PLUM

The purpose of the PLUM procedure is to model the dependence of an ordinal categorical response variable on a set of discrete and/or continuous independent variables.

Since the choice and the number of response categories can be quite arbitrary, it is essential to model the dependence such that the choice of the response categories does not affect the conclusion of the inference. That is, the final conclusion should be the same if any two or more adjacent categories of the old scale are combined. Such considerations lead to modeling the dependence of the response on the independent variables by means of the cumulative response probability.

Notations

Y	The ordinal response variable, which takes integer values from 1 to J, $J \geq 2$.
J	The number of categories of the ordinal response.
m	The number of subpopulations.
$\mathbf{X}^A$	$m \times p^A$ matrix with vector-element x_i^A, the observed values at the ith subpopulation, determined by the independent variables specified in the command.
$\mathbf{X}$	$m \times p$ matrix with vector-element x_i, the observed values of the location model's independent variables at the ith subpopulation.
$\mathbf{Z}$	$m \times q$ matrix with vector-element z_i, the observed values of the scale model's independent variables at the ith subpopulation.
f_{ijs}	The frequency weight for the s-th observation which belongs to the cell corresponding to $Y = j$ at subpopulation i.
n_{ij}	The sum of frequency weights of the observations that belong to the cell corresponding to $Y = j$ at subpopulation i.
r_{ij}	The cumulative total up to and including $Y = j$ at subpopulation i.
n_i	The marginal frequency of subpopulation i.
n	The sum of all frequency weights.
γ_{ij}	The cumulative response probability up to and including $Y = j$ at subpopulation i.
π_{ij}	The cell response probability corresponding to $Y = j$ at subpopulation i.
θ	$(J-1)\times 1$ vector of threshold parameters in the location part of the model.
β	$p \times 1$ vector of location parameters in the location part of the model.
τ	$q \times 1$ vector of scale parameters in the scale part of the model.
$\mathbf{B} = (\theta^T, \beta^T, \tau^T)^T$	The $\{(J\text{-}1)+p+q\} \times 1$ vector of unknown parameters in the general model.
$\hat{\mathbf{B}} = (\hat{\theta}^T, \hat{\beta}^T, \hat{\tau}^T)^T$	The $\{(J\text{-}1)+p+q\} \times 1$ vector of maximum likelihood estimates of the parameters in the general model.
$\breve{\mathbf{B}} = (\breve{\theta}^T, \breve{\beta}^T)^T$	The $\{(J\text{-}1)+p\} \times 1$ vector of maximum likelihood estimates of the parameters in the location-only model.
$\hat{\gamma}_{ij}$	The cumulative response probability estimate based on the maximum likelihood estimate $\hat{\mathbf{B}}$ in the general model.

$\breve{\gamma}_{ij}$	The cumulative response probability estimate based on the maximum likelihood estimate $\breve{\mathbf{B}}$ in the location-only model.
$\hat{\pi}_{ij}$	The cell response probability estimate based on the maximum likelihood estimate $\hat{\mathbf{B}}$ in the general model.
$\breve{\pi}_{ij}$	The cell response probability estimate based on the maximum likelihood estimate $\breve{\mathbf{B}}$ in the location-only model.
$\hat{e}$	Number of non-redundant parameters in the general model. If all parameters are non-redundant, $\hat{e}$ = (J-1) + p+ q.
$\breve{e}$	Number of non-redundant parameters in the location-only model. If all parameters are non-redundant, $\breve{e}$ = (J-1) + p.

Data Aggregation

Observations with negative or missing frequency weights are discarded. Observations are aggregated by the definition of subpopulations. Subpopulations are defined by the cross-classifications of the set of independent variables specified in the command.

Let n_i be the marginal count of subpopulation i,

$$n_i = \sum_{j=1}^{J} n_{ij}$$

If there is no observation for the cell of $Y = j$ at subpopulation i, it is assumed that $n_{ij} = 0$, provided that $n_i \neq 0$. A non-negative scalar $\delta \in [0, 1)$ may be added to any zero cell (i.e., cell with $n_{ij} = 0$) if its marginal count n_i is nonzero. The value of δ is zero by default.

