

IBM SPSS Bootstrapping 20



注：この情報とサポートされている製品をご使用になる前に、「注意事項」（p. 43）の一般情報をお読みください。

本版は IBM® SPSS® Statistics 20 , および新版で指示されるまで後続するすべてのリリースおよび変更に対して適用されます。

Adobe 製品のスクリーンショットは Adobe Systems Incorporated の許可を得て転載しています。

Microsoft 製品のスクリーンショットは Microsoft 社の許可を得て転載しています。

Licensed Materials - Property of IBM

© Copyright IBM Corporation 1989, 2011.

U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

はじめに

IBM® SPSS® Statistics は、データ分析の包括的システムです。Bootstrapping は、このマニュアルで説明されている追加の分析手法を提供するオプションのアドオン モジュールです。Bootstrapping アドオン モジュールは SPSS Statistics Core システムと組み合わせて使用し、Core システムに完全に統合されます。

IBM Business Analytics について

IBM Business Analytics ソフトウェアは、意思決定者がビジネス パフォーマンスを向上させるために信頼する完全で、一貫した正確な情報を提供します。ビジネス インテリジェンス、予測分析、財務実績および戦略管理、および 分析アプリケーションの包括的なポートフォリオを利用することによって、現在の実績を明確、迅速に理解し、将来の結果を予測することができます。豊富な業界のソリューション、実績ある実例、専門サービスと組み合わせ、さまざまな規模の組織が、高い生産性を実現、意思決定を自信を持って自動化し、より良い決定をもたらします。

このポートフォリオの一部として、IBM SPSS Predictive Analytics ソフトウェアを使用する組織は、将来のイベントを予測し、その洞察に基づいて積極的に行動し、より優れた業績を実現することができます。全世界の企業、政府、学術分野のお客様が IBM SPSS の技術を活用し、不正行為を減少させ、リスクを軽減させながら、顧客の獲得、保持、成長において、競争優位を高めることができます。IBM SPSS ソフトウェアを日々の業務に取り入れることによって、組織は業務目標を達成し、大きな競争的優位を獲得することができるよう、意思決定を方向付け、自動化することができるようになります。お問い合わせは、<http://www.ibm.com/spss> を参照してください。

テクニカル サポート

テクニカル サポートのサービスをご利用いただけます。IBM Corp. 製品の使用方法や、対応しているハードウェア環境へのインストールに関して問い合わせることもできます。テクニカル サポートの詳細については、IBM Corp. Web サイト (<http://www.ibm.com/support>) を参照してください。連絡の際は、所属団体名、サポート契約などを確認できるよう、あらかじめ手元にご用意ください。

学生向けテクニカル サポート

IBM SPSS ソフトウェア製品の Student 版、アカデミック版、Grad パック版を使用している学生の場合、学生用の特別オンライン ページ、[Solutions for Education \(http://www.ibm.com/spss/rd/students/\)](http://www.ibm.com/spss/rd/students/) ページを参照してください。大学提供の IBM SPSS ソフトウェアのコピーを使用している場合、大学の IBM SPSS 製品コーディネータにお問い合わせください。

カスタマ サービス

配送やアカウントに関するご質問は、お近くの営業所にお問い合わせください。お問い合わせの際には、シリアル番号をご用意ください。

トレーニング セミナー

IBM Corp. では一般公開およびオンサイトで トレーニング セミナーを実施しています。セミナーでは実践的な講習を行います。セミナーは主要都市で定期的開催されます。セミナーに関する詳細については、<http://www.ibm.com/software/analytics/spss/training> を参照してください。

追加の出版物

Marija Noruš による『SPSS Statistics: Guide to Data Analysis』、『SPSS Statistics: Statistical Procedures Companion』、『SPSS Statistics: Advanced Statistical Procedures Companion』が Prentice Hall から出版されました。補助的な資料としてご利用いただけます。これらの出版物には、SPSS Statistics Base モジュール、Advanced Statistics モジュール、Regression モジュールの統計的手続きについて記載されています。初めてデータ分析を行う場合、高度なアプリケーションを使用する場合に応じて、この本は IBM® SPSS® Statistics が提供している機能を効率よく使用するための手助けとなります。出版物の内容、サンプルの図表などの詳細は、作者の Web サイトを参照してください。
<http://www.norusis.com>

内容

パート I: ユーザー ガイド

| | | |
|---|---------------------|---|
| 1 | ブートストラップの概要 | 1 |
| 2 | Bootstrapping | 3 |
| | ブートストラップをサポートする手順 | 5 |
| | BOOTSTRAP コマンドの追加機能 | 9 |

パート II: 例

| | | |
|---|--------------------------|----|
| 3 | ブートストラップ | 11 |
| | ブートストラップを使用した比率の信頼区間の取得 | 11 |
| | データの準備 | 11 |
| | 分析の実行 | 12 |
| | ブートストラップの指定 | 15 |
| | 統計 | 16 |
| | 度数分布表 | 17 |
| | ブートストラップを使用した中央値の信頼区間の取得 | 17 |
| | 分析の実行 | 17 |
| | 記述統計量 | 21 |
| | ブートストラップを使用したより適切な予測値の選択 | 21 |
| | データの準備 | 22 |
| | 分析の実行 | 22 |
| | パラメータ推定値 | 30 |
| | 推奨参考文献 | 31 |

付録

| | | |
|---|-----------|----|
| A | サンプル ファイル | 32 |
| B | 注意事項 | 43 |
| | 参考文献 | 46 |
| | 索引 | 48 |

パート I: ユーザー ガイド

ブートストラップの概要

データを収集する場合、サンプルを収集する母集団のプロパティに関心がある場合があります。サンプルから算出した推定値によってこれらの母集団パラメータについて推定します。たとえば、製品に付属する 従業員 data.sav データセットは、より大きい母集団の従業員からの無作為サンプルで、現在の給与の \$34,419.57 というサンプル平均値は、従業員の母集団における現在の平均給与の推定値です。さらに、推定値には、サイズ 474 のサンプルに対し \$784.311 の標準誤差があるため、現在の平均給与の 95% の信頼区間は \$32,878.40 ~ \$35,960.73 となります。これらの推定値はどれほど信頼できるものなのでしょうか。特定の「分かっている」母集団と正常に動作するパラメータの場合、サンプル推定のプロパティについて十分理解しており、これらの結果に自信があります。ブートストラップは、「不明な」母集団と正常に動作しないパラメータの推定値のパラメータについて、詳細情報を探します。

図 1-1
母平均に関するパラメトリック推定

| | | 統計量 | 標準誤差 |
|-------|---------------|-------------|-----------|
| 現在の給与 | 平均値 | \$34,419.57 | \$784.311 |
| | 平均値の 95% 信頼区間 | | |
| | 下限 | \$32,878.40 | |
| | 上限 | \$35,960.73 | |
| | 中央値 | \$28,875.00 | |

ブートストラップの機能

もっとも簡単に言えば、サンプルサイズが N のデータセットの場合、元のデータセットから置き換えた、サイズが N である B の「ブートストラップ」を取得し、これらの B ブートストラップ サンプルの推定値を計算します。これらの B ブートストラップ推定は、推定値について推定できるサイズ B のサンプルです。たとえば、従業員 data.sav から 1,000 件のブートストラップ サンプルを取得する場合、現在の給与のサンプル平均値に対するブートストラップ推定標準誤差の \$776.91 は、推定値 \$784.311 に代わるものです。

さらに、ブートストラップでは、パラメトリック推定では使用できない中央値の標準誤差や信頼区間を提供します。

図 1-2
サンプル平均に関するパラメトリック推定

| | 統計量 | 標準誤差 | ブートストラップ ^a | | | | |
|-------|---------------|-------------|-----------------------|----------|----------|-------------|-------------|
| | | | バイアス | 標準誤差 | 95% 信頼区間 | | |
| | | | | | 下限 | 上限 | |
| 現在の給与 | 平均値 | \$34,419.57 | \$784.311 | \$14.66 | \$776.91 | \$32,990.38 | \$36,026.06 |
| | 平均値の 95% 信頼区間 | | | | | | |
| | 下限 | \$32,878.40 | | | | | |
| | 上限 | \$35,960.73 | | | | | |
| | 中央値 | \$28,875.00 | | \$-13.22 | \$536.63 | \$27,750.00 | \$29,850.00 |

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

製品のブートストラップのサポート

ブートストラップは、ブートストラップをサポートする手続きのサブダイアログとして組み込まれています。ブートストラップをサポートする手続きの詳細は、「[ブートストラップをサポートする手順](#)」を参照してください。

ブートストラップがダイアログで要求されると、新しい個別の BOOTSTRAP コマンドが、ダイアログで生成された通常のシンタックスの他に貼り付けられます。BOOTSTRAP コマンドは、指定に従って、ブートストラップ サンプルを作成します。内部では、これらのブートストラップ サンプルがデータ エディタに明示的に示されていなくても、分割と同様に処理します。つまり、内部では、 $B \times N$ 件のケースがあり、ブートストラップ時にデータを処理する場合、状態バーのケース数は $1 \sim B \times N$ となります。出力管理システム (OMS) を使用して、「ブートストラップ分割」に分析を実行した結果を収集します。これらの結果はプールされ、プールされたブートストラップの結果は、手続きで生成された通常の実出力とともにビューアに表示されます。特定のケースの場合、「ブートストラップ分割 0」への参照を確認できます。これは元のデータセットです。

Bootstrapping

ブートストラップとは、平均値、中央値、比率、オッズ比、相関係数、または回帰係数など、推定に対する標準語さおよび信頼区間の頑強な推定を取得する方法です。仮説検定の構築にも使用することができます。ブートストラップは、これらの方法の想定が疑わしい場合（不均一残差の回帰モデルが小さいサンプルに適合する場合など）、またはパラメトリック推定が不可能または標準誤差の計算に非常に複雑な数式が必要な場合（中央値、4 分位、パーセンタイルの信頼区間を計算する場合など）、パラメトリック推定の代替として最も役立ちます。

例。ある電話会社では、毎月 27% の顧客が解約しています。解約数を減少させるために、管理者は事前定義された顧客グループで、解約率がどのように異なるかを知りたいと考えています。ブートストラップを使用して、1 つの解約率が、4 つの主な顧客タイプを説明するかどうかを指定できます。詳細は、[3 章 ブートストラップを使用した比率の信頼区間の取得 in IBM SPSS Bootstrapping 20](#) を参照してください。

従業員レコードを確認した場合、管理者は従業員の以前の職歴に関心を持ちます。職歴は偏っていますが、平均値は中央値に比べて、従業員の「一般的な」以前の職歴について適切な推定値とはなりません。ただし、パラメトリック信頼区間は、製品の中央値には使用できません。詳細は、[3 章 ブートストラップを使用した中央値の信頼区間の取得 in IBM SPSS Bootstrapping 20](#) を参照してください。

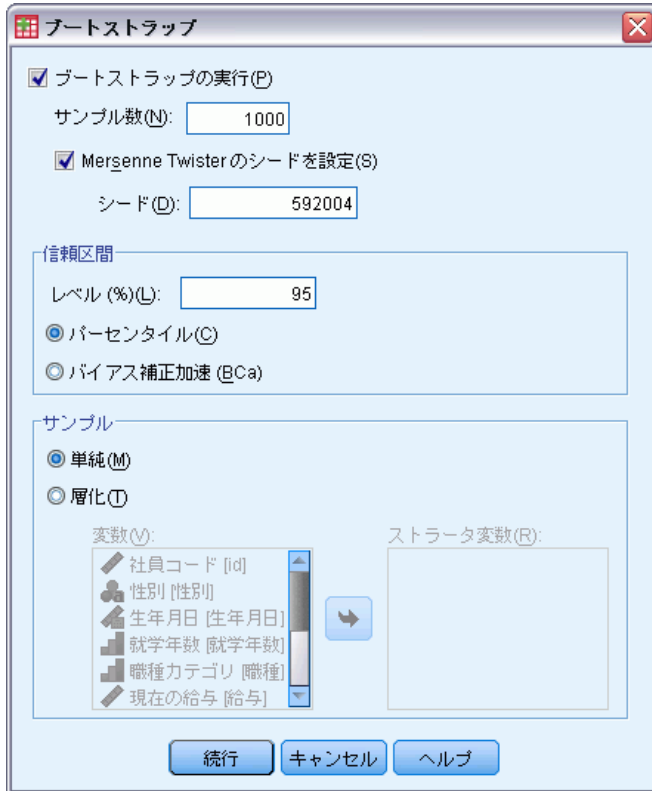
また、管理者は線型モデルを現在の給与と初任給との差分にあてはめて、従業員の給与の増加に関連する要素を決定したいとも考えます。線型モデルをブートストラップする場合、残差やワイルド ブートストラップなど、特別な再抽出方法を使用して、より正確な結果を取得できます。詳細は、[3 章 ブートストラップを使用したより適切な予測値の選択 in IBM SPSS Bootstrapping 20](#) を参照してください。

