LOGLINEAR

The LOGLINEAR procedure models cell frequencies using the multinomial response model and produces maximum likelihood estimates of parameters by the Newton-Raphson method. The contingency tables are converted to two-way $I \times J$ tables, with I and J being the dimensions of the independent and dependent categorical variables respectively.

Notation

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

n_{ij}	Observed frequency of cell (i, j)
I	Dimension of the row variable, associated with independent variables
J	Dimension of the column variable, associated with dependent variables
w_{ij}	Weight of cell (i, j)
β_k	Coefficients in the loglinear model; $1 \le k \le p$
$\beta_k^{(l)}$	Estimate of β_k at the lth iteration
$\hat{\beta}_k$	Final estimate of β_k
m_{ij}	Expected values of n_{ij}
$m_{ij}^{(l)}$	Estimate of m_{ij} at the lth iteration
$\hat{m}_{ij}$	Estimate of m_{ij} at the final iteration
$\hat{M}_{i.}$	$\sum_{j=1}^{J} \hat{m}_{ij}$
$\hat{M}_{.j}$	$\sum_{i=1}^{I} \hat{m}_{ij}$
M	$\sum_{j=1}^{J} \sum_{i=1}^{I} \hat{m}_{ij}$

Model

In the general LOGLINEAR model, the logarithms of the cell frequencies are formulated as a linear function of the parameters. The actual form of the model is determined by the contrast and the effects specified. The model has the form

$$y_{ij} \equiv \ln\left(\frac{m_{ij}}{w_{ij}}\right) = \lambda_i + \sum_{k=1}^{p} \beta_k x_{ijk} \qquad 1 \le i \le I, 1 \le j \le J$$

where λ_i are chosen so that $\sum_j m_{ij} = \sum_j n_{ij}$, and x_{ijk} are the independent variables in the linear model.

Contrasts

The values of x_{ijk} are determined by the types of contrasts specified in the procedure. The default contrast is DEVIATION.

Computational Algorithm

To estimate the coefficients, a series of weighted regressions is used for iterative calculations. The iterative process is outlined (also see Haberman, 1978) as follows:

(1) Obtain initial approximations $y_{ij}^{(0)}$ and use them to obtain $\beta_k^{(0)}$.

(2) Obtain the next approximations $y_{ij}^{(1)}$ and $m_{ij}^{(1)}$.

(3) Use the updated $y_{ij}^{(1)}$ in (2) to obtain the next approximations $\beta_k^{(1)}$.

(4) Repeat steps 2 and 3, replacing $\beta_k^{(l)}$ with $\beta_k^{(l+1)}$. Continue repeating this until convergence is achieved.

The computations begin with selection of initial approximations $m_{ij}^{(0)} = n_{ij} + \delta$ for m_{ij}. The default for δ is 0.5. If the model is saturated, δ is added to n_{ij}

permanently. So, for a saturated model, the observed counts n_{ij} are increased by δ. If the model is not saturated, δ is added to n_{ij} only at the initial step and is then subtracted at the second step.

The maximum likelihood estimates $\hat{\beta}_k$ of β_k are found by the Newton-Raphson method. Let $\beta^{(l)}$ be the column vector containing the ML estimates at the lth iteration; then

$$\beta^{(0)} = \left(C^{(0)}\right)^{-1} a^{(0)}$$

$$\beta^{(l+1)} = \beta^{(l)} + \left(C^{(l+1)}\right)^{-1} a^{(l+1)}, \quad \text{for } l \geq 0,$$

where the (k,l)-element of $C^{(l)}$ is

$$c_{kl}^{(l)} = \sum_{j=1}^{J} \sum_{i=1}^{I} \left(x_{ijk} - \theta_{ik}^{(l)} \right) \left(x_{ijl} - \theta_{il}^{(l)} \right) m_{ij}^{(l)}$$

with

$$\theta_{ik}^{(l)} = \frac{\sum_j m_{ij}^{(l)} x_{ijk}}{\sum_j m_{ij}^{(l)}} \quad \text{for } 1 \leq i \leq I,\ 1 \leq k \leq p$$

and the kth element of $a^{(0)}$ is

$$a_k^{(0)} = \sum_{i,j} x_{ijk} y_{ij}^{(0)} m_{ij}^{(0)} - \frac{\left(\sum_{i,j} x_{ijk} m_{ij}^{(0)} \right) \left(\sum_{i,j} y_{ij} m_{ij}^{(0)} \right)}{\sum_{i,j} m_{ij}^{(0)}}$$

and the kth element of $a^{(l)}$ is

$$a_k^{(l)} = \sum_{i,j} x_{ijk}\left(n_{ij} - m_{ij}^{(l)}\right) \quad \text{for } l \geq 1 .$$

The estimated cell means are updated by

$$m_{ij}^{(l)} = \frac{Tw_{ij}\exp\left(v_{ij}^{(l-1)}\right)}{\sum_{i,j} w_{ij}\exp\left(v_{ij}^{(l-1)}\right)} \quad \text{for } l \geq 1$$

where

$$T = \begin{cases} \sum_{i,j}\left(n_{ij} + \delta\right) & \text{if the model is saturated} \\ \sum_{i,j}\left(n_{ij}\right) & \text{otherwise} \end{cases}$$

and

$$v_{ij}^{(l-1)} = \sum_{k=1}^{p} x_{ijk}\beta_k^{(l-1)}$$

The iterative process stops when either the maximum number of iterations (default=20) is reached or

$$\max_{i,j}\left|v_{ij}^{(l+1)} - v_{ij}^{(l)}\right| < \varepsilon \quad \left(\text{with default } \varepsilon = 0.001\right).$$