Data Assumptions

Let $(n_{i1},\ldots,n_{iJ})^{\mathrm{T}}$ be the $J \times 1$ vector of counts for the categories of Y at subpopulation. It is assumed that each $(n_{i1},\ldots,n_{iJ})^{\mathrm{T}}$ is independently multinomial distributed with probability vector $(\pi_{i1},\ldots,\pi_{iJ})^{\mathrm{T}}$ of dimension $J \times 1$ and fixed total n_i.

Model

Let γ_{ij} = Prob($Y \leq j$| $\boldsymbol{x}_i$) be the cumulative response probability for Y, i.e.,

$$\gamma_{ij} = \sum_{l=1}^{j} \pi_{il}$$

for j = 1, …, J-1. Notice that $\gamma_{iJ} = 1$, hence only the first J-1 γ ’s are needed in the model.

General Model

The general model is given by

$$\eta_{ij} = \frac{\theta_j - \boldsymbol{\beta}^{T}\boldsymbol{x}_i}{\sigma(\boldsymbol{z}_i)}$$

where η_{ij} is related to the cumulative probability γ_{ij} by a link function link(γ_{ij}),

$$\eta_{ij} = \text{link}(\gamma_{ij}),$$

for j = 1, …, J-1 and i = 1, …, m. Possible forms of link(γ) are

$$\text{link}(\gamma) = \begin{cases} \log(\frac{\gamma}{1-\gamma}) & \text{Logit link} \\ \log(-\log(1-\gamma)) & \text{Complementary log-log link} \\ -\log(-\log(\gamma)) & \text{Negative Log-log link} \\ \Phi^{-1}(\gamma) & \text{Probit link} \\ \tan(\pi(\gamma-0.5)) & \text{Cauchit (Inverse Cauchy) link} \end{cases}.$$

The numerator in the right hand side of the general model specifies the *location* of the model, $\theta_j - \boldsymbol{\beta}^{T}\boldsymbol{x}_i$. In the location part of the model, $\boldsymbol{\theta}$ is the vector of thresholds. Values of the thresholds are subject to a monotonicity property $\theta_1 \leq \ldots \leq \theta_{J-1}$. $\boldsymbol{\beta}$ is the vector of location parameters.

The denominator, $\sigma(z_i)$, is the *scale*. Possible forms of $\sigma(z)$ are

$$\sigma(\boldsymbol{z}) = \begin{cases} 1 & \text{if unity scale is assumed} \\ \exp(\tau^{T}\boldsymbol{z}) & \text{if non-constant scale is assumed} \end{cases}.$$

$\boldsymbol{\tau}$ is the vector of scale parameters.

Location-Only Model

If $\sigma(z_i) = 1$ is assumed, then $\eta_{ij} = \theta_j - \boldsymbol{\beta}^{T}\boldsymbol{x}_i$. The general model is said to reduce to the *location-only model*. The parameter $\mathbf{B}$ reduces to $\mathbf{B} = (\boldsymbol{\theta}^{T}, \boldsymbol{\beta}^{T})^{T}$.

Log-likelihood Function

The log-likelihood of the model is

$$l = \sum_{i=1}^{m} \sum_{j=1}^{J-1} r_{ij}\varphi_{ij} - r_{i(j+1)} g(\varphi_{ij})$$

in which r_{ij} is the cumulative total

$$r_{ij} = \sum_{k=1}^{j} nk,$$

the argument φ_{ij} is given by

$$\varphi_{ij} = \log\left(\frac{\gamma_{ij}}{\gamma_{i\,j+1} - \gamma_{ij}}\right),$$

and the function $g(\varphi)$ is

$$g(\varphi) = \log(1 + \exp(\varphi)) = \log\left(\frac{\gamma_{i\,j+1}}{\gamma_{i\,j+1} - \gamma_{ij}}\right)$$

Notice that a constant term $c = \sum_{i=1}^{m} \log\{n_i!/(n_{i1}!\ldots n_{iJ}!)\}$ which is independent of the unknown parameters has been excluded here. Thus, *l* is in fact the *kernel* of the true log-likelihood function.