ブートストラップ サンプルの分析から得られた結果のブートストラップ サンプリングおよびプールは、多くの手順によりサポートされています。ブートストラップ分析を指定するコントロールは、ブートストラップをサポートする手順の共通サブダイアログとして直接統合されます。ブートストラップ ダイアログの設定は、手順を通じて続きます。そのため、ダイアログを介してブートストラップにより度数分析を実行すると、ブートストラップをサポートする手順について、ブートストラップがデフォルトで有効になります。

ブートストラップ分析を行うには

- ▶ メニューから、ブートストラップをサポートする手順を選択して、[ブートストラップ] をクリックします。

図 2-1
[ブートストラップ] ダイアログ ボックス



- ▶ [ブートストラップの実行] を選択します。

次のオプションを制御できます。

サンプル数。パーセンタイルおよび BCa 区間が生成された場合、少なくとも 1000 件のブートストラップ サンプルを使用することをお勧めします。正の整数を指定します。

Mersenne Twister のシードを設定。シードを設定すると、分析を複製することができます。このコントロールを使用すると、アクティブ ジェネレータとして Mersenne Twister を設定し、[乱数ジェネレータ] ダイアログの固定開始ポイントを指定することと同様の設定ができますが、このダイアログでシードを設定すると、乱数ジェネレータの現在の状態を保持し、分析が完了した後、その状態を復元します。

信頼区間。50 より大きく 100 より小さい信頼区間を指定します。パーセンタイル区間は、該当する信頼区間のパーセンタイルに対応するブートストラップ値を使用します。たとえば、95% パーセンタイルの信頼区間は、ブートストラップ値の 2.5 番目と 97.5 番目のパーセンタイルを区間の下限および上限として使用します（必要に応じて、ブートストラップ値を補間します）。バイアス修正と加速 (BCa) 区間は、より正確な区間に調整し、計算するためにはより多くの時間を必要とします。

サンプル。[単純] なサンプリング方法は、元のデータセットからの置換によるケースの再抽出です。[層化] 方法は、ストラータ変数のクロス分類によって定義されたストラータ内で、元のデータセットからの置換によるケースの再抽出です。層化ブートストラップ サンプリングは、ストラータ間の単位が非常に異なるのに対し、ストラータ内の単位が比較的同質である場合に役立ちます。

ブートストラップをサポートする手順

次の手順では、ブートストラップをサポートしています。

注：

- ブートストラップは、多重代入データセットを処理しません。データセットに `Imputation_` 変数がある場合、[ブートストラップ] ダイアログは無効になります。
- ブートストラップは、リストごとの削除を使用してケースの基本を決定します。つまり、分析変数のいずれかについて欠損値を持つケースは分析から削除されるため、ブートストラップが有効な場合、分析の手順で別の形式の欠損値処理を指定している場合であっても、リストごとの削除が有効となります。

Statistics Base のオプション

度数

- [統計量] テーブルでは、平均値、標準偏差、分散、中央値、歪度、尖度、およびパーセンタイルのブートストラップ推定をサポートしています。
- [度数] テーブルでは、パーセントのブートストラップ推定をサポートしています。

記述統計

- [記述統計量] テーブルでは、平均値、標準偏差、分散、歪度、および尖度のブートストラップ推定をサポートしています。

探索的

- [記述統計量] テーブルでは、平均値、5% トリム平均、標準偏差、中央値、歪度、尖度、および4 分位範囲のブートストラップ推定をサポートしています。
- [M 推定量] テーブルは、Huber の M 推定量、Tukey のバイウエイト、Hampel の M 推定量、および Andrew のウェーブのブートストラップ推定をサポートしています。
- [パーセンタイル] テーブルでは、パーセンタイルのブートストラップ推定をサポートしています。

クロス集計表

- [傾向測度] テーブルは、ラムダ、Goodman と Kruskal のタウ、不確定性係数、および Somers の d のブートストラップ推定をサポートしています。
- [対称性による類似度] テーブルは、ファイ、Cramer の V、分割係数、Kendall のタウ b、Kendall のタウ c、ガンマ、Spearman の相関、Pearson の R のブートストラップ推定をサポートしています。
- [リスク推定値] テーブルでは、オッズ比のブートストラップ推定をサポートしています。
- [Mantel-Haenszel 共通オッズ比] テーブルでは、ブートストラップ推定と $\ln(\text{推定})$ の有意度検定をサポートしています。

平均値

- [レポート] テーブルでは、平均値、中央値、グループ中央値、標準偏差、分散、尖度、歪度、調和平均、および幾何平均のブートストラップ推定をサポートしています。

1 サンプルの t 検定

- [統計量] テーブルでは、平均値、および標準偏差のブートストラップ推定をサポートしています。
- [検定] テーブルでは、ブートストラップ推定と平均値の差の有意度検定をサポートしています。

独立したサンプルの t 検定

- [グループ統計量] テーブルでは、平均値、および標準偏差のブートストラップ推定をサポートしています。
- [検定] テーブルでは、ブートストラップ推定と平均値の差の有意度検定をサポートしています。

対応のあるサンプルの t 検定

- [統計量] テーブルでは、平均値、および標準偏差のブートストラップ推定をサポートしています。

- [相関] テーブルでは、相関のブートストラップ推定をサポートしています。
- [検定] テーブルでは、平均値のブートストラップ推定をサポートしています。

一元配置分散分析

- [記述統計量] テーブルでは、平均値、および標準偏差のブートストラップ推定をサポートしています。
- [多重比較] テーブルでは、平均の差のブートストラップ推定をサポートしています。
- [対比検定] テーブルでは、ブートストラップ推定と対比の値の有意度検定をサポートしています。

GLM - 1 変量

- [記述統計量] テーブルでは、平均値、および標準偏差のブートストラップ推定をサポートしています。
- [パラメータ推定] テーブルでは、ブートストラップ推定と係数 B の有意度検定をサポートしています。
- [対比結果] テーブルでは、ブートストラップ推定と差分の有意度検定をサポートしています。
- [推定周辺平均:推定] テーブルでは、平均値のブートストラップ推定をサポートしています。
- [推定周辺平均:ペアごとの比較] テーブルでは、平均値のブートストラップ推定をサポートしています。
- [ポスト ホック テスト: 多重比較] テーブルでは、平均の差のブートストラップ推定をサポートしています。

2 変量の相関分析

- [記述統計量] テーブルでは、平均値、および標準偏差のブートストラップ推定をサポートしています。
- [相関] テーブルでは、相関のブートストラップ推定および有意検定をサポートしています。

注:

Pearson 相関のほか、ノンパラメトリック相関 (Kendall のタウ b または Spearman) が要求されて場合、ダイアログは、CORRELATIONS コマンド およびNONPAR CORR コマンドを、それぞれ個別の BOOTSTRAP コマンドとともに貼り付けます。同じブートストラップ サンプルを使用して、すべての相関を計算します。

プーリングの前、Fisher Z の変換が相関に適用されます。プーリングの前、逆の Z の変換が相関に適用されます。

偏相関分析

- [記述統計量] テーブルでは、平均値、および標準偏差のブートストラップ推定をサポートしています。
- [相関] テーブルでは、相関のブートストラップ推定をサポートしています。

線型回帰

- [記述統計量] テーブルでは、平均値、および標準偏差のブートストラップ推定をサポートしています。
- [相関] テーブルでは、相関のブートストラップ推定をサポートしています。
- [モデル要約] テーブルでは、Durbin-Watson のブートストラップ推定をサポートしています。
- [係数] テーブルでは、ブートストラップ推定と係数 B の有意度検定をサポートしています。
- [相関係数] テーブルでは、相関のブートストラップ推定をサポートしています。
- [残差統計量] テーブルでは、平均値、および標準偏差のブートストラップ推定をサポートしています。

順序回帰

- [パラメータ推定] テーブルでは、ブートストラップ推定と係数 B の有意度検定をサポートしています。

判別分析

- [標準化正準判別関数係数] テーブルでは、標準化係数のブートストラップ推定がサポートされています。
- [正準判別関数係数] テーブルでは、標準化されていない係数のブートストラップ推定がサポートされています。
- [分類関数係数] テーブルでは、係数のブートストラップ推定がサポートされています。

Advanced Statistics のオプション

GLM 多変量

- [パラメータ推定] テーブルでは、ブートストラップ推定と係数 B の有意度検定をサポートしています。

線型混合モデル

- [固定効果の推定値] テーブルでは、ブートストラップ推定と推定の有意度検定をサポートしています。
- [共分散パラメータの推定値] テーブルでは、ブートストラップ推定と推定の有意度検定をサポートしています。

一般化線型モデル

- [パラメータ推定] テーブルでは、ブートストラップ推定と係数 B の有意度検定をサポートしています。

Cox 回帰分析

- [式に含まれる変数] テーブルでは、ブートストラップ推定と係数 B の有意度検定をサポートしています。

Regression ,7fsIfsvfsVfs?fs“}

2 項ロジスティック回帰

- [式に含まれる変数] テーブルでは、ブートストラップ推定と係数 B の有意度検定をサポートしています。

多項ロジスティック回帰

- [パラメータ推定] テーブルでは、ブートストラップ推定と係数 B の有意度検定をサポートしています。

BOOTSTRAP コマンドの追加機能

コマンド シンタックス言語を使用して、次のことも実行できます。

- 残差およびワイルド ブートストラップ サンプリング (SAMPLING サブコマンド) の実行

複雑なシンタックス情報については、「コマンド シンタックス リファレンス」を参照してください。

パート II: 例

ブートストラップ

ブートストラップとは、平均値、中央値、比率、オッズ比、相関係数、または回帰係数など、推定に対する標準語さおよび信頼区間の頑強な推定を取得する方法です。仮説検定の構築にも使用することができます。ブートストラップは、これらの方法の想定が疑わしい場合（不均一残差の回帰モデルが小さいサンプルに適合する場合など）、またはパラメトリック推定が不可能または標準誤差の計算に非常に複雑な数式が必要な場合（中央値、4 分位、パーセンタイルの信頼区間を計算する場合など）、パラメトリック推定の代替として最も役立ちます。

ブートストラップを使用した比率の信頼区間の取得

ある電話会社では、毎月 27% の顧客が解約しています。解約数を減少させるために、管理者は事前定義された顧客グループで、解約率がどのように異なるかを知りたいと考えています。

この情報は、telco.sav に収集されています。詳細は、[A 付録 p.32 サンプル ファイル](#) を参照してください。ブートストラップを使用して、1 つの解約率が、4 つの主な顧客タイプを説明するかどうかを指定します。

注：この例では度数分布表手続きを使用し、Statistics Base オプションが必要です。

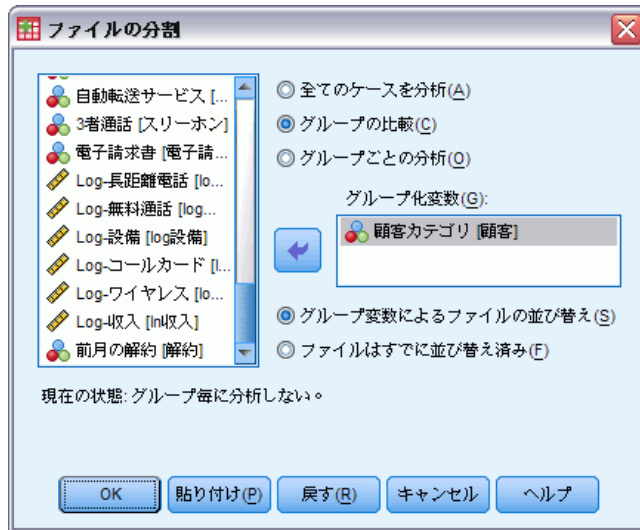
データの準備

まず、顧客カテゴリでファイルを分割する必要があります。

- ▶ ファイルを分割するには、データ エディタのメニューから次の項目を選択します。

データ > ファイルの分割(F)...

