GENLOG
Multinomial Loglinear and Logit Models

This chapter describes the algorithms used to calculate maximum-likelihood estimates for the multinomial loglinear model and the multinomial logit model. This algorithm is applicable only to aggregated data.

Notation

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

A	Generic categorical independent (explanatory) variable. Its categories are indexed by an array of integers.
B	Generic categorical dependent (response) variable. Its categories are indexed by an array of integers.
r	Number of categories of B.
c	Number of categories of A.
p	Number of nonredundant (nonaliased) parameters.
i	Generic index for the category of B.
j	Generic index for the categories of A.
k	Generic index for the parameter.
n_{ij}	Observed count in the ith response of B and the jth setting of A.
N_j	Marginal total count at the jth setting of A. It is equal to $\sum_{i=1}^{r} n_{ij}$.
N	Total observed count. It is equal to $\sum_{j=1}^{c}\sum_{i=1}^{r} n_{ij}$.
m_{ij}	Expected count.

π_{ij}	Probability of having an observation in the *i*th response of *B* and the *j*th setting of *A*. $0 \le \pi_{ij} \le 1$ and $\sum_{j=1}^{c}\sum_{i=1}^{r}\pi_{ij} = 1$.
z_{ij}	Cell structure value.
α_j	*j*th normalizing constant.
β_k	*k*th nonredundant parameter.
β	A vector of $\left(\beta_1,\ldots,\beta_p\right)'$.
x_{ijk}	An element in the *i*th row and the *k*th column of the design matrix for the *j* setting.

The same notation is used for both loglinear and logit models so that the methods are presented in a unified way. Conceptually, one can consider a loglinear model as a special case of a logit model where the explanatory variable has only one level (that is, $c = 1$).

Components of the Model

There are two components in a loglinear model: the random component and the systematic component.

Random Component

The random component describes the joint distribution of the counts.

- The counts $\left\{n_{1j},\ldots,n_{rj}\right\}$ at the *j*th setting of *A* have the multinomial $\left(N_j,\pi_{1j},\ldots,\pi_{rj}\right)$ distribution.

- The counts n_{ij} and $n_{i'j'}$ are independent if $j \ne j'$.

- The joint probability distribution of $\{n_{ij}\}$ is the product of these *c* independent multinomial distributions. The probability density function is

$$\prod_{j=1}^{c}\left(\frac{N_j!}{\prod_{i=1}^{r} n_{ij}!}\prod_{i=1}^{r}\pi_{ij}^{n_{ij}}\right) \qquad \textbf{(1)}$$

- The expected count is $\mathrm{E}(n_{ij}) = m_{ij} = N_j\pi_{ij}$.

- The covariance is

$$\mathrm{cov}(n_{ij}, n_{i'j'}) = \begin{cases} N_j\pi_{ij}(\delta_{ii'} - \pi_{i'j}) & \text{if } j = j' \\ 0 & \text{if } j \neq j' \end{cases}$$

where $\delta_{ab} = 1$ if $a = b$ and $\delta_{ab} = 0$ if $a \neq b$.

Systematic Component

The systematic component describes the linkage function between the expected counts and the parameters. The expected counts are themselves functions of other parameters. Explicitly, for $i = 1,\ldots,r$ and $j = 1,\ldots,c$,

$$m_{ij} = \begin{cases} z_{ij}e^{\alpha_j + v_{ij}} & \text{if } z_{ij} > 0 \\ 0 & \text{if } z_{ij} \leq 0 \end{cases}$$

where

$$v_{ij} = \sum_{k=1}^{p} x_{ijk}\beta_k$$

SPSS does not consider α_1 to α_c as parameters, but as normalizing constants.

Normalizing Constants

$$\alpha_j = \log\left(\frac{N_j}{\sum_{i=1}^{r} z_{ij} e^{v_{ij}}}\right) \qquad j = 1,\ldots,c \tag{2}$$

Cell Structure Values

The cell structure values play two roles in SPSS loglinear procedures, depending on their signs. If $z_{ij} > 0$, it is a usual weight for the corresponding cell and $\log(z_{ij})$ is sometimes called the **offset**. If $z_{ij} \leq 0$, a **structural zero** is imposed on the cell $(B = i, A = j)$. Contingency tables containing at least one structural zero are called **incomplete tables**. If $n_{ij} = 0$ but $z_{ij} > 0$, the cell $(B = i, A = j)$ contains a **sampling zero**. Although SPSS still considers a structural zero part of the contingency table, it is not used in fitting the model. Cellwise statistics are not computed for structural zeros.