Further details of the log-likelihood function can be found at the end of this chapter.

Derivatives of the Log-likelihood Function

Details of derivatives can be found at the end of this chapter.

First Derivative

The first derivative of *l* with respect to $\mathbf{B}_k, k = 1, \ldots, (J-1) + p + q$, is

$$\frac{\partial l}{\partial B_k} = \sum_{i=1}^{m} \sum_{j=1}^{J-1} \frac{\partial l_i}{\partial \varphi_{ij}} U_{ij} Q_{ijk},$$

in which

$$\frac{\partial l_i}{\partial \varphi_{ij}} = r_{ij} - r_{i(j+1)} \frac{\gamma_{ij}}{\gamma_{i\,j+1}},$$

$$U_{ij} = \frac{\gamma_{i\,j+1}}{\gamma_{ij}(\gamma_{i\,j+1} - \gamma_{ij})},$$

and

$$Q_{ij} = P_{ijk}\frac{\partial\gamma_{ij}}{\partial\eta_{ij}} - P_{i\,j+1k}\frac{\gamma_{ij}}{\gamma_{i\,j+1}}\frac{\partial\gamma_{i\,j+1}}{\partial\eta_{i\,j+1}},$$

in which

$$P_{ijk} = \frac{\partial\eta_{ij}}{\partial B_k} = \begin{cases} \dfrac{\delta_{jk}}{\exp(\ {}^{\mathrm{T}}\mathbf{z}_i)} & \text{if } 1 \le k \le (J-1) \\ \dfrac{-x_{i[k-(J-1)]}}{\exp(\ {}^{\mathrm{T}}\mathbf{z}_i)} & \text{if } (J-1)+1 \le k \le (J-1)+p \\ -z_{i[k-\{(J-1)+p\}]}\eta_{ij} & \text{if } (J-1)+p+1 \le k \le (J-1)+p+q \end{cases},$$

$\delta_{jk} = 1$ if $j = k$, 0 otherwise, and $P_{iJk} = 0$. For $i = 1, \ldots, m$, $j = 1, \ldots, J$-1,

$$\frac{\partial\gamma_{ij}}{\partial\eta_{ij}} = \begin{cases} \gamma_{ij}(1-\gamma_{ij}) & \text{Logit link} \\ -(1-\gamma_{ij})\log(1-\gamma_{ij}) & \text{Complementary log-log link} \\ -\gamma_{ij}\log(\gamma_{ij}) & \text{Negative Log-log link} \\ \phi(\Phi^{-1}(\gamma_{ij})) & \text{Probit link} \\ \cos^2(\pi(\gamma_{ij}-0.5))/\pi & \text{Cauchit link} \end{cases},$$

and $\partial\gamma_{iJ}/\partial\eta_{iJ} = 0$.

Second Derivative

The second derivative is

$$\frac{\partial^2 l}{\partial B_s \partial B_k} = \sum_{i=1}^{m}\sum_{j=1}^{J-1}\left(\frac{\partial^2 l_i}{\partial B_s \partial\varphi_{ij}}U_{ij}Q_{ijk} + \frac{\partial l_i}{\partial\varphi_{ij}}\frac{\partial U_{ij}}{\partial B_s}Q_{ijk} + \frac{\partial l_i}{\partial\varphi_{ij}}U_{ij}\frac{\partial Q_{ijk}}{\partial B_s}\right)$$

for $s, k = 1, \ldots, (J-1)+p+q$. The first term of the equation is

$$\frac{\partial^2 l_i}{\partial B_s \partial\varphi_{ij}}U_{ij}Q_{ijk} = -\frac{r_{i\,j+1}}{\gamma_{i\,j+1}}U_{ij}Q_{ijs}Q_{ijk}.$$

