CONJOINT

This procedure performs conjoint analysis using ordinary least squares.

Notation

The following notation is used throughout this chapter unless otherwise stated:

n	The total number of regular cards in the PLAN file.
p	The total number of factors.
d	The number of discrete factors.
l	The number of linear factors.
q	The number of quadratic factors.
m_i	The number of levels of levels of the *i*th discrete factor.
a_{ij}	The *j*th level of the *i*th discrete factor $(i = 1, \ldots, d)$.
x_i	The *i*th linear factor $(i = 1, \ldots, l)$.
z_i	The *i*th ideal or anti-ideal factor $(i = 1, \ldots, q)$.
r_i	The response for the *i*th card $(i = i, \ldots, n)$.
t	The total number of subjects being analyzed at the same time. (When /SUBJECT is specified, t is usually 1.)

Model

The model for the response r_i for the *i*th card from a subject is

$$r_i = \beta_0 + \sum_{j=1}^{p} u_{jk_{ji}}$$

where $u_{jk_{ji}}$ is the utility (part worth) associated with the k_{ji}th level of the *j*th factor on the *i*th card.

The Design Matrix

A design matrix X is formed from the values in the PLAN file. There is one row for each card in the PLAN file. The columns of the matrix are defined by each of the factor variables in the following manner:

- There is a column of 1s for the constant. This column is used for the estimate of β_0^*.
- For each discrete factor containing m_i levels, $m_i - 1$ columns are formed. Each column represents the deviation of one of the factor levels from the overall mean. There is a 1 in the column if that level of the factor was observed, $a-1$ if the last level of the factor was observed, or a 0 otherwise. These columns are used to estimate the $m_i - 1$ values of α_{ij}.
- For each linear factor, there is one column which is the centered value of that factor $\left(x_{ij} - \bar{x}_i\right)$. These columns are used to estimate the values for $\hat{\beta}_i$.
- For each quadratic factor there are two columns, one which contains the centered value of the factor $\left(z_{ij} - \bar{z}_i\right)$, the next which contains the square of the centered factor value $\left(z_{ij} - \bar{z}_i\right)^2$. These columns are used to estimate the values of $\hat{\gamma}^*$.

Converting Card Numbers to Ranks

If the observations are card numbers, they are converted to ranks. If card number i has a value of k, then $r_i = k$.

Estimation

The estimates

$$\left(\hat{\beta}_0^* \hat{\alpha} \hat{\beta} \hat{\gamma}^*\right)' = \left(X'X\right)^{-1} X'y$$

are computed by using a QR decomposition (see MANOVA) where

$$y_i = \begin{cases} r_i & \text{if responses are scores} \\ n+1-r_i & \text{if responses are ranks} \end{cases}$$

The variance-covariance matrix of these estimates is

$$\hat{\sigma}^2 (X'X)^{-1}$$

where

$$\hat{\sigma}^2 = \sum_{i=1}^{t} \sum_{j=1}^{n} \left(r_{ij} - \hat{r}_{ij} \right)^2 \Big/ (nt - d - l - 2q - 1)$$

The values of $\hat{\gamma}$ are computed by

$$\hat{\gamma}_{i1} = \hat{\gamma}^*_{i1} - 2\hat{\gamma}^*_{i2}\, \bar{z}_i$$

and

$$\hat{\gamma}_{i2} = \hat{\gamma}^*_{i2}$$

with variances

$$\text{var}\left(\hat{\gamma}_{j1}\right) = \text{var}\left(\hat{\gamma}^*_{j1}\right) - 4\bar{z}_j \,\text{cov}\left(\hat{\gamma}^*_{j1}, \hat{\gamma}_{j2}\right) + 4\bar{z}_j^2 \,\text{var}\left(\hat{\gamma}_{j2}\right)$$

and

$$\text{var}\left(\hat{\gamma}_{j2}\right) = \text{var}\left(\hat{\gamma}^*_{j2}\right)$$

where

$$\operatorname{cov}\left(\hat{\gamma}_{j1}, \hat{\gamma}_{j2}\right) = \operatorname{cov}\left(\hat{\gamma}^{*}_{j1}, \hat{\gamma}_{j2}\right) - 2\bar{z}^{2} \operatorname{var}\left(\hat{\gamma}_{j2}\right)$$