Maximum-Likelihood Estimation

The multinomial log-likelihood is

$$L(\beta) = L(\beta_1,\ldots,\beta_p) = \text{constant} + \sum_{j=1}^{c}\sum_{i=1}^{r} n_{ij} \log(m_{ij}) \tag{3}$$

Likelihood Equations

It can be shown that

$$\frac{\partial L}{\partial \beta_k} = \sum_{j=1}^{c}\sum_{i=1}^{r} (n_{ij} - m_{ij}) x_{ijk} \qquad \textbf{for } k = 1,\ldots,p$$

Let $\mathbf{g}(\beta) = \left(g_1(\beta),\ldots,g_p(\beta)\right)'$ be the $(p+1)$ gradient vector with

$$g_k(\beta) = \frac{\partial L}{\partial \beta_k}$$

The maximum-likelihood estimates $\hat{\beta} = \left(\hat{\beta}_1,\ldots,\hat{\beta}_p\right)^t$ are regarded as a solution to the vector of likelihood equations:

$$\mathbf{g}(\beta) = 0 \qquad \textbf{(4)}$$

Hessian Matrix

The likelihood equations are nonlinear functions of β. Solving them for $\hat{\beta}$ requires an iterative method. The Newton-Raphson method is used. It can be shown that

$$\frac{\partial^2 L}{\partial \beta_k \partial \beta_t} = -\sum_{j=1}^{c}\sum_{i=1}^{r} m_{ij}\left(x_{ijk} - \theta_{jk}\right)\left(x_{ijl} - \theta_{jl}\right)$$

where

$$\theta_{jk} = \frac{1}{N_j}\sum_{i=1}^{r} m_{ij}x_{ijk} \qquad j = 1,\ldots,c \text{ and } k = 1,\ldots,p \qquad \textbf{(5)}$$

Let $\mathbf{H}(\beta)$ be the $p \times p$ information matrix, where $-\mathbf{H}(\beta)$ is the Hessian matrix of (3). The elements of $\mathbf{H}(\beta)$ are

$$h_{kl}(\beta) = -\frac{\partial^2 L}{\partial \beta_k \partial \beta_l} \qquad k = 1,\ldots,p \text{ and } l = 1,\ldots,p \qquad \textbf{(6)}$$

Note: $\mathbf{H}(\beta)$ is a symmetric positive-definite matrix. The asymptotic covariance matrix of $\hat{\beta}$ is estimated by $\mathbf{H}^{-1}(\beta)$.

Newton-Raphson Method

Let $\beta^{(s)}$ denote the *s*th approximation for the solution to (4). By the Newton-Raphson method,

$$\beta^{(s+1)} = \beta^{(s)} + \mathbf{H}^{-1}\left(\beta^{(s)}\right)\mathbf{g}\left(\beta^{(s)}\right)$$

Define $\mathbf{q}(\beta) = \mathbf{H}(\beta)\beta + \mathbf{g}(\beta)$. Using (5) again, the *k*th element of $\mathbf{q}(\beta)$ is

$$q_k(\beta) = \sum_{j=1}^{c}\sum_{i=1}^{r}\eta_{ij}\left(x_{ijk} - \theta_{jk}\right) \qquad \textbf{(7)}$$

where

$$\eta_{ij} = \begin{cases} m_{ij}v_{ij} + \left(n_{ij} - m_{ij}\right) & \text{if } z_{ij} > 0 \text{ and } m_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\mathbf{H}\left(\beta^{(s)}\right)\beta^{(s+1)} = \mathbf{q}\left(\beta^{(s)}\right) \qquad \textbf{(8)}$$

Thus, given $\beta^{(s)}$, the $(s+1)$th approximation $\beta^{(s+1)}$ is found by solving the system of equations in (8).

Initial Values

SPSS uses the $\beta^{(0)}$, which corresponds to a saturated model as the initial value for β. Then the initial estimates for the expected cell counts are

$$m_{ij}^{(0)} = \begin{cases} n_{ij} + \Delta & \text{if } z_{ij} > 0 \\ 0 & \text{if } z_{ij} \le 0 \end{cases} \quad \textbf{(9)}$$

where $\Delta \ge 0$ is a constant.