The second term is

$$\frac{\partial l_i}{\partial \varphi_{ij}} \frac{\partial U_{ij}}{\partial B_s} Q_{ijk} = -\left(r_{ij} - r_{i\,j+1} \frac{\gamma_{ij}}{\gamma_{i\,j+1}} \right) \left(\frac{1}{\gamma_{ij}^2} U_{ij} Q_{ijs} + \frac{1}{(\gamma_{i\,j+1} - \gamma_{ij})^2} (U_{i\,j+1} Q_{i\,j+1s} - U_{ij} Q_{ijs}) \right) Q_{ijk}$$

To calculate the third term, notice that

$$\begin{aligned} \frac{\partial Q_{ijk}}{\partial B_s} = \quad & \frac{\partial P_{ijk}}{\partial B_s} \frac{\partial \gamma_{ij}}{\partial \eta_{ij}} + P_{ijk} \frac{\partial^2 \gamma_{ij}}{\partial B_s \partial \eta_{ij}} - \frac{\partial P_{i\,j+1k}}{\partial B_s} \frac{\gamma_{ij}}{\gamma_{i\,j+1}} \frac{\partial \gamma_{i\,j+1}}{\partial \eta_{i\,j+1}} \\ & - P_{i\,j+1k} \frac{Q_{ijs}}{\gamma_{i\,j+1}} \frac{\partial \gamma_{i\,j+1}}{\partial \eta_{i\,j+1}} - P_{i\,j+1l} \frac{\gamma_{ij}}{\gamma_{i\,j+1}} \frac{\partial^2 \gamma_{i\,j+1}}{\partial B_s \partial \eta_{i\,j+1}} \end{aligned}$$

where

$$\frac{\partial Pk}{\partial B_s} = \begin{cases} 0 & 1 \le k \le (J-1)+p \text{ and } 1 \le s \le (J-1)+p \\ -z_{i[s-\{(J-1)+p\}]} P_{ijk} & 1 \le k \le (J-1)+p \text{ and } (J-1)+p+1 \le s \le (J-1)+p+q \\ -z_{i[k-\{(J-1)+p\}]} P_{ijs} & (J-1)+p+1 \le k \le (J-1)+p+q \end{cases}$$

and $\partial P_{iJk} / \partial B_s = 0$. Moreover,

$$\frac{\partial^2 \gamma_{ij}}{\partial B_s \partial \eta_{ij}} = R_{ij} \frac{\partial \gamma_{ij}}{\partial \eta_{ij}} P_{ijs}$$

and $\partial^2 \gamma_{iJ} / \partial B_s \partial \eta_{iJ} = 0$. R_{ij} has the following form:

$$R_{ij} = \begin{cases} 1 - 2\gamma_{ij} & \text{Logit link} \\ 1 + \log(1 - \gamma_{ij}) & \text{Complementary log-log link} \\ -(1 + \log \gamma_{ij}) & \text{Negative Log-log link} \\ -\phi(\Phi^{-1}(\gamma_{ij}))\Phi^{-1}(\gamma_{ij}) & \text{Probit link} \\ \sin(2\pi\gamma_{ij}) & \text{Cauchit link} \end{cases} .$$

The third term can be calculated by applying these equations.

Expectation of the Second Derivative

For s, $k = 1, \ldots, (J - 1) + p + q$,

$$
\begin{aligned}
\mathrm{E}\left(\frac{\partial^2 l}{\partial B_s \partial B_k}\right) &= \sum_{i=1}^{m}\sum_{j=1}^{J-1}\mathrm{E}\left(\frac{\partial^2 l_i}{\partial B_s \partial \varphi_{ij}} U_{ij} Q_{ijk}\right) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{J-1}\mathrm{E}\left(-\frac{r_{i\,j+1}}{\gamma_{i\,j+1}} U_{ij} Q_{ijs} Q_{ijk}\right). \\
&= -\sum_{i=1}^{m} n_i \sum_{j=1}^{J-1} U_{ij} Q_{ijs} Q_{ijk}
\end{aligned}
$$