図 3-1
[ファイルの分割] ダイアログ ボックス



- ▶ [グループの比較] を選択します。
- ▶ グループの基準とする変数として [顧客カテゴリ] を選択します。
- ▶ [OK] をクリックします。

分析の実行

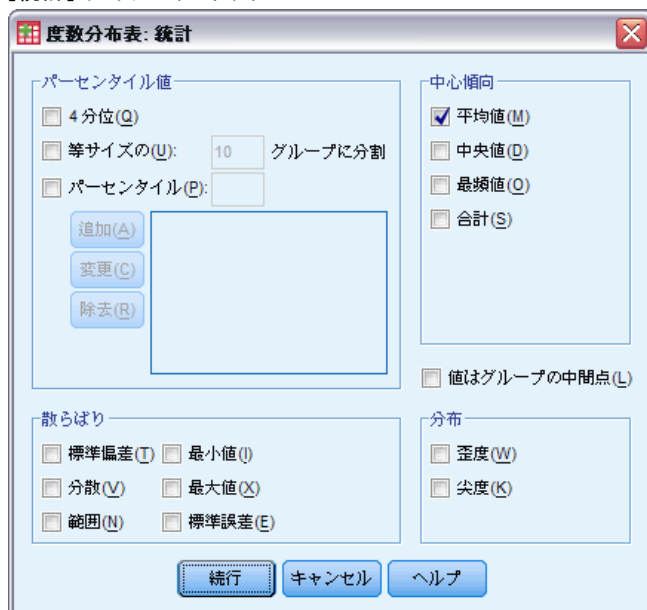
- ▶ 比率のブートストラップ信頼区間を取得するには、メニューから次の項目を選択します。
分析(A) > 記述統計 > 度数分布表...

図 3-2
[度数分布表]メイン ダイアログ



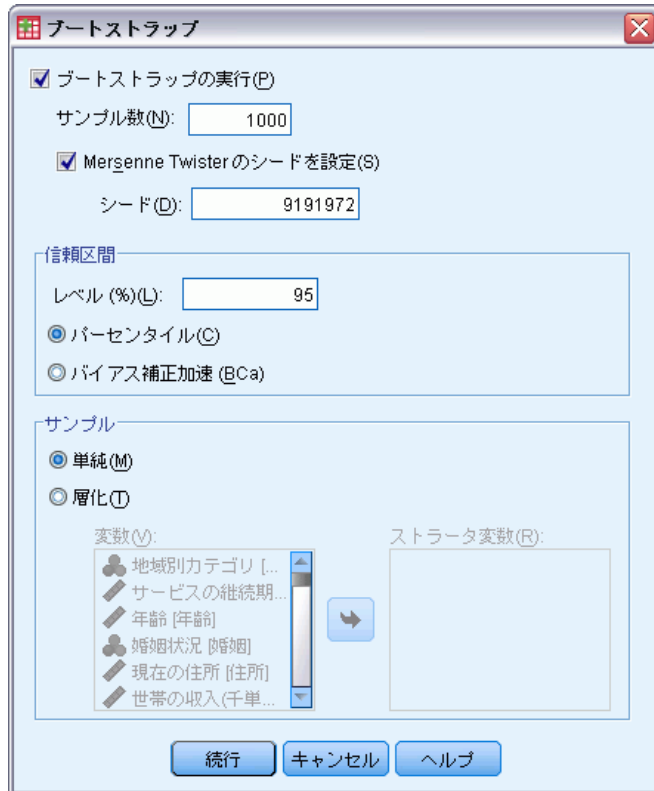
- ▶ 分析の変数として、前月の解約数 [解約] を選択します。
- ▶ [統計量] をクリックします。

図 3-3
[統計] ダイアログ ボックス



- ▶ [中心傾向] グループの [平均] を選択します。
- ▶ [続行] をクリックします。
- ▶ [度数分布表] ダイアログ ボックスで [ブートストラップ] をクリックします。

図 3-4
[ブートストラップ] ダイアログ ボックス



- ▶ [ブートストラップの実行] を選択します。
- ▶ この例の結果を正確に複製するには、[Mersenne Twister のシードを設定] を選択し、シードとして9191972 と入力します。
- ▶ [続行] をクリックします。
- ▶ [度数分布表] ダイアログ ボックスで [OK] をクリックします。

以上の選択により、次のコマンド シンタックスが生成されます。

```
SORT CASES BY custcat.
SPLIT FILE LAYERED BY custcat.
PRESERVE.
SET RNG=MT MTINDEX=9191972.
SHOW RNG.
BOOTSTRAP
  /SAMPLING METHOD=SIMPLE
  /VARIABLES INPUT=churn
  /CRITERIA CILEVEL=95 CITYPE=PERCENTILE NSAMPLES=1000
  /MISSING USERMISSING=EXCLUDE.
FREQUENCIES VARIABLES=churn
  /STATISTICS=MEAN
  /ORDER=ANALYSIS.
```

RESTORE.

- SORT CASES コマンドおよび SPLIT FILE コマンドが変数 `custcat` でファイルを分割します。
- PRESERVE コマンドおよび RESTORE コマンドは、乱数ジェネレータの現在の状態を「記憶」し、ブートストラップが終了するとシステムをその状態に復元します。
- SET コマンドは、乱数ジェネレータを Mersenne Twister に、インデックスを 9191972 に設定し、ブートストラップの結果が正確に複製できるようにします。SHOW コマンドは、参照用に出力のインデックスを表示します。
- BOOTSTRAP コマンドは、単純な再抽出を使用して、1,000 件のブートストラップ サンプルを要求します。
- 変数 `churn` を使用して、再抽出のケース基準を決定します。この変数の欠損値を持つケースは分析から除外されます。
- BOOTSTRAP の後の FREQUENCIES の手続きは、各ブートストラップのサンプルで実行されます。
- STATISTICS サブコマンドは、元のデータの変数 `churn` の平均を生成します。また、度数分布表の平均値および割合にプールされた統計量を作成します。

ブートストラップの指定

図 3-5
ブートストラップの指定

| | | |
|----------|---------|-------|
| サンプリング方法 | 単純 | |
| サンプル数 | | 1000 |
| 信頼区間レベル | | 95.0% |
| 信頼区間型 | パーセンタイル | |

ブートストラップの指定の表には、再抽出時に使用された設定が含まれ、また目的の分析が実行されたかどうかを確認するために参照できます。

統計

図 3-6
比率のブートストラップ信頼区間を表示した統計量

前月の解約

| 顧客カテゴリ | | | | Statistic | ブートストラップ ^a | | | |
|------------|----|-----|-----|-----------|-----------------------|------|----------|----|
| | | | | | バイアス | 標準誤差 | 95% 信頼区間 | |
| | | | | | | | 下限 | 上限 |
| ベーシック サービス | 度数 | 有効 | 266 | 0 | 0 | 266 | 266 | |
| | | 欠損値 | 0 | 0 | 0 | 0 | 0 | |
| | | 平均値 | .31 | .00 | .03 | .26 | .37 | |
| E-サービス | 度数 | 有効 | 217 | 0 | 0 | 217 | 217 | |
| | | 欠損値 | 0 | 0 | 0 | 0 | 0 | |
| | | 平均値 | .27 | .00 | .03 | .21 | .34 | |
| プラス サービス | 度数 | 有効 | 281 | 0 | 0 | 281 | 281 | |
| | | 欠損値 | 0 | 0 | 0 | 0 | 0 | |
| | | 平均値 | .16 | .00 | .02 | .12 | .20 | |
| トータル サービス | 度数 | 有効 | 236 | 0 | 0 | 236 | 236 | |
| | | 欠損値 | 0 | 0 | 0 | 0 | 0 | |
| | | 平均値 | .37 | .00 | .03 | .31 | .44 | |

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

統計表には、顧客カテゴリのレベルごとに、前月の解約数の平均値が表示されています。前月の解約数の値は 0 および 1 のみであるため（1 は解約した顧客を示す）、平均値は解約者数の比率と等しくなります。[統計] 列には、元のデータセットを使用して、度数分布表が通常生成する値が示されます。[ブートストラップ] 列は、ブートストラップ アルゴリズムで作成されます。

- [バイアス] は、ブートストラップ サンプルのこの統計量の平均値と [統計] 列の値との差です。この場合、[前月の解約数] の平均値は、1000 件のブートストラップにすべてについて計算され、またこれらの平均値の平均が計算されます。
- Std. [標準誤差] は、1000 件のブートストラップ サンプル全体の [前月の解約数] の平均値の標準誤差です。
- 95% のブートストラップ信頼区間の下限は、1000 件のブートストラップ サンプルが昇順に並んでいる場合、前月の解約数の 25 番目および 26 番目の平均値の補間となります。上限は、975 番目と 976 番目の平均値の補間となります。

表の結果は、解約率は顧客の種類によって異なることを示します。具体的には、プラス サービスの顧客の信頼区間は、他の信頼区間と重ならず、これらの顧客が平均的に解約の可能性が引くことを示します。

値が 2 つしかないカテゴリ変数进行处理する場合、これらの信頼区間は、1 サンプルのノンパラメトリック検定の手続きまたは 1 サンプルの t 検定の手続きで生成された変数の代わりとなります。

度数分布表

図 3-7
比率のブートストラップ信頼区間を表示した度数分布表

| 顧客カテゴリ | 有効 | いいえ | 度数 | パーセント | 有効パーセン ト | 累積パーセン ト | パーセントのブートストラップ ^a | | | |
|------------|----|-----|-----|-------|-------------|-------------|-----------------------------|------|----------|-------|
| | | | | | | | バイアス | 標準誤差 | 95% 信頼区間 | |
| | | | | | | | | | 下限 | 上限 |
| ベーシック サービス | 有効 | いいえ | 183 | 68.8 | 68.8 | 68.8 | .0 | 2.8 | 63.2 | 74.4 |
| | | | 83 | 31.2 | 31.2 | 100.0 | .0 | 2.8 | 25.6 | 36.8 |
| | | | 266 | 100.0 | 100.0 | | .0 | .0 | 100.0 | 100.0 |
| E-サービス | 有効 | いいえ | 158 | 72.8 | 72.8 | 72.8 | .1 | 3.1 | 66.4 | 78.8 |
| | | | 59 | 27.2 | 27.2 | 100.0 | -.1 | 3.1 | 21.2 | 33.6 |
| | | | 217 | 100.0 | 100.0 | | .0 | .0 | 100.0 | 100.0 |
| プラス サービス | 有効 | いいえ | 237 | 84.3 | 84.3 | 84.3 | .0 | 2.1 | 80.1 | 88.3 |
| | | | 44 | 15.7 | 15.7 | 100.0 | .0 | 2.1 | 11.7 | 19.9 |
| | | | 281 | 100.0 | 100.0 | | .0 | .0 | 100.0 | 100.0 |
| トータル サービス | 有効 | いいえ | 148 | 62.7 | 62.7 | 62.7 | .0 | 3.2 | 56.4 | 69.1 |
| | | | 88 | 37.3 | 37.3 | 100.0 | .0 | 3.2 | 30.9 | 43.6 |
| | | | 236 | 100.0 | 100.0 | | .0 | .0 | 100.0 | 100.0 |

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

度数分布表には、各カテゴリの割合の信頼区間（比率 × 100%）が表示されるため、すべてのカテゴリ変数に使用できます。比較可能な信頼区間は、製品の他の部分には使用できません。

ブートストラップを使用した中央値の信頼区間の取得

従業員レコードを確認した場合、管理者は従業員の以前の職歴に関心を持ちます。職歴は偏っていますが、平均値は中央値に比べて、従業員の「一般的な」以前の職歴について適切な推定値とはなりません。ただし、ブートストラップを使用しない場合、中央値の信頼区間は通常、製品の統計手続きには使用できません。

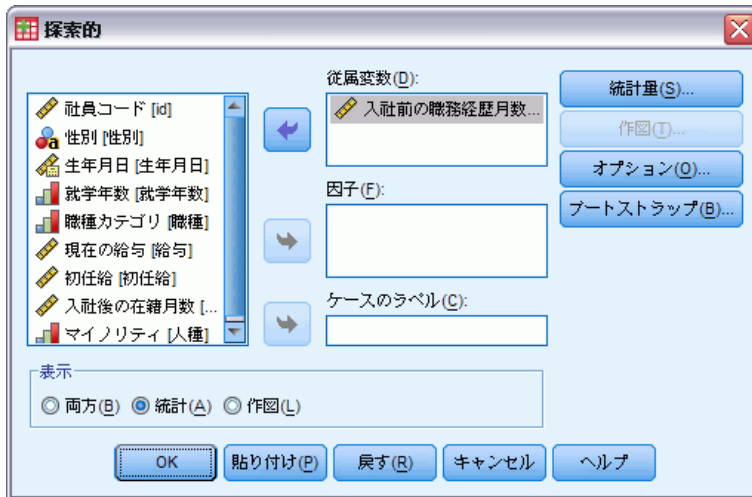
この情報は、Employee data.sav に収集されています。詳細は、[A 付録 p.32 サンプル ファイル](#) を参照してください。ブートストラップを使用した中央値の信頼区間の取得

注：この例では探索的分析手続きを使用し、Statistics Base オプションが必要です。

分析の実行

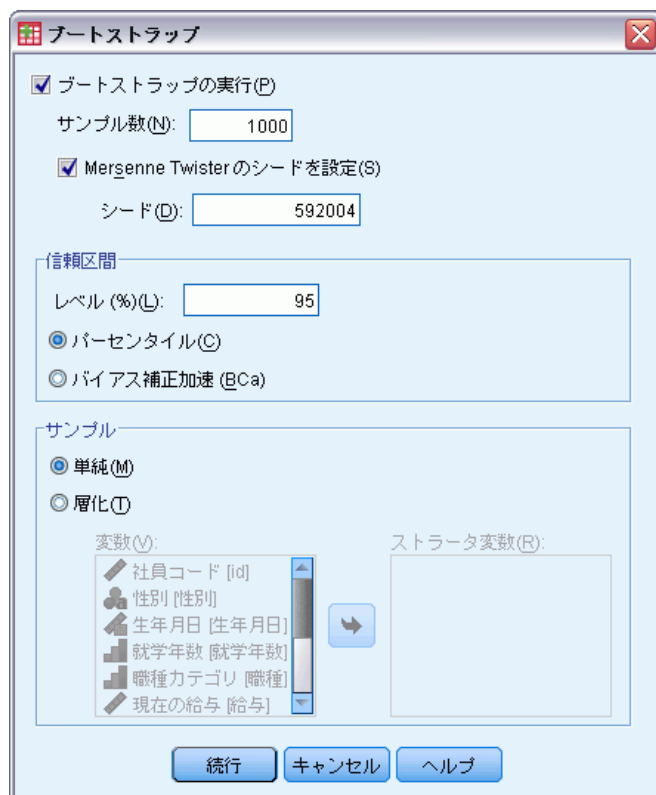
- ▶ 中央値のブートストラップ信頼区間を取得するには、メニューから次の項目を選択します。
分析(A) > 記述統計 > 探索的...