Note: For saturated models, SPSS adds Δ to n_{ij} if $z_{ij} > 0$. This is done to avoid numerical problems in case some observed counts are 0. We advise users to set Δ to 0 whenever all observed counts (other than structural zeros) are positive.

The initial values for other quantities are

$$\theta_{jk}^{(0)} = \frac{1}{N_j} \sum_{i=1}^{r} m_{ij}^{(0)} x_{ijk} \quad \textbf{(10)}$$

and

$$\eta_{ij}^{(0)} = \begin{cases} m_{ij}^{(0)} \log\left(m_{ij}^{(0)} / z_{ij}\right) + \left(n_{ij} - m_{ij}^{(0)}\right) & \text{if } z_{ij} > 0 \text{ and } m_{ij}^{(0)} > 0 \\ 0 & \text{otherwise} \end{cases} \quad \textbf{(11)}$$

Stopping Criteria

SPSS checks the following conditions for convergence:

1. $\max_{i,j}\left(\left|m_{ij}^{(s+1)} - m_{ij}^{(s)}\right| / m_{ij}^{(s)}\right) < \varepsilon$ provided that $m_{ij}^{(s)} > 0$
2. $\max_{i,j}\left(\left|m_{ij}^{(s+1)} - m_{ij}^{(s)}\right|\right) < \varepsilon$
3. $\sqrt{\left(\sum_{k=1}^{p} g_k^2\left(\hat{\beta}\right)\right) / p} < \varepsilon$

The iteration is said to be **converged** if either conditions 1 and 3 or conditions 2 and 3 are satisfied. If $p = 0$, then condition 3 will be automatically satisfied. The iteration is said to be **not converged** if neither pair of conditions is satisfied within the maximum number of iterations.

Algorithm

The iteration process uses the following steps:

1. Calculate $m_{ij}^{(0)}$ using (9), $\theta_{jk}^{(0)}$ using (10), and $n_{ij}^{(0)}$ using (11).
2. Set $s = 0$.
3. Calculate $\mathbf{H}\left(\beta^{(s)}\right)$ using (6) evaluated at $m_{ij} = m_{ij}^{(s)}$; calculate $\mathbf{q}\left(\beta^{(s)}\right)$ using (7) evaluated at $n_{ij} = n_{ij}^{(s)}$.
4. Solve for $\beta^{(s+1)}$ using (8).
5. Calculate $v_{ij}^{(s+1)} = \sum_{k=1}^{p} x_{ijk}\beta_k^{(s+1)}$ and

$$m_{ij}^{(s+1)} = \begin{cases} N_j\left(z_{ij}e^{v_{ij}^{(s+1)}} \Big/ \left(\sum_{t=1}^{r} z_{tj}e^{v_{ij}^{(s+1)}}\right)\right) & \text{if } z_{ij} > 0 \\ 0 & \text{if } z_{ij} \le 0 \end{cases}$$

6. Check whether the stopping criteria are satisfied. If yes, stop iteration and declare convergence. Otherwise continue.
7. Increase *s* by 1 and check whether the maximum iteration has been reached. If yes, stop iteration and declare the process not converged. Otherwise repeat steps 3-7.

Estimated Normalizing Constants

Using (2), the maximum-likelihood estimate for α_j is

$$\hat{\alpha}_j = \log\left(\frac{N_j}{\sum_{i=1}^{r} z_{ij} e^{\hat{v}_{ij}}}\right) \qquad j = 1,\ldots,c$$

where

$$\hat{v}_{ij} = \sum_{k=1}^{p} x_{ijk}\hat{\beta}_k$$

Estimated Cell Counts

The estimated expected count is

$$\hat{m}_{ij} = \begin{cases} N_j\left(z_{ij}e^{\hat{v}_{ij}} / \left(\sum_{t=1}^{r} z_{tj}e^{\hat{v}_{tj}}\right)\right) & \text{if } z_{ij} > 0 \\ 0 & \text{if } z_{ij} \le 0 \end{cases}$$

Goodness-of-Fit Statistics

The Pearson chi-square statistic is

$$X^2 = \sum_{j=1}^{c}\sum_{i=1}^{r} X_{ij}{}^2$$

where

$$X_{ij}{}^2 = \begin{cases} \left(n_{ij} - \hat{m}_{ij}\right)^2 / \hat{m}_{ij} & \text{if } z_{ij} > 0, n_{ij} > 0, \text{ and } \hat{m}_{ij} > 0 \\ \text{SYSMIS} & \text{if } z_{ij} > 0, n_{ij} > 0, \text{ and } \hat{m}_{ij} = 0 \\ 0 & \text{if } z_{ij} \le 0 \text{ or } n_{ij} = \hat{m}_{ij} \end{cases}$$

If any $X_{ij}{}^2$ is system missing, then X^2 is also system missing.