Parameter Estimation

Maximum Likelihood Estimate

To obtain the maximum likelihood estimate of $\mathbf{B}$, a Fisher Scoring iterative estimation method or Newton-Raphson iterative estimation method can be used. Let $\mathbf{B}^{(t)}$ be the parameter vector at iteration t and $\partial l / \partial \mathbf{B}^{(t)}$ be a vector of the first derivatives of l evaluated at $\mathbf{B} = \mathbf{B}^{(t)}$. Moreover, let $\mathbf{A}^{(t)}$ be a $\{(J\text{-}1)+p+q\} \times \{(J\text{-}1)+p+q\}$ matrix such that

$$
\left[\mathbf{A}^{(t)}\right]_{sk} = \begin{cases} -\left.\dfrac{\partial^2 l}{\partial B_s \partial B_k}\right|_{\mathbf{B}=\mathbf{B}^{(t)}} & \text{Newton - Raphson approach} \\ -\mathrm{E}\left.\left(\dfrac{\partial^2 l}{\partial B_s \partial B_k}\right)\right|_{\mathbf{B}=\mathbf{B}^{(t)}} & \text{Fisher Scoring approach} \end{cases}
$$

For a location-only model, the corresponding formulas use the first $(J\text{-}1)+p$ elements of $\partial l / \partial \mathbf{B}^{(t)}$ and the upper $\{(J\text{-}1)+p\} \times \{(J\text{-}1)+p\}$ submatrix of $\mathbf{A}^{(t)}$.

The parameter vector $\mathbf{B}$ at iteration $t+1$ is updated by $\mathbf{B}^{(t+1)}$ where

$$
\mathbf{A}^{(t)}\mathbf{B}^{(t+1)} = \mathbf{A}^{(t)}\mathbf{B}^{(t)} + \xi \frac{\partial l}{\partial \mathbf{B}^{(t)}}
$$

and $\xi > 0$ is a stepping scalar such that $l\left(\mathbf{B}^{(t+1)}\right) - l\left(\mathbf{B}^{(t)}\right) \geq 0$.

Stepping

Use step-halving method if $l\left(\mathbf{B}^{(t+1)}\right) - l\left(\mathbf{B}^{(t)}\right) < 0$. Let V be the maximum number of steps in step-halving, the set of values of ξ is $\{1/2^v$: $v = 0, \ldots, V\text{-}1\}$.

Starting Values of the Parameters

Location-Only Model

If a location-only model is specified, set $\mathbf{B}^{(0)} = \left(\theta^{(0)\mathrm{T}}, \mathbf{0}^{\mathrm{T}}\right)^{\mathrm{T}}$ where

$$\theta_j^{(0)} = \mathrm{link}\left(\frac{\sum_{i=1}^{m}\sum_{k=1}^{j} n_{ik}}{\sum_{i=1}^{m} n_i}\right)$$

for $j = 1, \ldots, J$-1.

General Model

If a general model is specified, first ignore the scale part (i.e., by assuming that $\boldsymbol{\tau} = 0$ and treat the model as if it is a location-only model) and use $\mathbf{B}^{(0)} = \left(\theta^{(0)\mathrm{T}}, \mathbf{0}^{\mathrm{T}}\right)^{\mathrm{T}}$ as the starting value to obtain the maximum likelihood estimate $\breve{\mathbf{B}} = \left(\breve{\theta}^{\mathrm{T}}, \breve{\beta}^{\mathrm{T}}\right)^{\mathrm{T}}$. After $\breve{\mathbf{B}}$ is obtained, find the maximum likelihood estimate $\hat{\mathbf{B}} = \left(\hat{\theta}^{\mathrm{T}}, \hat{\beta}^{\mathrm{T}}, \hat{\tau}^{\mathrm{T}}\right)^{\mathrm{T}}$ of the general model by starting at $\left(\breve{\theta}^{\mathrm{T}}, \breve{\beta}^{\mathrm{T}}, \mathbf{0}^{\mathrm{T}}\right)^{\mathrm{T}}$.

The above practice is essentially the same as taking $\mathbf{B}^{(0)} = \left(\theta^{(0)\mathrm{T}}, \mathbf{0}^{\mathrm{T}}, \mathbf{0}^{\mathrm{T}}\right)^{\mathrm{T}}$. The advantage is that the maximum likelihood estimate $\breve{\mathbf{B}}$ can be obtained in the process of finding $\hat{\mathbf{B}}$.