The value for $\hat{\beta}_0$ is calculated by

$$\hat{\beta}_0 = \hat{\beta}_0^{*} - \sum \hat{\beta}_i \bar{x}_i - \sum \left(\hat{\gamma}_{i1}\bar{z}_i + \hat{\gamma}_{i2}\bar{z}_i^{2}\right)$$

with variance

$$\operatorname{var}\left(\hat{\beta}_0\right) = a\Sigma_a^{-1}a'$$

where

$$a = \left(1, -\bar{x}_1, \ldots, -\bar{x}_l, -\bar{z}_1, \bar{z}_1^{2}, \ldots, -\bar{z}_q, \bar{z}_q^{2}\right)$$

and

$$\Sigma_a = \begin{pmatrix} \operatorname{var}\hat{\beta}_0^{*} & \operatorname{cov}\left(\hat{\beta}_0^{*}, \hat{\beta}_1\right) & \operatorname{cov}\left(\hat{\beta}_0^{*}, \hat{\gamma}_{q1}^{*}\right) & \operatorname{cov}\left(\hat{\beta}_0^{*}, \hat{\gamma}_{q2}^{*}\right) \\ & \operatorname{var}\hat{\beta}_1 & & \\ & & \operatorname{var}\hat{\gamma}_{q1}^{*} & \\ & & & \operatorname{var}\hat{\gamma}_{q2}^{*} \end{pmatrix}$$

Utility (Part Worth) Values

Discrete Factors

$$\hat{u}_{jk} = \begin{cases} \hat{a}_{jk} & \text{for } k = 1, \ldots, m_j - 1 \\ -\sum_{j=1}^{m_j - 1} \hat{a}_{jk} & \text{for } k = m_j \end{cases}$$

Linear Factors

$$\hat{u}_{jk} = \hat{\beta}_j x_k$$

Ideal or Anti-ideal Factors

$$\hat{u}_{jk} = \hat{\gamma}_{j1} z_{jk} + \hat{\gamma}_{j2} z_{jk}^2$$

Standard Errors of Part Worths

The standard error of part worth $u_{jk} = \sqrt{\text{var}(u_{jk})}$ where $\text{var}(u_{jk})$ is defined below:

Discrete Factors

$$\text{var}(u_{jk}) = \begin{cases} \text{var}(\hat{\alpha}_{jk}) & \text{for } k = 1, \ldots, m_j - 1 \\ \sum_{i=1}^{m_j - 1} \text{var}(\hat{\alpha}_{jk}) - 2 \sum_{i=1}^{m_j - 1} \sum_{l<i} \text{cov}(\hat{\alpha}_{ji}, \hat{\alpha}_{jl}) & \text{for } k = m_j \end{cases}$$

Linear Factors

$$\operatorname{var}\left(u_{jk}\right) = x_k^2 \operatorname{var}\left(\hat{\beta}_j\right)$$

Ideal and Anti-ideal Factors

$$\operatorname{var}\left(u_{jk}\right) = z_k^2 \operatorname{var}\left(\hat{\gamma}_{j1}\right) + 2z_k^3 \operatorname{cov}\left(\hat{\gamma}_{j1}, \hat{\gamma}_{j2}\right) + z_k^4 \operatorname{var}\left(\hat{\gamma}_{j2}\right)$$

Importance Scores

The importance score for factor *i* is

$$IMP_i = 100 \frac{RANGE_i}{\sum_{i=1}^{p} RANGE_i}$$

where $RANGE_i$ is the highest minus lowest utility for factor *i*.

If there is a SUBJECT command, the importance for each factor is calculated separately for each subject, and these are then averaged.

Predicted Scores

$$\hat{r}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{u}_{jk_{ji}}$$

where $\hat{u}_{jk_{ji}}$ is the estimated utility (part worth) associated with the k_{ji}th level of the *j*th factor.

Correlations

Pearson and Kendall correlations are calculated between predicted $(\hat{r}_i)$ and the observed (r_i) responses. See the CORRELATIONS and NONPAR CORR chapters for algorithms. Pearson correlations for holdouts are not calculated.

Simulations

Each person is assigned a probability p_i for each simulation *i*. The probabilities are all computed based on the predicted score $(\hat{r}_i)$ for that product. The probabilities are computed as follows:

Max Utility

$$p_i = \begin{cases} 1 & \text{if } \hat{r}_i = \max(\hat{r}_i) \\ 0 & \text{otherwise} \end{cases}$$

BTL

$$p_i = \frac{\hat{r}_i}{\sum_j \hat{r}_j}$$

Logit

$$p_i = \frac{e^{\hat{r}_i}}{\sum_j e^{\hat{r}_j}}$$

Probabilities are averaged across respondents for the grouped simulation results. For the BTL and Logit methods, only subjects having all positive $\hat{r}_i$ values are used.