Computed Statistics

Correlation Matrix of Parameter Estimates

Let C be the final $C^{(l)}$ and $H = C^{-1}$. The correlation between $\hat{\beta}_i$ and $\hat{\beta}_j$ is computed as

$$\frac{h_{ij}}{\sqrt{h_{ii}h_{jj}}}$$

Goodness of Fit

The Pearson chi-square is computed as

$$\chi^2 = \sum_{i,j} \frac{\left(n_{ij} - \hat{m}_{ij}\right)^2}{\hat{m}_{ij}}$$

and the likelihood-ratio chi-square is

$$L = 2\sum_{i,j} n_{ij} \ln\left(\frac{n_{ij}}{\hat{m}_{ij}}\right)$$

The degrees of freedom are $I \times (J-1) - p - E$, where E is the number of cells with $n_{ij}w_{ij} \le 0$ and p is the number of coefficients in the model.

Residuals

Residuals

$$residual_{ij} = n_{ij} - \hat{m}_{ij}$$

Standardized Residuals

$$standard\ residual_{ij} = \frac{n_{ij} - \hat{m}_{ij}}{\sqrt{\hat{m}_{ij}}}$$

Adjusted Residuals

$$adjusted\ residual_{ij} = \frac{n_{ij} - \hat{m}_{ij}}{\sqrt{s_{ij}}}$$

where

$$s_{ij} = \hat{m}_{ij}\left[1 - \frac{\hat{m}_{ij}}{T} - \hat{m}_{ij}\sum_{k,l}\left(x_{ijk} - \hat{\theta}_{ik}\right)\left(x_{ijl} - \hat{\theta}_{il}\right)h_{kl}\right]$$

$$\hat{\theta}_{ik} = \frac{\sum_{j}\hat{m}_{ij}x_{ijk}}{\sum_{j}\hat{m}_{ij}}$$

Generalized Residuals

Consider a linear combination of the cell counts

$$\sum_{i,j} d_{ij} n_{ij}$$

The estimated expected value is computed as

$$\sum_{i,j} d_{ij} \hat{m}_{ij}$$

Two generalized residuals are computed.

Residuals

$$residual = \sum_{i,j} d_{ij} n_{ij} - \sum_{i,j} d_{ij} \hat{m}_{ij}$$

Adjusted Residuals

$$adjusted\ residual = \frac{\sum_{i,j} d_{ij} n_{ij} - \sum_{i,j} d_{ij} \hat{m}_{ij}}{\sqrt{C_1}}$$

where

$$C_1 = \sum_{i,j} d_{ij}^2 \hat{m}_{ij} - \sum_{i} \left[\frac{\left(\sum_{j} \hat{m}_{ij} d_{ij} \right)^2}{\sum_{j} \hat{m}_{ij}} \right] - \sum_{k=1}^{p} \sum_{l=1}^{p} f_k f_l h_{kl}$$

$$f_k = \sum_{i,j} \hat{m}_{ij} d_{ij} \left(x_{ijk} - \hat{\theta}_{ik} \right)$$

Analysis of Dispersion

Following Haberman (1982), define

$S(Y) =$ Total dispersion

$S(Y|X) =$ Conditional dispersion

$S(X) =$ Dispersion due to fit

$R = \dfrac{S(X)}{S(Y)} =$ Measure of association

For entropy

$$S(Y) = -M \sum_{j=1}^{J} \hat{p}_j \ln\left(\hat{p}_j\right)$$

$$S(Y|X) = -\sum_{i=1}^{I} \hat{M}_{i\bullet} \sum_{j=1}^{J} \hat{p}_{i|j} \ln\left(\hat{p}_{i|j}\right)$$

$$S(X) = S(Y) - S(Y|X)$$

For concentration

$$S(Y) = M \times \left(1 - \sum_{j=1}^{J} \hat{p}_j^2\right)$$

$$S(Y|X) = \sum_{i=1}^{I} \hat{M}_{i\bullet} \left(1 - \sum_{j=1}^{J} \hat{p}_{i|j}^2\right)$$

$$S(X) = S(Y) - S(Y|X)$$

where

$$\hat{p}_j = \frac{\hat{M}_{\bullet j}}{M}$$

$$\hat{p}_{j|i} = \frac{\hat{m}_{ij}}{\hat{M}_{i\bullet}}$$

Haberman (1977) shows that, under the hypothesis that *Y* and *X* are independent,

$$\psi_E = 2S(X) \rightarrow \chi^2_{I(J-1)}$$

in the case of entropy, and

$$\psi_C = \frac{M(J-1)S(X)}{S(Y)} \rightarrow \chi^2_{I-1}$$

in the case of concentration.

References

Haberman, S. J. 1977. Maximum likelihood estimates in exponential response models. *Annals of Statistics*, 5: 815–841.

Haberman, S. J. 1978. *Analysis of qualitative data*, Volume 1. New York: Academic Press.

Haberman, S. J. 1982. Analysis of dispersion of multinomial responses. *Journal of the American Statistical Association*, 77: 568–580.