図 3-8
[探索的分析]メイン ダイアログ



- ▶ 従属変数として [経験 (月) 数 [経験]] を選択します。
- ▶ [表示] グループの [統計] を選択します。
- ▶ [ブートストラップ] をクリックします。

図 3-9
[ブートストラップ] ダイアログ ボックス



- ▶ [ブートストラップの実行] を選択します。
- ▶ この例の結果を正確に複製するには、[Mersenne Twister のシードを設定] を選択し、シードとして**592004** と入力します。
- ▶ より正確な区間を（より多くの処理時間を費やして）取得するには、[バイアス補正加速 (BCa)] を選択します。
- ▶ [続行] をクリックします。
- ▶ [探索的分析] ダイアログ ボックスで [OK] をクリックします。

以上の選択により、次のコマンド シンタックスが生成されます。

```
PRESERVE.
SET RNG=MT MTINDEX=592004.
SHOW RNG.
BOOTSTRAP
  /SAMPLING METHOD=SIMPLE
  /VARIABLES TARGET=prevexp
  /CRITERIA CILEVEL=95 CITYPE=BCA NSAMPLES=1000
  /MISSING USERMISSING=EXCLUDE.
EXAMINE VARIABLES=prevexp
```

```
/PLOT NONE  
/STATISTICS DESCRIPTIVES  
/CINTERVAL 95  
/MISSING LISTWISE  
/NOTOTAL.  
RESTORE.
```

- PRESERVE コマンドおよび RESTORE コマンドは、乱数ジェネレータの現在の状態を「記憶」し、ブートストラップが終了するとシステムをその状態に復元します。
- SET コマンドは、乱数ジェネレータを Mersenne Twister に、インデックスを 592004 に設定し、ブートストラップの結果が正確に複製できるようにします。SHOW コマンドは、参照用に出力のインデックスを表示します。
- BOOTSTRAP コマンドは、単純な再抽出を使用して、1000 件のブートストラップ サンプルを要求します。
- VARIABLES サブコマンドは、変数 prevexp を使用して、再抽出用のケース基準を決定します。この変数の欠損値を持つケースは分析から除外されます。
- CRITERIA サブコマンドは、ブートストラップ サンプルの数を要求するほか、デフォルトのパーセンタイル区間の代わりにバイアス補正加速ブートストラップ区間を要求します。
- BOOTSTRAP の後の EXAMINE の手続きは、各ブートストラップのサンプルで実行されます。
- PLOT サブコマンドは、プロット出力をオフにします。
- 他のすべてのオプションには、それぞれのデフォルト値が設定されます。

記述統計量

図 3-10
ブートストラップ信頼区間を表示する記述統計量の表

| | 統計量 | 標準誤差 | ブートストラップ ^a | | | | |
|------------|---------------|-----------|-----------------------|--------|--------------|----------|-----------|
| | | | バイアス | 標準誤差 | BCa 95% 信頼区間 | | |
| | | | | | 下限 | 上限 | |
| 入社前の職務経歴月数 | 平均値 | 95.86 | 4.804 | | | | |
| | 平均値の 95% 信頼区間 | 下限 | 86.42 | | | 86.39 | 105.20 |
| | | 上限 | 105.30 | | | | |
| | 5%トリム平均 | 84.64 | | .02 | 4.94 | 75.38 | 94.21 |
| | 中央値 | 55.00 | | -.11 | 3.66 | 50.00 | 60.00 |
| | 分散 | 10938.281 | | 18.783 | 977.081 | 8954.509 | 13057.229 |
| | 標準偏差 | 104.586 | | -.015 | 4.689 | 94.644 | 114.245 |
| | 最小値 | 0 | | | | | |
| | 最大値 | 476 | | | | | |
| | 範囲 | 476 | | | | | |
| | 4分位範囲 | 121 | | -1 | 10 | 103 | 137 |
| | 歪度 | 1.510 | .112 | .006 | .110 | 1.284 | 1.768 |
| | 尖度 | 1.696 | .224 | .040 | .463 | .823 | 2.876 |

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

記述統計表は、さまざまな統計量と、これらの統計量のブートストラップ信頼区間を示します。中央値のブートストラップ信頼区間（86.39, 105.20）は、パラメトリック信頼区間（86.42, 105.30）と類似しており、「通常の」従業員がおおよそ 7 ～ 9 年の経験があることを示します。ただし、[経験（月数）] の分布は歪んでおり、平均値は中央値と比べて「通常の」現在の給与を示す指標としては適切ではありません。中央値のブートストラップ信頼区間（50.00, 60.00）は、平均値の信頼区間に比べて値の範囲が狭く、小さくなり、「通常の」従業員にはおおよそ 4 ～ 9 年の経験があることを示します。ブートストラップを使用すると、通常の経験をより適切に示す値の範囲を取得できます。

ブートストラップを使用したより適切な予測値の選択

従業員の記録を確認した場合、管理者は線型モデルを現在の給与と初任給との差分にあてはめて、従業員の給与の増加に関連する要素を決定したいとも考えます。線型モデルをブートストラップする場合、残差やワイルド ブートストラップなど、特別な再抽出方法を使用して、より正確な結果を取得できます。

この情報は、Employee data.sav に収集されています。詳細は、A 付録 p.32 サンプル ファイル を参照してください。

注：この例では GLM 1 変量分散分析手続きを使用し、Statistics Base オプションが必要です。

データの準備

まず、現在の給与と初任給との差を計算する必要があります。

- ▶ メニューから次の項目を選択します。
変換(T) > 変数の計算(C)...

図 3-11
[変数の計算] ダイアログ ボックス



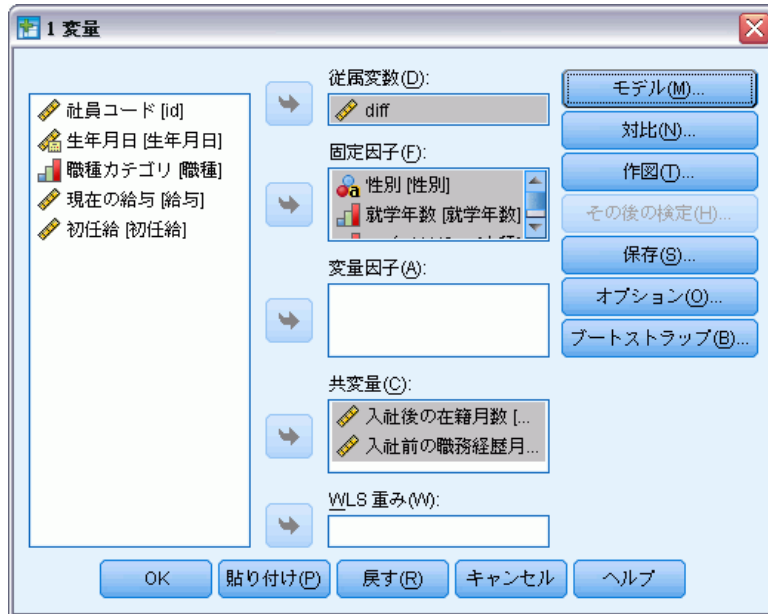
- ▶ 目標変数として diff と入力します。
- ▶ 数式として salary-salbegin と入力します。
- ▶ [OK] をクリックします。

分析の実行

ワイルド残差ブートストラップで GLM 1 変量分散分析を実行する場合、まず残差を作成する必要があります。

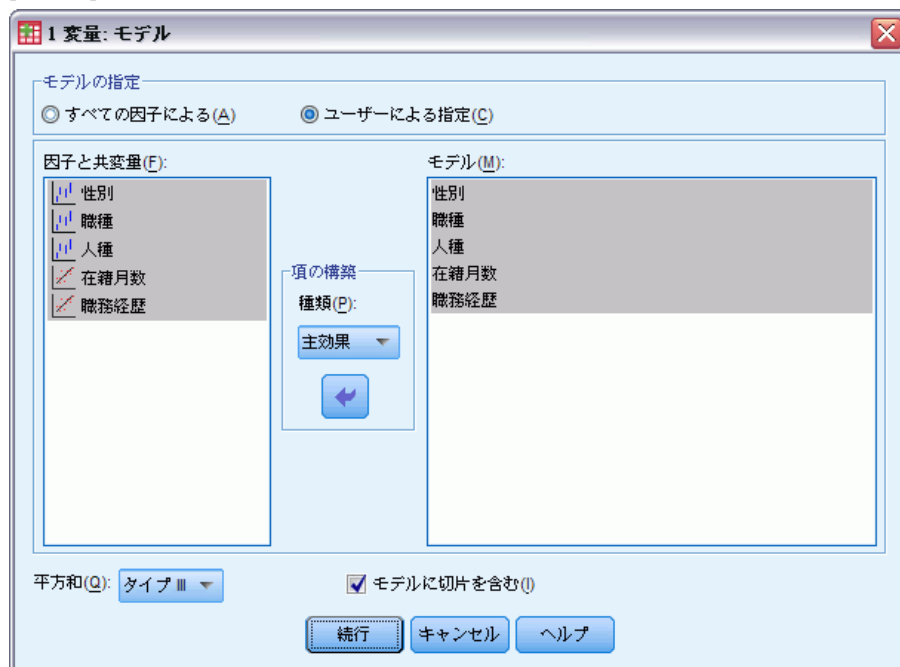
- ▶ メニューから次の項目を選択します。
分析(A) > 一般線型モデル > 1 変量...