The likelihood-ratio chi-square statistic is

$$G^2 = 2\sum_{j=1}^{c}\sum_{i=1}^{r} G_{ij}{}^2$$

where

$$G_{ij}{}^2 = \begin{cases} n_{ij}\left(\log\left(n_{ij} / \hat{m}_{ij}\right)\right) & \text{if } z_{ij} > 0, n_{ij} > 0 \text{ and } \hat{m}_{ij} > 0 \\ \text{SYSMIS} & \text{if } z_{ij} > 0, n_{ij} > 0 \text{ and } \hat{m}_{ij} = 0 \\ 0 & \text{if } z_{ij} > 0, n_{ij} = 0, \text{ and } \hat{m}_{ij} \ge 0; \\ & z_{ij} \le 0 \text{ or } n_{ij} = \hat{m}_{ij} \end{cases}$$

If any $G_{ij}{}^2$ is system missing, then G^2 is also system missing.

Degrees of Freedom

The degrees of freedom for each statistic is defined as $a = c(r-1) - p - E$, where E is the number of cells with $z_{ij} \le 0$ or $\hat{m}_{ij} = 0$.

Significance Level

The significance level (or the p value) for the Pearson chi-square statistic is $\text{Prob}\left(\chi_a^2 > X^2\right)$ and that for the likelihood-ratio chi-square statistic is $\text{Prob}\left(\chi_a^2 > G^2\right)$. In both cases, χ_a^2 is the central chi-square distribution with a degrees of freedom.

Analysis of Dispersion (Logit Models Only)

SPSS provides the analysis of dispersion based on two types of dispersion: entropy and concentration. The following definitions are used:

S(A)	Dispersion due to the model
S(B|A)	Dispersion due to residuals
S(B)	Total dispersion
R=S(A)/S(B)	Measure of association

where $S(A) + S(B|A) = S(B)$. Also define

$$\hat{\pi}_i = \frac{\sum_{j=1}^{c} \hat{m}_{ij}}{\sum_{j=1}^{c} N_j}$$

$$\hat{\pi}_{i|j} = \frac{\hat{m}_{ij}}{N_j}$$

The bounds are $0 \le \hat{\pi}_i \le 1$ and $0 \le \hat{\pi}_{i|j} \le 1$.

Entropy

$$S(B) = -N \sum_{i=1}^{r} S_i(B)$$

where

$$S_i(B) = \begin{cases} \hat{\pi}_i \log(\pi_i) & \text{if } 0 < \hat{\pi}_i \le 1 \\ 0 & \text{if } \hat{\pi}_i = 0 \end{cases}$$

and

$$S(B|A) = -\sum_{j=1}^{c} N_j \sum_{i=1}^{r} S_{ij}(B|A)$$

where

$$S_{ij}(B|A) = \begin{cases} \hat{\pi}_{i|j} \log(\hat{\pi}_{i|j}) & \text{if } 0 < \hat{\pi}_{i|j} \le 1 \\ 0 & \text{if } \hat{\pi}_{i|j} = 0 \end{cases}$$

Concentration

$$S(B) = N\left(1 - \sum_{i=1}^{r} \hat{\pi}_i^2\right)$$

$$S(B|A) = \sum_{j=1}^{c} N_j \left(1 - \sum_{i=1}^{r} \hat{\pi}_{i|j}^2\right)$$

Degrees of Freedom

Source of Dispersion	Measure	Degrees of Freedom
Due to model	*S(A)*	$f(r-1)$
Due to residuals	*S(B|A)*	$(N-f-1)(r-1)$
Total	*S(B)*	$(N-1)(r-1)$

where *f* equals *p* minus the number of nonredundant columns (in the design matrix) associated with the main effects of the dependent factors.

Residuals

Goodness-of-fit statistics provide only broad summaries of how models fit data. The pattern of lack of fit is revealed in cell-by-cell comparisons of observed and fitted cell counts.