Ordinal Adjustments for the Threshold Parameters

If the monotonicity property $\theta_1 \le \ldots \le \theta_{J-1}$ is not preserved at the end of any iteration, ad hoc adjustment will be taken before the next iteration starts. If $\theta_j^{(t)} > \theta_{j+1}^{(t)}$ for some *j*, then both $\theta_j^{(t)}$ and $\theta_{j+1}^{(t)}$ are set to $\left(\theta_j^{(t)} + \theta_{j+1}^{(t)}\right)/2$ before the next iteration. This value is then compared with $\theta_{j+2}^{(t)}$ and so on.

Convergence Criteria

Given two convergence criteria $\varepsilon_k > 0$ and $\varepsilon_p > 0$, the iteration is considered to be converged if one of the following criteria are satisfied:

1. $\left|l(\mathbf{B}^{(t+1)}) - l(\mathbf{B}^{(t)})\right| < \varepsilon_k$.
2. $\max_i \left|\mathbf{B}_i^{(t+1)} - \mathbf{B}_i^{t}\right| < \varepsilon_p$.

Statistics

Model Information

Final Model, General

The value of –2log-likelihood of the model is given by

$$-2l(\hat{\mathbf{B}})$$

where $l(\hat{\mathbf{B}})$ is the value of the log-likelihood evaluated at $\hat{\mathbf{B}}$.

Final Model, Location-Only

If unity scale is assumed, the general model reduces to the location-only model. The value of –2log-likelihood of the model is given by

$$-2l(\breve{\mathbf{B}}).$$

Initial Model, Intercept-Only

In the initial model, when the intercepts are the only parameters in the model, the parameter vector is $\mathbf{B}^{(0)} = \left(\theta^{(0)\,\mathrm{T}}, \mathbf{0}^{\mathrm{T}}, \mathbf{0}^{\mathrm{T}}\right)^{\mathrm{T}}$. The value of the –2log-likelihood is

$$-2l(\mathbf{B}^{(0)}).$$

Model Chi-Square

The value of the Model Chi-square statistic is given by the difference between any two nesting models of interest.

General Model versus Intercept-Only Model

The following statistic is available when a general model is specified. The Model Chi-square statistic is given by

$$-2l(\mathbf{B}^{(0)}) - 2l(\hat{\mathbf{B}}).$$

Under that null hypothesis that $H_0\!:\beta = \mathbf{0}$ and $\tau = \mathbf{0}$, the Model Chi-square is asymptotically chi-squared distributed with $\hat{e} - (J-1)$ degrees of freedoms.

Location-Only Model versus Intercept-Only Model

The following statistic is available when a location-only model is specified. The Model Chi-square statistic is given by

$$-2l(\mathbf{B}^{(0)}) - 2l(\breve{\mathbf{B}}).$$

Under that null hypothesis that $H_0\!:\!\boldsymbol{\beta} = \mathbf{0}$, the Model Chi-square is asymptotically chi-squared distributed with $\breve{e}\; - (J-1)$ degrees of freedoms.

General Model versus Location-Only Model

The following statistic is available when a general model is specified. The Model Chi-square statistic is given by

$$-2l(\breve{\mathbf{B}}) - 2l(\hat{\mathbf{B}}).$$

Under that null hypothesis that $H_0\!:\!\boldsymbol{\tau} = \mathbf{0}$, the Model Chi-square is asymptotically chi-squared distributed with $\hat{e} - \breve{e}$ degrees of freedoms.

Likelihood Ratio Test for Equal Slopes Assumption

For location-only model, a likelihood ratio test of parallel lines in the location is performed. If the regression lines are not parallel, the location can be specified as