図 3-12
[GLM 1 変量] メイン ダイアログ



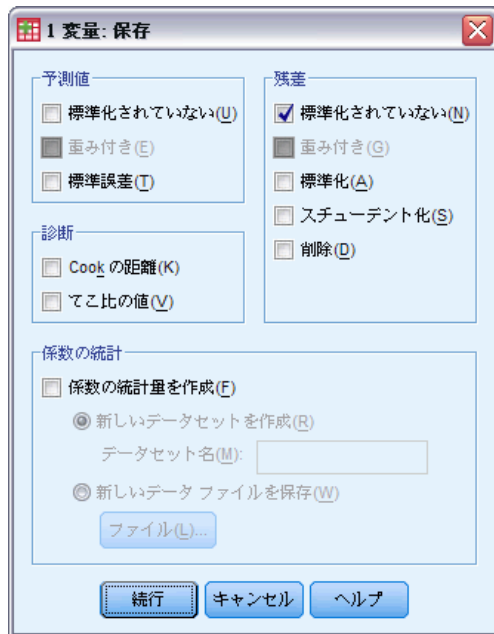
- ▶ [従属変数] として「diff」を選択します。
- ▶ 因子として、「性別 [性別]」、「雇用カテゴリ [雇用カテゴリ]」、および「人種 [人種]」を選択します。
- ▶ [共変量] として「入社後の月数 [勤続期間]」と「経験 (月数)」を選択します。
- ▶ [モデル] をクリックします。

図 3-13
[モデル] ダイアログ ボックス



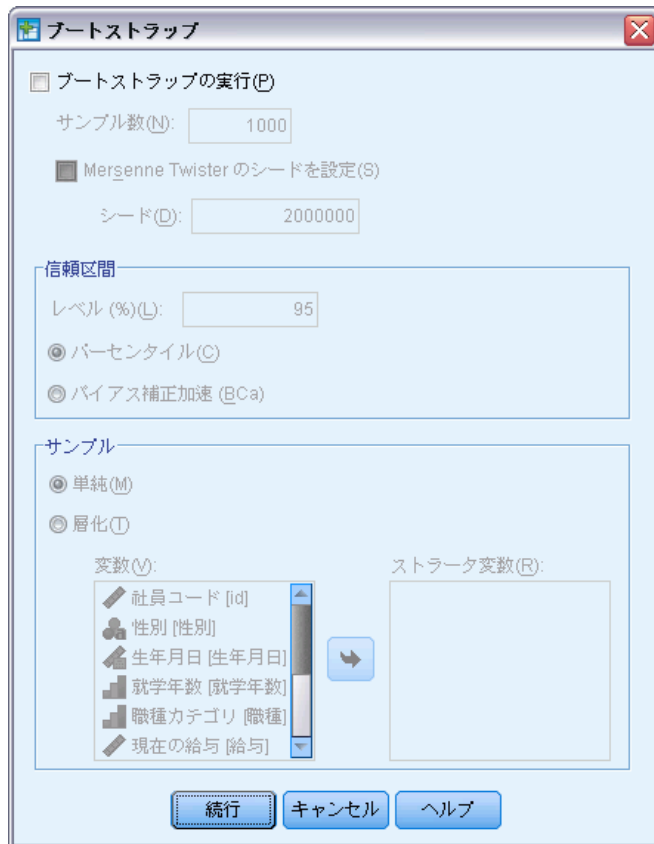
- ▶ [カスタム] を選択し、[項の構築] ドロップダウンから、[主効果] を選択します。
- ▶ モデル項として「性別」～「経験」を選択します。
- ▶ [続行] をクリックします。
- ▶ [GLM 1 変量] ダイアログ ボックスで [保存] をクリックします。

図 3-14
[保存] ダイアログ ボックス



- ▶ [残差] グループの [標準化されていない] を選択します。
- ▶ [続行] をクリックします。
- ▶ [GLM 1 変量] ダイアログ ボックスで [ブートストラップ] をクリックします。

図 3-15
[ブートストラップ] ダイアログ ボックス

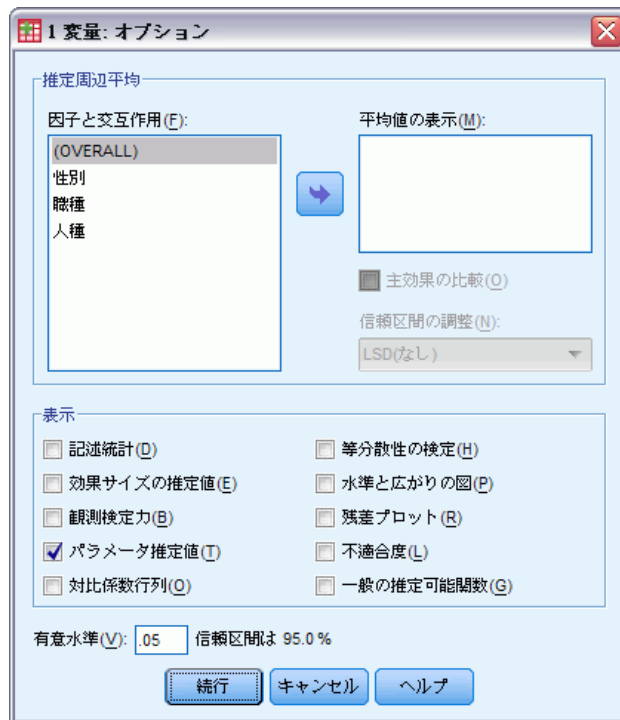


ブートストラップ設定は、ダイアログを通じてブートストラップのサポートを継続します。ブートストラップが有効な場合、新しい変数のデータセットへの保存はサポートされていないため、ブートストラップがオフになっていることを確認する必要があります。

- ▶ 必要に応じて [ブートストラップの実行] をオフにします。
- ▶ [GLM 1 変量] ダイアログ ボックスで [OK] をクリックします。データセットには、新しい変数 RES_1 が追加され、このモデルの標準化されていない残差が含まれます。
- ▶ [GLM 1 変量] ダイアログ ボックスをもう一度呼び出し、[保存] をクリックします。

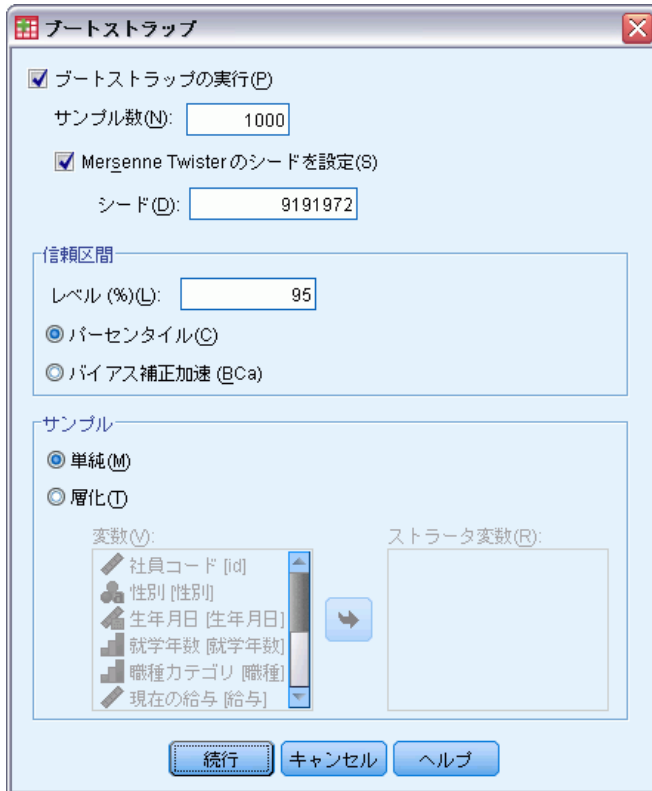
- ▶ [GLM 1 変量] ダイアログ ボックスで [標準化されていない] をオフにし、[続行] をクリックし、[オプション] をクリックします。

図 3-16
[オプション] ダイアログ ボックス



- ▶ [表示] グループの [パラメータ推定] を選択します。
- ▶ [続行] をクリックします。
- ▶ [GLM 1 変量] ダイアログ ボックスで [ブートストラップ] をクリックします。

図 3-17
[ブートストラップ] ダイアログ ボックス



- ▶ [ブートストラップの実行] を選択します。
- ▶ この例の結果を正確に複製するには、[Mersenne Twister のシードを設定] を選択し、シードとして9191972 と入力します。
- ▶ ダイアログを介してワイルド ブートストラップを実行するオプションがないため、[GLM 1 変量] ダイアログ ボックスで [続行] をクリックして、[貼り付け] をクリックします。

以上の選択により、次のコマンド シンタックスが生成されます。

```
PRESERVE.
SET RNG=MT MTINDEX=9191972.
SHOW RNG.
BOOTSTRAP
/SAMPLING METHOD=SIMPLE
/VARIABLES TARGET=diff INPUT=gender jobcat minority jobtime prevexp
/CRITERIA CILEVEL=95 CITYPE=PERCENTILE NSAMPLES=1000
/MISSING USERMISSING=EXCLUDE.
UNIANOVA diff BY gender jobcat minority WITH jobtime prevexp
```

```

/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/PRINT=PARAMETER
/CRITERIA=ALPHA(.05)
/DESIGN=gender jobcat minority jobtime prevexp.
RESTORE.

```

ワイルド ブートストラップ サンプルングを実行するには、`SAMPLING` サブコマンドの `METHOD` キーワードを編集して、`METHOD=WILD (RESIDUALS=RES_1)` を読み込みます。

コマンド シンタックスの最後のセットは次のようになります。

```

PRESERVE.
SET RNG=MT MTINDEX=9191972.
SHOW RNG.
BOOTSTRAP
/SAMPLING METHOD=WILD(RESIDUALS=RES_1)
/VARIABLES TARGET=diff INPUT=gender jobcat minority jobtime prevexp
/CRITERIA CILEVEL=95 CITYPE=PERCENTILE NSAMPLES=1000
/MISSING USERMISSING=EXCLUDE.
UNIANOVA diff BY gender jobcat minority WITH jobtime prevexp
/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/PRINT=PARAMETER
/CRITERIA=ALPHA(.05)
/DESIGN=gender jobcat minority jobtime prevexp.
RESTORE.

```

- `PRESERVE` コマンドおよび `RESTORE` コマンドは、乱数ジェネレータの現在の状態を「記憶」し、ブートストラップが終了するとシステムをその状態に復元します。
- `SET` コマンドは、乱数ジェネレータを Mersenne Twister に、インデックスを 9191972 に設定し、ブートストラップの結果が正確に複製できるようにします。`SHOW` コマンドは、参照用に出力のインデックスを表示します。
- `BOOTSTRAP` コマンドは、ワイルド サンプルングと `RES_1` を残差を含む変数として使用する 1000 件のブートストラップ サンプルを要求します。
- `VARIABLES` サブコマンドは、`diff` が線型モデルの目標変数となるよう指定します。この目標変数と変数 `gender`、`jobcat`、`minority`、`jobtime`、および `prevexp` を使用して、再抽出のケース基準を決定します。これらの変数の欠損値を持つレコードは分析から除外されます。

- CRITERIA サブコマンドは、ブートストラップ サンプルの数を要求するほか、デフォルトのパーセンタイル区間の代わりにバイアス補正加速ブートストラップ区間を要求します。
- BOOTSTRAP の後の UNIANOVA の手続きは、各ブートストラップのサンプルで実行され、元のデータのパラメータ推定を作成します。また、モデル係数のプールされた統計量が作成されます。

パラメータ推定値

図 3-18
パラメータ推定値(M)

従属変数:diff

| パラメータ | B | 標準誤差 | t 値 | 有意確率 | 95% 信頼区間 | |
|--------|----------------|----------|---------|------|------------|------------|
| | | | | | 下限 | 上限 |
| 切片 | 22789.014 | 2920.700 | 7.803 | .000 | 17049.673 | 28528.355 |
| [性別=f] | -4085.253 | 726.416 | -5.624 | .000 | -5512.701 | -2657.804 |
| [性別=m] | 0 ^a | . | . | . | . | . |
| [職種=1] | -17717.706 | 939.798 | -18.853 | .000 | -19564.463 | -15870.949 |
| [職種=2] | -13101.918 | 1780.683 | -7.358 | .000 | -16601.061 | -9602.776 |
| [職種=3] | 0 ^a | . | . | . | . | . |
| [人種=0] | 1332.363 | 819.349 | 1.626 | .105 | -277.705 | 2942.431 |
| [人種=1] | 0 ^a | . | . | . | . | . |
| 在籍月数 | 145.539 | 32.586 | 4.466 | .000 | 81.505 | 209.572 |
| 職務経歴 | -21.423 | 3.575 | -5.993 | .000 | -28.447 | -14.398 |

a.このパラメータは冗長であるためゼロに設定されます。

パラメータ推定値の表には通常、モデル項のブートストラップされていないパラメータ推定値が表示されます。[minority=0] の 0.105 の有意値は 0.05 より大きくなり、人種が給与の増加に影響を与えないことを示します。

図 3-19
ブートストラップ パラメータ推定値

従属変数:diff

| パラメータ | B | ブートストラップ ^a | | | | |
|--------|------------|-----------------------|----------|-----------|--------------|------------|
| | | バイアス | 標準誤差 | 有意確率 (両側) | BCa 95% 信頼区間 | |
| | | | | | 下限 | 上限 |
| 切片 | 22789.014 | -95.084 | 3280.762 | .001 | 16235.191 | 28822.089 |
| [性別=f] | -4085.253 | 32.480 | 622.971 | .001 | -5452.970 | -2824.132 |
| [性別=m] | 0 | 0 | 0 | . | . | . |
| [職種=1] | -17717.706 | 46.324 | 1454.230 | .001 | -20825.508 | -14594.879 |
| [職種=2] | -13101.918 | 47.958 | 1753.311 | .001 | -16713.392 | -9627.308 |
| [職種=3] | 0 | 0 | 0 | . | . | . |
| [人種=0] | 1332.363 | -10.592 | 651.144 | .012 | 99.516 | 2587.473 |
| [人種=1] | 0 | 0 | 0 | . | . | . |
| 在籍月数 | 145.539 | .707 | 35.285 | .001 | 79.220 | 217.588 |
| 職務経歴 | -21.423 | -.065 | 2.859 | .001 | -27.535 | -16.054 |

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

パラメータ推定値表のブートストラップをご覧ください。[標準誤差] 列では、定数項と同じく、いくつかの係数のパラメトリック標準誤差がブートストラップ推定値と比べて小さいため、信頼区間が広がります。[minority=0] のようないくつかの係数について、パラメトリック標準誤差が大きいため、ブートストラップの結果で報告されている 0.006 の有意値が 0.05 より小さく、人種にかかわらず、従業員間の給与の観測された差分が偶然によるものではないことを示します。管理者は、この差について、考えられる原因を特定するためにさらに調査する必要があると理解します。

推奨参考文献

ブートストラップの詳細は、次のテキストを参照してください。

Davison, A. C., および D. V. Hinkley. 2006. *Bootstrap Methods and their Application*. : Cambridge University Press.

