento, 393
  - exportación, 396
  - generación de nodos nuevos desde, 393
  - HTML, 396
  - impresión, 393
- ROI
  - gráficos, 341, 352
- rotación de gráficos 3D, 250
- SAS
  - configuración de opciones de importación, 52
- selección de filas (casos), 70
- significación
  - fuerza de correlación, 433
- símbolo de agrupación
  - formatos de presentación de los números, 153

- símbolo decimal, 27–28, 151
  - formatos de presentación de los números, 153
  - nodo de exportación Archivo plano, 472
- Sintaxis de XPath, 54
- sistemas de coordenadas
  - transformación, 379
- sobrescritura de tablas de base de datos, 451
- SPLOM, 262
  - ejemplo, 279, 283
- SPSS Modeler Server, 2
- suavizamiento
  - nodo Gráfico, 303
- suavizamiento LOESS
  - nodo Gráfico, 303
- suavizamiento LOESS *Consulte suavizamiento LOESS*
  - nodo Gráfico, 303
- suma
  - nodo Val. globales, 447
  - resultado de estadísticos, 434
- Supernodos, 505
  - acercamiento, 516
  - almacenamiento, 524
  - anidación, 510
  - bloqueo, 513–514
  - carga, 524
  - configuración de los parámetros, 518
  - creación, 508
  - creación de cachés para, 522
  - desbloqueo, 514
  - edición, 516
  - procesos, 523
  - protección por contraseña, 513–515
  - supernodos de origen, 506
  - supernodos de proceso, 506
  - supernodos terminales, 507
  - tipos de, 505
  - uso de comentarios con, 517
- superposición de gráfico de colores, 244
- superposición de gráfico de formas, 244
- superposición de gráfico de paneles, 244, 246
- superposición de gráfico de tamaño, 244
- superposición de gráficos, 244
- sustitución de valores de campos, 177
  
- tablas
  - almacenamiento como texto, 400
  - almacenamiento de resultados, 399
  - uniones, 90
- tabulación cruzada
  - nodo Matriz, 403–404
- tamaño
  - en visualizaciones, 245
- tamaño de confirmación, 461
- texto
  - codificación, 28, 31, 473
  - datos, 25, 29
  - delimitado, 25
  - tipo, 32
  - tipo de conjunto, 137
  - tipo de marca, 137, 147
  - tipo de uso, 32, 137
  - tipos de campos, 65, 135
    - en visualizaciones, 254
  - tipos de datos, 29, 65, 106, 135, 137
    - creación de una instancia, 141
  - tipos de etiquetas
    - nodo de fuente IBM SPSS Data Collection, 40
  - tipos de gráficos
    - graphboard, 259
  - tipos de variables
    - en visualizaciones, 254
  - tramos en gráficos, 356
  - transformaciones
    - reclasificar, 185, 189
    - recodificación, 185, 189
  - transparencia
    - en visualizaciones, 245
  - transparencia en gráficos, 244
  - transposición de datos, 211–212
  - tratamiento de los elementos vacíos, 65, 135, 143
    - nodo Intervalos, 190
    - valores de relleno, 177
  - truncado de nombres de campos, 155, 157
  
  - umbrales
    - visualización de umbrales de intervalos, 199
  - unión de conjuntos de datos, 100
  - unión exterior, 90
  - unión interior, 90
  - uniones, 89–90, 92
    - exteriores parciales, 94
  - uniones parciales, 90, 94
  - Utilidad de conversión de mapas, 290, 292
  
  - vacíos, 422
    - en tablas Matriz, 404
  - valor clave para la agregación, 82
  - valor de cuartil para la agregación, 82
  - valor de la media de los registros, 82
  - valor de la media para la agregación, 82
  - valor de la mediana para la agregación, 82
  - valor de semilla
    - muestreo y registros, 76, 207
  - valor de semilla aleatoria
    - registros de muestreo, 76, 207
  - valor de varianza para la agregación, 82
  - valor máximo para la agregación, 82
  - valor mínimo para la agregación, 82
  - Valor *p*
    - importancia, 439
  - valores
    - especificación, 143
    - etiquetas de valor y de campo, 65, 135, 143

- lectura, 142
- valores de agrupación, 311
- valores de selección, 356, 360, 363
- valores definidos por el usuario como perdidos
  - en tablas Matriz, 404
- valores esperados
  - nodo Matriz, 404
- valores falsos, 147
- valores globales, 446
- valores no definidos, 92
- valores nulos
  - datos mezclados, 34, 63
  - en tablas Matriz, 404
- valores perdidos, 106, 143, 147
  - en nodos Agregar, 82
  - en tablas Matriz, 404
  - gestión, 422
  - relleno, 422
- valores perdidos del sistema, 422
  - en tablas Matriz, 404
- valores perdidos por el usuario, 422
- valores predefinidos, conexión de base de datos, 19
- valores sumados, 82
- valores vacíos
  - en tablas Matriz, 404
- valores verdaderos, 147
- Variables de códigos
  - nodo de fuente IBM SPSS Data Collection, 40
- variables de series temporales, 237
- Variables del sistema
  - nodo de fuente IBM SPSS Data Collection, 40
- Variables SourceFile
  - nodo de fuente IBM SPSS Data Collection, 40
- varianza
  - resultado de estadísticos, 434
- varios campos
  - selección, 169
- varita mágica en gráficos, 363
- ver
  - resultado en formato HTML en un explorador, 396
- vista de modelo
  - en preparación automática de datos, 120
- visualización
  - gráficos y diagramas, 243
- visualización de mapas
  - ejemplo, 281
- visualizaciones
  - categorías, 376
  - colores y tramas, 370
  - copia, 383
  - edición, 367
  - ejes, 374
  - escalas, 374
  - estilos de guión, 370
  - forma de puntos, 371
  - formatos de numeración, 373
  - márgenes, 373
  - modo de edición, 367
  - paneles, 376, 378
  - posición de la leyenda, 383
  - relación de aspecto de puntos, 371
  - relleno, 373
  - rotación de puntos, 371
  - texto, 369
  - transformación de los sistemas de coordenadas, 379
  - transparencia, 370
  - transponer, 376, 378–379
- visualizaciones de mapas
  - creación, 266
- zoom, 516