$$\eta_{ij} = \theta_j - \boldsymbol{\beta}_j^{\mathrm{T}} \mathbf{x}_i$$

for j = 1, …, J-1. That is, the location parameters $\boldsymbol{\beta}_j$ (or slopes) vary with the levels of the response. The parameter for the above "non-parallel" location-only model is $\mathbf{B} = (\boldsymbol{\theta}^{\mathrm{T}}, \boldsymbol{\beta}_j^{\mathrm{T}}, \ldots, \boldsymbol{\beta}_{J-1}^{\mathrm{T}})^{\mathrm{T}}$ which is of dimension $\{(J\text{-}1)+(J\text{-}1)p\} \times 1$. The first derivative $\partial l / \partial \mathbf{B}$ of the log-likelihood is the same as in the "parallel" model, except that $P_{ijk} = \partial \eta_{ij} / \partial B_k$ is replaced by the following:

$$P_{ijk} = \frac{\partial \eta_{ij}}{\partial B_k} = \begin{cases} \delta_{jk} & 1 \le k \le (J-1) \\ -x_{i[k-\{(J-1)+sp\}]} & (J-1)+sp \le k \le (J-1)+sp+p, \quad s = 1, \ldots, (J-2) \end{cases}.$$

Similarly, the expected value of the second derivative is the same as in the parallel model, except that the P_{ijk} is replaced by the above equation.

To test the null hypothesis of parallelism $H_0\!:\!\boldsymbol{\beta}_1 = \ldots = \boldsymbol{\beta}_{J-1}$, find the maximum likelihood estimate $\breve{\mathbf{B}}$ of the parallel location-only model and the maximum likelihood estimate $\breve{\breve{\mathbf{B}}}$ of the non-parallel model. The Model Chi-square statistic is given by

$$-2l(\breve{\mathbf{B}}) - 2l(\breve{\breve{\mathbf{B}}}).$$

Under the null hypothesis, the Model Chi-square statistic is asymptotically chi-squared distributed with (k-2)p degrees of freedoms.

Pseudo R Squares

Cox and Snell's R Square

The Cox and Snell's R^2 for a general model is

$$R_{\mathrm{CS}}^2 = 1 - \left(\frac{L(\mathbf{B}^{(0)})}{L(\hat{\mathbf{B}})} \right)^{\frac{2}{n}}$$

Replace $\hat{\mathbf{B}}$ by $\breve{\mathbf{B}}$ for a location-only model.

Nagelkerke's R Square

The Nagelkerke's R^2 is

$$R_{\mathrm{N}}^2 = \frac{R_{\mathrm{CS}}^2}{1 - L(\mathbf{B}^{(0)})^{2/n}}$$

McFadden's R Square

The McFadden's R^2 for a general model is

$$R_{\mathrm{M}}^2 = 1 - \left(\frac{l(\hat{\mathbf{B}})}{l(\mathbf{B}^{(0)})} \right)$$

Replace $\hat{\mathbf{B}}$ by $\breve{\mathbf{B}}$ for a location-only model.

Predicted Cell Counts & Cumulative Totals

Predicted Cell Counts

The estimated cell response probability based on the maximum likelihood estimate for the general model is

$$\hat{\pi}_{ij} = \begin{cases} \hat{\gamma}_{i1} & j = 1 \\ \hat{\gamma}_{ij} - \hat{\gamma}_{i\,j-1} & j = 2, \ldots, J-1 \\ 1 - \hat{\gamma}_{i\,J-1} & j = J \end{cases}$$

At each subpopulation i, the predicted count for response category $Y = j$ is

$$\hat{n}_{ij} = n_i \hat{\pi}_{ij}$$

The (raw) residual is $n_{ij} - \hat{n}_{ij}$ and the standardized residual is $(n_{ij} - \hat{n}_{ij}) / \sqrt{n_i \hat{\pi}_{ij} (1 - \hat{\pi}_{ij})}$.

Replace $\hat{\gamma}_{ij}$ by $\breve{\gamma}_{ij}$, $\hat{\pi}_{ij}$ by $\breve{\pi}_{ij}$, and $\hat{n}_{ij}$ by $\breve{n}_{ij}$ for a location-only model.

Predicted Cumulative Totals, General Model

The predicted cumulative total up to and including $Y = j$ is

$$\hat{r}_{ij} = n_i \hat{\gamma}_{ij},$$

The (raw) residual is $r_{ij} - \hat{r}_{ij}$ and the standardized residual is $(r_{ij} - \hat{r}_{ij}) / \sqrt{n_i \hat{\gamma}_{ij} (1 - \hat{\gamma}_{ij})}$.