Shao, J., および D. Tu. 1995. *The Jackknife and Bootstrap*. New York: Springer.

サンプル ファイル

製品とともにインストールされるサンプル ファイルは、インストールディレクトリの Samples サブディレクトリにあります。[サンプル] サブディレクトリ内に次の各言語の別のフォルダがあります。英語、フランス語、ドイツ語、イタリア語、日本語、韓国語、ポーランド語、ロシア語、簡体字中国語、スペイン語、そして繁体中国語です。

すべてのサンプル ファイルが、すべての言語で使用できるわけではありません。サンプル ファイルがある言語で使用できない場合、その言語のフォルダには、サンプル ファイルの英語バージョンが含まれています。

説明

以下は、このドキュメントのさまざまな例で使用されているサンプル ファイルの簡単な説明です。

- **accidents.sav**。与えられた地域での自動車事故の危険因子を年齢および性別ごとに調べている保険会社に関する架空のデータ ファイルです。各ケースが、年齢カテゴリと性別のクロス分類に対応します。
- **adl.sav**。脳卒中患者に提案される治療の効果を特定するための取り組みに関する架空のデータ ファイルです。医師団は、女性の脳卒中患者たちを、2 つのグループのいずれかにランダムに割り当てました。一方のグループは標準的な理学療法を受け、もう一方のグループは感情面の治療も追加で受けました。治療の 3 か月後に、各患者が日常生活の一般的な行動をどの程度とることができるかを、順序変数として得点付けしました。
- **advert.sav**。広告費とその売上成果の関係を調べるための小売業者の取り組みに関する架空のデータ ファイルです。この小売業者は、そのために、過去の売上と、それに関する広告費のデータを収集しました。
- **aflatoxin.sav**。収穫物によって濃度が大きく異なる毒物であるアフラトキシンを、トウモロコシの収穫物に関して検定することに関する架空のデータ ファイルです。ある穀物加工業者は、8 つそれぞれの収穫物から 16 のサンプルを受け取って、10 億分の 1 単位でアフラトキシン レベルを測定しました。
- **anorectic.sav**。拒食行動または過食行動の標準的な症状の特定を目指して、調査員 (Van der Ham, Meulman, Van Strien, および Van Engeland, 1997) が、摂食障害を持つ大人 55 人の調査を行いました。各患者が 4 年間で 4 回診察を受けたので、観測値は合計で 220 になりました。観

測値ごとに、16 種類の症状に関して患者の得点が記録されました。患者 71 (2 回目)、患者 76 (2 回目)、患者 47 (3 回目) の症状の得点が見つからなかったもので、残っている 217 回分の観測値が有効です。

- **bankloan.sav.** 債務不履行率を低減させるための銀行の取り組みに関する架空のデータ ファイルです。このファイルには、過去の顧客および見込み客 850 人に関する財務情報と人口統計情報が含まれています。最初の 700 ケースは、以前に貸付を行った顧客です。残りの 150 ケースは見込み顧客で、これらの顧客に関して銀行は信用リスクの良し悪しを分類する必要があります。
- **bankloan_binning.sav.** 過去の顧客 5,000 人に関する財務情報と人口統計情報を含む架空のデータ ファイルです。
- **behavior.sav.** 52 人の学生に 15 の状況と 15 の行動の組み合わせについて、0 = 「非常に適切」から 9 = 「非常に不適切」までの 10 段階でランク付けするよう依頼した研究があります (Price および Bouffard, 1974)。個人間の平均を取ったため、値は非類似度としてみなされます。
- **behavior_ini.sav.** このデータ ファイルには、behavior.sav の 2 次元の解の初期配置が含まれています。
- **brakes.sav.** 高性能自動車のディスク ブレーキを生産している工場での品質管理に関する架空のデータ ファイルです。このデータ ファイルには、8 台の機械で生産した 16 個のディスクの直径測定値が含まれています。ブレーキの目標の直径は 322 ミリメートルです。
- **breakfast.sav.** 21 人の Wharton School MBA の学生およびその配偶者に、15 種類の朝食を好みの順に (1 = 「最も好き」から 15 = 「最も嫌い」まで) ランク付けするよう依頼した研究があります (Green および Rao, 1972)。調査対象者の嗜好は、「すべて」から「スナックとドリンクのみ」まで、6 つの異なるシナリオに基づいて記録されました。
- **breakfast-overall.sav.** このデータ ファイルには、最初のシナリオ (「すべて」) のみの朝食の好みが含まれています。
- **broadband_1.sav.** 全国規模のブロードバンド サービスの地域ごとの契約者数を含む架空のデータ ファイルです。このデータ ファイルには、85 地域の月々の契約者数が 4 年間分含まれています。
- **broadband_2.sav.** このデータ ファイルは broadband_1.sav と同じですが、データが 3 か月分追加されています。
- **car_insurance_claims.sav.** 他の場所 (McCullagh および Nelder, 1989) で表示および分析される、自動車の損害請求に関するデータセットです。逆リンク関数を使用して従属変数の平均値を保険契約者の年齢、車種、製造年の線型結合と関連付けることにより、平均請求数はガンマ分布としてモデリングできます。申請された請求の数は、尺度重み付けとして使用できます。
- **car_sales.sav.** このデータ ファイルには、自動車のさまざまな車種やモデルの架空の売上推定値、定価、仕様が含まれています。定価と仕様はそれぞれ、edmunds.com と製造元のサイトから入手しました。

- **car_sales_uprepared.sav.** 変換したバージョンのフィールドを含まない car_sales.sav の修正したバージョンです。
- **carpet.sav.** 一般的な例 (Green および Wind, 1973) としては、新しいカーペット専用洗剤を市販することに関心のある企業が消費者の嗜好に関する 5 種類の因子 (パッケージのデザイン、ブランド名、価格、サービスシール、料金の払い戻し) の影響について調べたい場合があります。パッケージのデザインには、3 つの因子レベルがあります。それぞれ塗布用ブラシの位置が異なります。また、3 つのブランド名 (K2R、Glory、および Bissell)、3 つの価格水準があり、最後の 2 つの因子のそれぞれに対しては 2 つのレベル (「なし」または「あり」) があります。10 人の消費者が、これらの因子により定義された 22 個のプロファイルに順位を付けます。変数「嗜好」には、各プロファイルの平均順位の序列が含まれています。順位が低いほど、嗜好度は高くなります。この変数には、各プロファイルの嗜好測定値がすべて反映されます。
- **carpet_prefs.sav.** このデータ ファイルは carpet.sav と同じ例に基づいていますが、10 人の消費者それぞれから収集した実際のランキングが含まれています。消費者は、22 種類の製品プロファイルを、一番好きなものから一番嫌いなものまで順位付けすることを依頼されています。変数 PREF1 から PREF22 には、carpet_plan.sav で定義されている、関連するプロファイルの ID が含まれています。
- **catalog.sav.** このデータ ファイルには、あるカタログ会社が販売した 3 つの製品の、架空の月間売上高が含まれています。5 つの予測変数のデータも含まれています。
- **catalog_seasfac.sav.** このデータ ファイルは catalog.sav と同じですが、季節性の分解手続きとそれに付随する日付変数から計算した一連の季節因子が追加されています。
- **cellular.sav.** 解約率を削減するための携帯電話会社の取り組みに関する架空のデータ ファイルです。解約の傾向スコアは、0 ~ 100 の範囲でアカウントに適用されます。スコアリングが 50 以上のアカウントはプロバイダの変更を考えている場合があります。
- **ceramics.sav.** 新しい上質の合金に標準的な合金より高い耐熱性があるかどうかを特定するための、ある製造業者の取り組みに関する架空のデータ ファイルです。各ケースが 1 つの合金の別々のテストを表し、軸受けの耐熱温度が記録されます。
- **cereal.sav.** 880 人を対象に、朝食の好みについて、年齢、性別、婚姻状況、ライフスタイルが活動的かどうか (週 2 回以上運動するか) を含めて調査した、架空のデータ ファイルです。各ケースが別々の回答者を表します。
- **clothing_defects.sav.** ある衣料品工場での品質管理工程に関する架空のデータ ファイルです。工場で生産される各ロットから、調査員が衣料品のサンプルを取り出し、不良品の数を数えます。

- **coffee.sav.** このデータ ファイルは、6 つのアイスコーヒー ブランド (Kennedy, Riquier, および Sharp, 1996) について受けた印象に関連しています。回答者は、アイス コーヒーに対する 23 の各印象属性に対して、その属性が言い表していると思われるすべてのブランドを選択しました。機密保持のため、6 つのブランドを AA、BB、CC、DD、EE、および FF で表しています。
- **contacts.sav.** 企業のコンピュータ営業グループの担当者リストに関する架空のデータ ファイルです。各担当者は、所属する会社の部門および会社のランクによって分類されています。また、最新の販売金額、最後の販売以降の経過時間、担当者の会社の規模も記録されています。
- **creditpromo.sav.** 最近のクレジット カード プロモーションの有効性を評価するための、あるデパートの取り組みに関する架空のデータ ファイルです。このために、500 人のカード所有者がランダムに選択されました。そのうち半分には、今後 3 か月間の買い物に関して利率を下げることをプロモーションする広告を送付しました。残り半分には、通常どおりの定期的な広告を送付しました。
- **customer_dbase.sav.** 自社のデータ ウェアハウスにある情報を使用して、反応がありそうな顧客に対して特典を提供するための、ある会社の取り組みに関する架空のデータ ファイルです。顧客ベースのサブセットをランダムに選択して特典を提供し、顧客の反応が記録されています。
- **customer_information.sav.** 名前や住所など、顧客の連絡先情報を含む架空のデータ ファイルです。
- **customer_subset.sav.** customer_dbase.sav の 80 件のケースのサブセット。
- **debate.sav.** 政治討論の出席者に対して行った調査の、討論の前後それぞれの回答に関する架空のデータ ファイルです。各ケースが別々の回答者に対応します。
- **debate_aggregate.sav.** debate.sav 内の回答を集計する、架空のデータ ファイルです。各ケースが、討論前後の好みのクロス分類に対応しています。
- **demo.sav.** 月々の特典を送付することを目的とした、購入顧客のデータベースに関する架空のデータ ファイルです。顧客が特典に反応したかどうか、さまざまな人口統計情報と共に記録されています。
- **demo_cs_1.sav.** 調査情報のデータベースをコンパイルするための、ある会社の取り組みの最初のステップに関する架空のデータ ファイルです。各ケースが別々の都市に対応し、地域、地方、地区、および都市の ID が記録されています。
- **demo_cs_2.sav.** 調査情報のデータベースをコンパイルするための、ある会社の取り組みの第 2 のステップに関する架空のデータ ファイルです。各ケースが、最初のステップで選択した都市の別々の世帯単位に対応し、地域、地方、地区、都市、区画、および単位の ID が記録されます。計画の最初の 2 つの段階からの抽出情報も含まれています。