Simple Residuals

The **simple residual** of the *(i,j)*th cell is

$$r_{ij} = \begin{cases} n_{ij} - \hat{m}_{ij} & \text{if } z_{ij} > 0 \\ \text{SYSMIS} & \text{if } z_{ij} \le 0 \end{cases}$$

Standardized Residuals

The **standardized residual** for the *(i,j)*th cell is

$$r_{ij}^{S} = \begin{cases} \left(n_{ij} - \hat{m}_{ij}\right) / \sqrt{\hat{m}_{ij}\left(1 - \hat{m}_{ij} / N_j\right)} & \text{if } z_{ij} > 0 \text{ and } 0 < \hat{m}_{ij} < N_j \\ 0 & \text{if } z_{ij} > 0 \text{ and } n_{ij} = \hat{m}_{ij} \\ \text{SYSMIS} & \text{otherwise} \end{cases}$$

The standardized residuals are also known as **Pearson residuals** even though $\sum_{j=1}^{c}\sum_{i=1}^{r}\left(r_{ij}^{S}\right)^2 \neq X^2$. Although the standardized residuals are asymptotically normal, their asymptotic variances are less than 1.

Adjusted Residuals

The **adjusted residual** is the simple residual divided by its estimated standard error. Its definition and applications first appeared in Haberman (1973) and re-appeared on page 454 of Haberman (1979). This statistic for the *(i,j)*th cell is

$$r_{ij}^{A} = \begin{cases} \left(n_{ij} - \hat{m}_{ij}\right) / \sqrt{s_{ij}} & \text{if } z_{ij} > 0 \text{ and } \hat{m}_{ij} > 0 \\ 0 & \text{if } z_{ij} > 0 \text{ and } \mathrm{n_{ij}} = \hat{m}_{ij} \\ \text{SYSMIS} & \text{otherwise} \end{cases}$$

where

$$s_{ij} = \hat{m}_{ij}\left(1 - \frac{\hat{m}_{ij}}{N_j} - \hat{m}_{ij}\sum_{k=1}^{p}\sum_{l=1}^{p}\left(x_{ijk} - \hat{\theta}_{jk}\right)\left(x_{ijl} - \hat{\theta}_{jl}\right)h^{kl}\right)$$

h^{kl} is the *(k,l)*th element of $\mathbf{H}^{-1}\left(\hat{\beta}\right)$. The adjusted residuals are asymptotically standard normal.

Deviance Residuals

Pierce and Schafer (1986) and McCullagh and Nelder (1989) define the signed square root of the individual contribution to the G^2 statistic as the **deviance residual**. This statistic for the *(i,j)*th cell is

$$r_{ij}^{D} = \mathrm{sign}\left(n_{ij} - \hat{m}_{ij}\right)\sqrt{d_{ij}}$$

where

$$d_{ij} = \begin{cases} 2\left(n_{ij}\left(\log\left(n_{ij} / \hat{m}_{ij}\right)\right) - \left(n_{ij} - \hat{m}_{ij}\right)\right) & \text{if } z_{ij} > 0, \hat{m}_{ij} > 0, \text{ and } n_{ij} > 0 \\ 2\hat{m}_{ij} & \text{if } z_{ij} > 0, \hat{m}_{ij} \geq 0, \text{ and } \mathrm{n_{ij}} = 0 \\ 0 & \text{if } z_{ij} > 0 \text{ and } \mathrm{n_{ij}} = \hat{m}_{ij} \\ \text{SYSMIS} & \text{otherwise} \end{cases}$$

For multinomial sampling, the individual contribution to the G^2 statistic is only $2n_{ij}\log\left(n_{ij} / \hat{m}_{ij}\right)$, but this is negative when $n_{ij} < \hat{m}_{ij}$. Thus, an extra term $2\left(n_{ij} - \hat{m}_{ij}\right)$ is added to it so that $d_{ij} > 0$ for all *i* and *j*. However, we still have $\sum_{j=1}^{c}\sum_{i=1}^{r}\left(r_{ij}^{D}\right)^2 = G^2$.

Generalized Residual

Consider a linear combination of the cell counts $\sum_{j=1}^{c}\sum_{i=1}^{r} d_{ij} n_{ij}$, where d_{ij} are real numbers.