- **demo_cs.sav**。コンプレックス サンプル計画を使用して収集された調査情報を含む架空のデータ ファイルです。各ケースが別々の世帯単位に対応し、さまざまな人口統計情報および抽出情報が記録されています。
- **dmdata.sav**。これは、ダイレクト マーケティング企業の人口統計情報および購入情報を含む架空のデータです。dmdata2.sav には、テストメールを受け取った連絡先のサブセットの情報を含み、dmdata3.sav には、テストメールを受け取らなかった残りの連絡先に関する情報を含みます。
- **dietstudy.sav**。この架空のデータ ファイルには、” Stillman diet” (Rickman, Mitchell, Dingman, および Dalen, 1974) の研究結果が含まれています。各ケースが別々の被験者に対応し、被験者のダイエット前後の体重 (ポンド単位) と、トリグルセリド レベル (mg/100 ml 単位) が記録されています。
- **dvdplayer.sav**。新しい DVD プレーヤーの開発に関する架空のデータ ファイルです。プロトタイプを使用して、マーケティング チームはフォーカス グループ データを収集しました。各ケースが別々の調査対象ユーザーに対応し、ユーザーの人口統計情報と、プロトタイプに関する質問への回答が記録されています。
- **german_credit.sav**。このデータ ファイルは、カリフォルニア大学アーバイン校の Repository of Machine Learning Databases (Blake および Merz, 1998) にある “German credit” データセットから取ったものです。
- **grocery_1month.sav**。この架空のデータ ファイルは、grocery_coupons.sav データ ファイルの週ごとの購入を「ロールアップ」して、各ケースが別々の顧客に対応するようにしたものです。その結果、週ごとに変わっていた変数の一部が表示されなくなり、買物の総額が、調査を行った 4 週間の買物額の合計になっています。
- **grocery_coupons.sav**。顧客の購買習慣に関心を持っている食料雑貨店チェーンが収集した調査データを含む架空のデータ ファイルです。各顧客を 4 週間に渡って追跡し、各ケースが別々の顧客の週に対応しています。その週に食料品に費やした金額も含め、顧客がいつどこで買物をするかに関する情報が記録されています。
- **guttman.sav**。Bell (Bell, 1961) は、予想される社会グループを示す表を作成しました。Guttman (Guttman, 1968) は、この表の一部を使用しました。この表では、社会相互作用、グループへの帰属感、メンバとの物理的な近接性、関係の形式化などを表す 5 個の変数が、理論上の 7 つの社会グループと交差しています。このグループには、観衆 (例、フットボールの試合の観戦者)、視聴者 (例、映画館または授業の参加者)、公衆 (例、新聞やテレビの視聴者)、暴徒 (観衆に似ているが、より強い相互作用がある)、第一次集団 (親密な関係)、第二次集団 (自発的な集団)、および近代コミュニティ (物理的により密接した近接性と特化されたサービスの必要性によるゆるい同盟関係) があります。

- **health_funding.sav**。医療用資金（人口 100 人あたりの金額）、罹患率（人口 10,000 人あたりの人数）、医療サービス機関への訪問率（人口 10,000 人あたりの人数）のデータを含む、架空のデータ ファイルです。各ケースが別々の都市を表します。
- **hivassay.sav**。HIV 感染を発見する迅速な分析方法を開発するための、ある製薬研究所の取り組みに関する架空のデータ ファイルです。分析の結果は、8 段階の濃さの赤で表現され、色が濃いほど感染の可能性が高くなります。研究所では 2,000 件の血液サンプルに関して試験を行い、その半数が HIV に感染しており、半分は感染していませんでした。
- **hourlywagedata.sav**。管理職から現場担当まで、またさまざまな経験レベルの看護師の時給に関する架空のデータ ファイルです。
- **insurance_claims.sav**。不正請求の恐れがある、疑いを区別するためにモデルを作成する必要がある保険会社の仮説データ ファイルです。各ケースがそれぞれの請求を表します。
- **insure.sav**。10 年満期の生命保険契約に対し、顧客が請求を行うかどうかを示す危険因子を調査している保険会社に関する架空のデータ ファイルです。データ ファイルの各ケースは、年齢と性別が一致する、請求を行った契約と行わなかった契約のペアを表します。
- **judges.sav**。訓練を受けた審判（および 1 人のファン）が 300 件の体操の演技に対して付けた得点に関する架空のデータ ファイルです。各行が別々の演技を表し、審判たちは同じ演技を見ました。
- **kinship_dat.sav**。Rosenberg と Kim (Rosenberg および Kim, 1975) は、15 種類の親族関係用語（祖父、祖母、父、母、叔父、叔母、兄弟、姉妹、いとこ、息子、娘、甥、姪、孫息子、孫娘）の分析を行いました。Rosenberg と Kim は、大学生の 4 つのグループ（女性 2 組、男性 2 組）に、類似性に基づいて上記の用語を並べ替えるよう依頼しました。2 つのグループ（女性 1 組、男性 1 組）には、1 回目と違う条件に基づいて、2 回目の並べ替えをするように頼みました。このようにして、合計で 6 つの「ソース」が取得できました。各ソースは、15 × 15 の近接行列に対応します。この近接行列のセルの数は、ソースの人数から、ソース内でオブジェクトを分割した回数を引いたものです。
- **kinship_ini.sav**。このデータ ファイルには、kinship_dat.sav の 3 次元の解の初期配置が含まれています。
- **kinship_var.sav**。このデータ ファイルには、kinship_dat.sav の解の次元の解釈に使用できる独立変数である性別、世代、および(ation), and 親等が含まれています。特に、解の空間をこれらの変数の線型結合に制限するために使用できます。
- **marketvalues.sav**。1999 ~ 2000 年の間の、イリノイ州アルゴンキンの新興住宅地での住宅売上に関するデータ ファイルです。これらの売上は、公開レコードの問題となります。

- **nhis2000_subset.sav**。National Health Interview Survey (NHIS) は、米国民を対象とした人口ベースの大規模な調査です。全国の代表的な世帯サンプルについて対面式で調査が行われます。各世帯のメンバーに関して、人口統計情報、健康に関する行動および状態の観測値が得られます。このデータ ファイルには、2000 年の調査から得られた情報のサブセットが含まれています。National Center for Health Statistics。National Health Interview Survey, 2000。一般使用データおよびドキュメント。ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/。2003 年にアクセス。
- **ozone.sav**。データには、残りの変数からオゾン濃度を予測するための、6 個の気象変数に対する 330 個の観測値が含まれています。それまでの研究者 (Breiman および Friedman (F), 1985)、(Hastie および Tibshirani, 1990) が、他の研究者と共に、これらの変数間に非線型性を確認しています。この場合、標準的な回帰アプローチは使用できません。
- **pain_medication.sav**。この架空のデータ ファイルには、慢性関節炎を治療する抗炎症薬の臨床試験の結果が含まれています。特に興味深いことは、薬の効果が出るまでの時間と、既存の薬剤との比較です。
- **patient_los.sav**。この架空のデータ ファイルには、心筋梗塞 (MI、または「心臓発作」) の疑いで入院した患者の治療記録が含まれています。各ケースが別々の患者に対応し、入院に関連する多くの変数が記録されています。
- **patlos_sample.sav**。この架空のデータ ファイルには、心筋梗塞 (MI、または「心臓発作」) の治療中に血栓溶解剤を投薬された患者のサンプルの治療記録が含まれています。各ケースが別々の患者に対応し、入院に関連する多くの変数が記録されています。
- **poll_cs.sav**。市民の法案支持率を議会開会前に特定するための、世論調査員の取り組みに関する架空のデータ ファイルです。各ケースは登録有権者に対応しています。ケースごとに、有権者が居住している郡、町、区域が記録されています。
- **poll_cs_sample.sav**。この架空のデータ ファイルには、poll_cs.sav の有権者のサンプルが含まれています。サンプルは、poll_csplan 計画ファイルで指定されている計画に従って抽出され、このデータ ファイルには包含確率およびサンプル重み付けが記録されています。ただし、抽出計画では確率比例 (PPS) 法を使用するため、結合選択確率を含むファイル (poll_jointprob.sav) もあります。サンプル抽出後、有権者の人口統計および法案に関する意見に対応する追加の変数が収集され、データ ファイルに追加されました。
- **property_assess.sav**。限られたリソースで資産価値評価を最新に保つための、郡の評価担当者の取り組みに関する架空のデータ ファイルです。各ケースは、前年に郡内で売却された資産に対応します。データ ファイル内の各ケースでは、資産が存在する町、最後に訪問した評価

担当者、その評価からの経過時間、当時行われた評価、および資産の売却価値が記録されています。

- **property_assess_cs.sav**。限られたリソースで資産価値評価を最新に保つための、州の評価担当者の取り組みに関する架空のデータ ファイルです。各ケースは州内の資産に対応します。データ ファイル内の各ケースでは、資産が存在する郡、町、および区域、最後の評価からの経過時間、および当時行われた評価が記録されています。
- **property_assess_cs_sample.sav**。この架空のデータ ファイルには、property_assess_cs.sav の資産のサンプルが含まれています。サンプルは、property_assess_csplan 計画ファイルで指定されている計画に従って抽出され、このデータ ファイルには包含確率およびサンプル重み付けが記録されています。サンプル抽出後、現在の価値変数が収集され、データ ファイルに追加されました。
- **recidivism.sav**。管轄地域での累犯率を把握するための、政府の法執行機関の取り組みに関する架空のデータ ファイルです。各ケースは元犯罪者に対応し、人口統計情報、最初の犯罪の詳細、初犯から 2 年以内の場合は 2 回目の逮捕までの期間が記録されています。
- **recidivism_cs_sample.sav**。管轄地域での累犯率を把握するための、政府の法執行機関の取り組みに関する架空のデータ ファイルです。各ケースは 2003 年の 7 月に最初の逮捕から釈放された元犯罪者に対応し、人口統計情報、最初の犯罪の詳細、2006 年 7 月までの 2 回目の逮捕のデータが記録されています。犯罪者は recidivism_cs.plan で指定された抽出計画に従って抽出された部門から選択されます。調査では確率比例 (PPS) 法を採用したため、結合選択確率を保持したファイル (recidivism_cs_jointprob.sav) も用意されています。
- **rfm_transactions.sav**。購入日、購入品目、各取引のマネタリー量など、購買取引データを含む架空のデータ ファイルです。
- **salesperformance.sav**。2 つの新しい販売トレーニング コースの評価に関する架空のデータ ファイルです。60 人の従業員が 3 つのグループに分けられ、全員が標準のトレーニングを受けます。さらに、グループ 2 は技術トレーニングを、グループ 3 は実践的なチュートリアルを受けます。トレーニング コースの最後に各従業員がテストを受け、得点が記録されました。データ ファイルの各ケースは別々の訓練生を表し、割り当てられたグループと、テストの得点が記録されています。
- **satisf.sav**。ある小売業者が 4 箇所の店舗で行った満足度調査に関する架空のデータ ファイルです。合計で 582 人の顧客を調査し、各ケースは 1 人の顧客からの回答を表します。
- **screws.sav**。このデータ ファイルには、ねじ、ボルト、ナット、鋸 (びょう) (Hartigan, 1975) の特性に関する情報が含まれています。
- **shampoo_ph.sav**。あるヘアケア製品工場での品質管理に関する架空のデータ ファイルです。定期的に、6 つの異なる製品が測定され、pH が記録されます。目標範囲は 4.5 ~ 5.5 です。

- **ships.sav.** 他の場所 (McCullagh など, 1989) で表示および分析される、波による貨物船への損害に関するデータセットです。件数は、船舶の種類、建造期間、およびサービス期間によって、ポワゾン率で発生するものとしてモデリングできます。因子のクロス分類によって形成されたテーブルの各セルのサービス月数の集計によって、危険にさらされる確率の値が得られます。
- **site.sav.** 業務拡大に向けて新たな用地を選択するための、ある会社の取り組みに関する架空のデータ ファイルです。2 人のコンサルタントを雇って、用地を別々に評価させました。広範囲のレポートに加えて、各用地を「良い」、「普通」、「悪い」のいずれかで集計しました。
- **smokers.sav.** このデータ ファイルは、1998 年の National Household Survey of Drug Abuse から抜粋したものであり、アメリカの世帯の確率サンプルです。(<http://dx.doi.org/10.3886/ICPSR02934>) したがって、このデータ ファイルを分析する場合は、まず人口の傾向を反映させてデータを重み付けする必要があります。
- **stocks.sav** このデータ ファイルには、1 年あたりの在庫価格、量が含まれています。
- **stroke_clean.sav.** この架空のデータ ファイルには、[データの準備] オプションの手続きを使用して整理した後の、医療データベースの状態が含まれています。
- **stroke_invalid.sav.** この架空のデータ ファイルには、医療データベースの初期状態が含まれており、データ入力にいくつかエラーがあります。
- **stroke_survival.** この架空のデータ ファイルは、虚血性脳卒中で数回の困難に直面した後リハビリ プログラムを終えた患者の生存時間に関するものです。脳卒中後、心筋梗塞の発生、虚血性脳卒中、または出血性脳卒中が注意され、イベントの時間が記録されます。脳卒中後に実施されたリハビリ プログラムの最後まで生存した患者のみが含まれるため、サンプルは左側が切り捨てられます。
- **stroke_valid.sav.** この架空のデータ ファイルには、[データの検証] 手続きを使用して確認した後の、医療データベースの状態が含まれています。異常である可能性のあるケースが含まれています。
- **survey_sample.sav.** このデータ ファイルには、人口統計データおよびさまざまな態度指標などの調査データが含まれています。これは「1998 NORC General Social Survey」の変数のサブセットに基づいていますが、いくつかのデータ値が変更され、追加の架空変数がデモの目的で追加されています。
- **telco.sav.** 顧客ベースにおける解約率を削減するための電気通信会社の取り組みに関する架空のデータ ファイルです。各ケースが別々の顧客に対応し、人口統計やサービス利用状況などのさまざまな情報が記録されています。

- **telco_extra.sav.** このデータ ファイルは telco.sav データ ファイルに似ていますが、「期間」および対数変換された顧客支出の属性が削除され、標準化された対数変換顧客支出の変数に置き換えられています。
- **telco_missing.sav.** このデータ ファイルは telco.sav データ ファイルのサブセットですが、一部の人口統計データ値が欠損値に置き換えられています。
- **testmarket.sav.** この架空のデータ ファイルは、新しいメニューを追加しようというファースト フード チェーンの計画に関連しています。新製品をプロモーションするためのキャンペーンには 3 つの候補があるため、新メニューはいくつかのランダムに選択した市場にある場所で紹介されます。場所ごとに別々のプロモーションを使用し、最初の 4 週間の新メニューの週間売上高が記録されます。各ケースが場所と週に対応します。
- **testmarket_1month.sav.** この架空のデータ ファイルは、testmarket.sav データ ファイルの週ごとの売上を「ロールアップ」して、各ケースが別々の場所に対応するようにしたものです。その結果、週ごとに変わっていた変数の一部が表示されなくなり、売上高が、調査を行った 4 週間の売上高の合計になっています。
- **tree_car.sav.** これは、人口統計および自動車購入価格のデータを含む架空のデータ ファイルです。
- **tree_credit.sav.** これは、人口統計および銀行ローン履歴のデータを含む架空のデータ ファイルです。
- **tree_missing_data.sav.** これは、人口統計および銀行ローン履歴のデータと、多数の欠損値を含む架空のデータ ファイルです。
- **tree_score_car.sav.** これは、人口統計および自動車購入価格のデータを含む架空のデータ ファイルです。
- **tree_textdata.sav.** 尺度および値ラベルを割り当てる前の、変数のデフォルトの状態を示すことを主な目的とする、変数を 2 つだけ含む単純なデータ ファイルです。
- **tv-survey.sav.** テレビ スタジオで実施された、ヒットした番組の放送期間を延長するかどうかを検討する調査に関する架空のデータ ファイルです。906 人の回答者に、さまざまな条件下でこの番組を視聴するかどうかを質問しました。各行は別々の回答者を表し、各列は別々の条件を表します。
- **ulcer_recurrence.sav.** このファイルには、潰瘍の再発を防ぐための 2 つの治療の有効性を比較するように計画された調査の情報の一部が含まれています。これは区間調査の良い例であり、他の場所 (Collett, 2003) で表示および分析されています。
- **ulcer_recurrence_recoded.sav.** このファイルでは、ulcer_recurrence.sav の情報が、単に調査終了時のイベント確率ではなく調査の区間ごとのイベント確率をモデリングできるように再編成されています。これは他の場所 (Collett など, 2003) で表示および分析されています。

- **verd1985.sav.** このデータ ファイルは調査 (Verdegaal, 1985) に関連しています。8 つの変数に対する 15 人の被験者の回答を記録しました。対象となる変数が 3 つのグループに分類されます。グループ 1 には「年齢」と「婚姻」、グループ 2 には「ペット」と「新聞」、グループ 3 には「音楽」と「居住地域」がそれぞれ含まれます。「ペット」は多重名義として尺度化され、「年齢」は順序として尺度化されます。また、その他のすべての変数は単一名義として尺度化されます。
- **virus.sav.** 自社のネットワーク上のウィルスの影響を特定するための、インターネット サービス プロバイダ (ISP) の取り組みに関する架空のデータ ファイルです。この ISP は、ネットワーク上の感染した E メール トラフィックの (およその) パーセンテージを、発見の瞬間から脅威が阻止されるまで追跡しました。
- **wheeze_steubenville.sav.** これは、子供 (Ware, Dockery, Spiro III, Speizer, および Ferris Jr., 1984) に対する大気汚染の健康上の影響の長期調査から得られたサブセットです。このデータには、オハイオ州スビューベンビルの 7 歳、8 歳、9 歳、10 歳の子供を対象に行った、喘鳴の状態の反復 2 値測定と、調査の初年に母親が喫煙していたかどうかの固定記録が含まれています。
- **workprog.sav.** 体の不自由な人をより良い仕事に就かせようとする政府の事業プログラムに関する架空のデータ ファイルです。プログラムの参加者候補のサンプルが追跡されました。その中には、ランダムに選ばれてプログラムに登録された人と、そうでない人がいました。各ケースが別々のプログラム参加者を表します。
- **worldsales.sav** このデータ ファイルには、大陸および製品ごとの販売収益が含まれています。

注意事項

この情報は、世界各国で提供される製品およびサービス向けに作成されています。

IBMはこのドキュメントで説明する製品、サービス、機能は他の国では提供していない場合があります。現在お住まいの地域で利用可能な製品、サービス、および、情報については、お近くの IBM の担当者にお問い合わせください。IBM 製品、プログラム、またはサービスに対する参照は、IBM 製品、プログラム、またはサービスのみが使用することができることを説明したり意味するものではありません。IBM の知的所有権を侵害しない機能的に同等の製品、プログラム、またはサービスを代わりに使用することができます。ただし、IBM 以外の製品、プログラム、またはサービスの動作を評価および確認するのはユーザーの責任によるものです。

IBMは、本ドキュメントに記載されている内容に関し、特許または特許出願中の可能性があります。本ドキュメントの提供によって、これらの特許に関するいかなる権利も使用者に付与するものではありません。ライセンスのお問い合わせは、書面にて、下記住所に送ることができます。

IBM Director of Licensing, IBM Corporation, North Castle Drive,
Armonk, NY 10504-1785, U. S. A.

2 バイト文字セット (DBCS) 情報についてのライセンスに関するお問い合わせは、お住まいの国の IBM Intellectual Property Department に連絡するか、書面にて下記宛先にお送りください。

神奈川県大和市下鶴間1623番14号 日本アイ・ビー・エム株式会社 法務・知的財産 知的財産権ライセンス渉外

以下の条項は、イギリスまたはこのような条項が法律に反する他の国では適用されません。 International Business Machines は、明示的または黙示的に関わらず、第三者の権利の侵害しない、商品性または特定の目的に対する適合性の暗黙の保証を含むがこれに限定されない、いかなる保証なく、本出版物を「そのまま」提供します一部の州では、特定の取引の明示的または暗示的な保証の免責を許可していないため、この文が適用されない場合があります。

この情報には、技術的に不適切な記述や誤植を含む場合があります。情報については変更が定期的に行われます。これらの変更は本書の新版に追加されます。IBM は、本書に記載されている製品およびプログラムについて、事前の告知なくいつでも改善および変更を行う場合があります。

IBM 以外の Web サイトに対するこの情報内のすべての参照は、便宜上提供されているものであり、決してそれらの Web サイトを推奨するものではありません。これらの Web サイトの資料はこの IBM 製品の資料に含まれるものではなく、これらの Web サイトの使用はお客様の責任によるものとします。

IBM はお客様に対する一切の義務を負うことなく、自ら適切と考える方法で、情報を使用または配布することができるものとします。

本プログラムのライセンス取得者が (i) 別途作成されたプログラムと他のプログラム（本プログラムを含む）との間の情報交換および (ii) 交換された情報の相互利用を目的とした本プログラムに関する情報の所有を希望する場合、下記住所にお問い合わせください。

IBM Software Group, Attention:Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

上記のような情報は、該当する条項および条件に従い、有料で利用できるものとします。

本ドキュメントに記載されている許可されたプログラムおよびそのプログラムに使用できるすべてのライセンス認証された資料は、IBM Customer Agreement、IBM International Program License Agreement、および当社とかわした同等の契約の条件に基づき、IBM によって提供されます。

IBM 以外の製品に関する情報は、それらの製品の供給業者、公開済みの発表、または公開で使用できるソースから取得しています。IBM は、それらの製品のテストは行っておらず、IBM 以外の製品に関連する性能、互換性、またはその他の要求については確認できません。IBM 以外の製品の性能に関する質問は、それらの製品の供給業者に通知する必要があります。

この情報には、日常の業務処理で用いられるデータや報告書の例が含まれています。できる限り詳細に説明するため、例には、個人、企業、ブランド、製品などの名前が使用されています。これらの名称はすべて架空のものであり、実際の企業で使用される名称および住所とは一切関係ありません。

この情報をソフトコピーでご覧になっている場合は、写真やカラーのイラストが表示されない場合があります。

商標

IBM、IBM ロゴ、および [ibm.com](http://www.ibm.com)、SPSS は、世界の多くの国で登録された IBM Corporation の商標です。IBM の商標の現在のリストは、<http://www.ibm.com/legal/copytrade.shtml> を参照してください。

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel、Intel のロゴ、Intel Inside、Intel Inside のロゴ、Intel Centrino、Intel Centrino のロゴ、Celeron、Intel Xeon、Intel SpeedStep、Itanium、および Pentium は、米国およびその他の国の Intel Corporation または関連会社の商標または登録商標です。

Java およびすべての Java ベースの商標およびロゴは、米国およびその他の国の Sun Microsystems, Inc. の商標です。

Linux は、米国およびその他の国における Linus Torvalds の登録商標です。

Microsoft、Windows、Windows NT、および Windows のロゴは、米国およびその他の国における Microsoft 社の商標です。

UNIX は、米国およびその他の国における The Open Group の登録商標です。

この製品は、WinWrap Basic (Copyright 1993-2007, Polar Engineering and Consulting, <http://www.winwrap.com>) を使用します。

その他の製品名およびサービス名等は、IBM または他の会社の商標です。

Adobe 製品のスクリーンショットは Adobe Systems Incorporated の許可を得て転載しています。

Microsoft 製品のスクリーンショットは Microsoft 社の許可を得て転載しています。



参考文献

- Bell, E. H. 1961. Social foundations of human behavior: Introduction to the study of sociology. New York: Harper & Row.
- Blake, C. L., および C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- Breiman, L., および J. H. Friedman(F). 1985. Estimating optimal transformations for multiple regression and correlation. Journal of the American Statistical Association, 80, .
- Collett, D. 2003. Modelling survival data in medical research, 2 ed. Boca Raton: Chapman & Hall/CRC.
- Davison, A. C., および D. V. Hinkley. 2006. Bootstrap Methods and their Application. : Cambridge University Press.
- Green, P. E., および V. Rao. 1972. Applied multidimensional scaling. Hinsdale, Ill.: Dryden Press.
- Green, P. E., および Y. Wind. 1973. Multiattribute decisions in marketing:A measurement approach. Hinsdale, Ill.: Dryden Press.
- Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. Psychometrika, 33, .
- Hartigan, J. A. 1975. Clustering algorithms. New York: John Wiley and Sons.
- Hastie, T., および R. Tibshirani. 1990. Generalized additive models. London: Chapman and Hall.
- Kennedy, R., C. Riquier, および B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. Journal of Targeting, Measurement, and Analysis for Marketing, 5, .
- McCullagh, P., および J. A. Nelder. 1989. Generalized Linear Models, 2nd ed. London: Chapman & Hall.
- Price, R. H., および D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. Journal of Personality and Social Psychology, 30, .
- Rickman, R., N. Mitchell, J. Dingman, および J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. Journal of the American Medical Association, 228, .
- Rosenberg, S., および M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. Multivariate Behavioral Research, 10, .

Shao, J., および D. Tu. 1995. The Jackknife and Bootstrap. New York: Springer.

Van der Ham, T., J. J. Meulman, D. C. Van Strien, および H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, .

Verdegaal, R. 1985. Meer sets analyse voor kwalitatieve gegevens (in Dutch). Leiden: Department of Data Theory, University of Leiden.

Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, およ
び B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and
respiratory health of children living in six cities. *American Review
of Respiratory Diseases*, 129, .

索引

商標, 44

サンプル ファイル
位置, 32

法律に関する注意事項, 43

中央値の信頼区間
SEPARATE から LAYERED に変更します。 , 21
比率の信頼区間
SEPARATE から LAYERED に変更します。 ,
16-17

パラメータ推定値
SEPARATE から LAYERED に変更します。 , 30

ブートストラップ, 3, 11
サポートされた手順, 5
中央値の信頼区間, 21
比率の信頼区間, 16-17
パラメータ推定値, 30
ブートストラップの指定, 15
ブートストラップの指定
SEPARATE から LAYERED に変更します。 , 15