The estimated expected value is

$$\sum_{j=1}^{c}\sum_{i=1}^{r} d_{ij}\hat{m}_{ij}$$

The simple residual for this linear combination is

$$\sum_{j=1}^{c}\sum_{i=1}^{r} d_{ij}\left(n_{ij}-\hat{m}_{ij}\right)$$

The standardized residual for this linear combination is

$$\frac{\sum_{j=1}^{c}\sum_{i=1}^{r} d_{ij}\left(n_{ij}-\hat{m}_{ij}\right)}{\sqrt{\sum_{j=1}^{c}\left(\sum_{i=1}^{r} d_{ij}^{2}\hat{m}_{ij}-\left(\sum_{i=1}^{r} d_{ij}m_{ij}\right)^{2}/N_{j}\right)}}$$

The adjusted residual for this linear combination is, as given on page 420 of Haberman (1979),

$$\frac{\sum_{j=1}^{c}\sum_{i=1}^{r} d_{ij}\left(n_{ij}-\hat{m}_{ij}\right)}{\sqrt{V}}$$

where

$$V = \sum_{j=1}^{c}\sum_{i=1}^{r} d_{ij}^2 \hat{m}_{ij} - \sum_{j=1}^{c} \frac{1}{N_j}\left(\sum_{i=1}^{r} d_{ij}\hat{m}_{ij}\right)^2 - \sum_{k=1}^{p}\sum_{l=1}^{p} f_k f_l h^{kl}$$

$$f_k = \sum_{j=1}^{c}\sum_{i=1}^{r} d_{ij}\hat{m}_{ij}\left(x_{ijk} - \theta_{ik}\right)$$

Generalized Log-Odds Ratio

Consider a linear combination of the natural logarithm of cell counts

$$\sum_{j=1}^{c}\sum_{i=1}^{r} d_{ij}\log\left(m_{ij}\right) \qquad \textbf{(12)}$$

where d_{ij} are real numbers with the restriction

$$\sum_{i=1}^{r} d_{ij} = 0 \qquad j = 1,\ldots,c$$

The quantity in (12) is estimated by

$$\sum_{j=1}^{c}\sum_{i=1}^{r} d_{ij}\log\left(\hat{m}_{ij}\right) = \sum_{j=1}^{c}\sum_{i=1}^{r} d_{ij}\log\left(z_{ij}\right) + \sum_{j=1}^{c}\sum_{i=1}^{r}\sum_{k=1}^{p} d_{ij}x_{ijk}\hat{\beta}_k \qquad \textbf{(13)}$$

The variance of (13) is

$$\text{var}\left(\sum_{j=1}^{c}\sum_{i=1}^{r} d_{ij}\log\left(\hat{m}_{ij}\right)\right) = \sum_{k=1}^{p}\sum_{l=1}^{p} w_k w_l h^{kl} \qquad \textbf{(14)}$$

where

$$w_k = \sum_{j=1}^{c}\sum_{i=1}^{r} d_{ij} x_{ijk} \qquad k = 1,\ldots,p$$

Wald Statistic

The null hypothesis is

$$H_0: \sum_{j=1}^{c}\sum_{i=1}^{r} d_{ij}\log\left(m_{ij}\right) = 0$$

The Wald statistic is

$$W = \frac{\left(\sum_{j=1}^{c}\sum_{i=1}^{r} d_{ij}\log\left(\hat{m}_{ij}\right)\right)^2}{\sum_{k=1}^{p}\sum_{l=1}^{p} w_k w_l h^{kl}}$$

Under H_0, W asymptotically distributes as a chi-square distribution with 1 degree of freedom. The significance level is $\text{Prob}\left(\chi_1^2 \geq W\right)$. *Note*: W will be system missing if (14) is 0.

Asymptotic Confidence Interval

The asymptotic $(1-\alpha)\times 100\%$ confidence interval for (12) is

$$\sum_{j=1}^{c}\sum_{i=1}^{r} d_{ij}\log\left(\hat{m}_{ij}\right) \pm z_{\alpha/2}\sqrt{\sum_{k=1}^{p}\sum_{l=1}^{p} w_k w_l h^{kl}}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ point of the standard normal distribution. The default value of α is 0.05.

Aggregated Data

This section shows how data are aggregated for a multinomial distribution. The following notation is used in this section:

v_{ij}	Number of SPSS cases for $B=i \quad (i=1,\ldots,r)$ and $A=j \quad (j=1,\ldots,c)$
n_{ijs}	sth SPSS caseweight for $B=i$ and $A=j \quad (s=1,\ldots,v_i)$
x_{ijs}	Covariate
z_{ijs}	Cell weight
c_{ijs}	GRESID coefficient
e_{ijs}	GLOR coefficient
v_{ij}^{+}	Number of positive z_{ijs} (cell weights) for $1 \le s \le v_{ij}$

The cell count is

$$n_{ij} = \begin{cases} \sum_{1 \le s \le v_{ij}}^{*} n_{ijs}^{+} & \text{if } v_{ij}^{+} > 0 \\ 0 & \text{if } v_{ij} = 0 \text{ or } v_{ij}^{+} = 0 \end{cases}$$

where

$$n_{ijs}^{+} = \begin{cases} n_{ijs} & \text{if } n_{ijs} > 0 \text{ and } z_{ijs} > 0 \\ 0 & \text{if } n_{ijs} \le 0 \text{ and } z_{ijs} > 0 \end{cases}$$

and $\sum_{1 \le s \le v_{ij}}^{*}$ means summation over the range of s with the terms $z_{ijs} > 0$.

The cell weight value is

$$z_{ij} = \begin{cases} \sum_{1 \le s \le v_{ij}}^{*} n_{ijs}^{+} z_{ijs} / n_{ij} & \text{if } n_{ij} > 0 \text{ and } v_{ij}^{+} > 0 \\ \sum_{1 \le s \le v_{ij}}^{*} z_{ijs} / v_{ij}^{+} & \text{if } n_{ij} = 0 \text{ and } v_{ij}^{+} > 0 \\ 0 & \text{if } v_{ij}^{+} = 0 \\ 1 & \text{if } v_{ij} = 0 \end{cases}$$

If no variable is specified as the cell weight variable, then all cases have unit cell weights by default.

The cell covariate value is

$$x_{ij} = \begin{cases} \sum_{1 \le s \le v_{ij}}^{*} n_{ijs}^{+} x_{ijs} / n_{ij} & \text{if } n_{ij} > 0 \text{ and } v_{ij} > 0 \\ \sum_{1 \le s \le v_{ij}}^{*} x_{ijs} / v_{ij}^{+} & \text{if } n_{ij} = 0 \text{ and } v_{ij}^{+} > 0 \\ 0 & \text{if } v_{ij}^{+} = 0 \text{ or } v_{ij} = 0 \end{cases}$$

The cell GRESID coefficient is

$$c_{ij} = \begin{cases} \sum_{1 \le s \le v_{ij}}^{*} n_{ijs}^{+} c_{ijs} / n_{ij} & \text{if } n_{ij} > 0 \text{ and } v_{ij} > 0 \\ \sum_{1 \le s \le v_{ij}}^{*} c_{ijs} / v_{ij}^{+} & \text{if } n_{ij} = 0 \text{ and } v_{ij}^{+} > 0 \\ 0 & \text{if } v_{ij}^{+} \text{ or } v_{ij} = 0 \end{cases}$$

There are no defaults for the GRESID coefficients.

The cell GLOR coefficient is

$$e_{ij} = \begin{cases} \sum_{1 \le s \le v_{ij}}^{*} n_{ijs}^{+} e_{ijs} / n_{ij} & \text{if } n_{ij} > 0 \text{ and } v_{ij} > 0 \\ \sum_{1 \le s \le v_{ij}}^{*} e_{ijs} / v_{ij}^{+} & \text{if } n_{ij} = 0 \text{ and } v_{ij}^{+} > 0 \\ 0 & \text{if } v_{ij}^{+} = 0 \text{ or } v_{ij} = 0 \end{cases}$$

There are no defaults for the GLOR coefficients.

References

Agresti, A. 1990. *Categorical data analysis*. New York: John Wiley & Sons, Inc.

Christensen, R. 1990. *Log-linear models*. New York: Springer-Verlag.

Haberman, S. J. 1973. The analysis of residuals in cross-classified tables. *Biometrics*, 29: 205–220.

Haberman, S. J. 1979. *Analysis of qualitative data*, Volume 2. New York: Academic Press.

McCullagh, P., and Nelder, J. A. 1989. Generalized linear models. 2nd ed. London: Chapman and Hall.

Pierce, D. A., and Schafer, D. W. 1986. Residuals in generalized linear models. *Journal of the American Statistical Association*, 81: 977–986.