Replace $\hat{\gamma}_{ij}$ by $\breve{\gamma}_{ij}$ and $\hat{r}_{ij}$ by $\breve{r}_{ij}$ for a location-only model.

Goodness of Fit Measures

Pearson Goodness of Fit Measure

The Pearson goodness of fit measure for a general model is

$$X^2 = \sum_{i=1}^{m} \sum_{j=1}^{J} \frac{(n_{ij} - n_i \hat{\pi}_{ij})^2}{n_i \hat{\pi}_{ij}},$$

Under the null hypothesis, the Pearson goodness-of-fit statistic is asymptotically chi-squared distributed with $m(J - 1) - \hat{e}$ degrees of freedom.

Replace $\hat{\pi}_{ij}$ by $\breve{\pi}_{ij}$ and $\hat{e}$ by $\breve{e}$ for a location-only model.

Deviance Goodness of Fit Measure

The Deviance goodness of fit measure for a general model is

$$D = 2 \sum_{i=1}^{m} \sum_{j=1}^{J} n_{ij} \log\left(\frac{n_{ij}}{n_i \hat{\pi}_{ij}} \right)$$

Under the null hypothesis, the Deviance goodness-of-fit statistic is asymptotically chi-squared distributed with $m(J - 1) - \hat{e}$ degrees of freedom.

Replace $\hat{\pi}_{ij}$ by $\breve{\pi}_{ij}$ and $\hat{e}$ by $\breve{e}$ for a location-only model.

Covariance and Correlation Matrices

The estimate of the covariance matrix of $\hat{\mathbf{B}}$ is

$$\mathrm{Cov}(\hat{\mathbf{B}}) = \begin{cases} -\left.\dfrac{\partial^2 l}{\partial\mathbf{B}\partial\mathbf{B}}\right|_{\mathbf{B}=\hat{\mathbf{B}}} & \text{Newton-Raphson method} \\ -\mathrm{E}\left.\left(\dfrac{\partial^2 l}{\partial\mathbf{B}\partial\mathbf{B}}\right)\right|_{\mathbf{B}=\hat{\mathbf{B}}} & \text{Fisher Scoring method} \end{cases}.$$

Let $\hat{\boldsymbol{\sigma}}$ be the $\{(J\text{-}1)+p+q\}\times 1$ vector of the square roots of the diagonal elements in $\mathrm{Cov}(\hat{\mathbf{B}})$. The estimate of the correlation matrix of $\hat{\mathbf{B}}$ is

$$\mathrm{Cor}(\hat{\mathbf{B}}) = \mathrm{Diag}(\hat{\boldsymbol{\sigma}}^{-1})\mathrm{Cov}(\hat{\mathbf{B}})\mathrm{Diag}(\hat{\boldsymbol{\sigma}}^{-1}).$$

Replace $\hat{\mathbf{B}}$ by $\breve{\mathbf{B}}$ and $\hat{\boldsymbol{\sigma}}$ by $\breve{\boldsymbol{\sigma}}$ (a $\{(J-1)+p\}\times 1$ vector) for a location-only model.

Parameter Statistics

An estimate of the standard deviation of $\hat{B}_k$ is $\hat{\sigma}_k$. The Wald statistic for $\hat{B}_k$ is

$$\mathrm{Wald}_k = \frac{\hat{B}_k}{\hat{\sigma}_k}$$

Under the null hypothesis that $H_0\!:B_k = 0$, Wald_k is asymptotically chi-squared distributed with 1 degree of freedom.

Based on the asymptotic normality of the parameter estimate, a 100(1-α)% Wald confidence interval for $\hat{B}_k$ is

$$\hat{B}_k \pm z_{1-\alpha/2}\hat{\sigma}_k$$

where $z_{1-\alpha/2}$ is the upper (1-α/2)100$^{\text{th}}$ percentile of the standard normal distribution.

Replace $\hat{B}_k$ by $\breve{B}_k$ and $\hat{\sigma}_k$ by $\breve{\sigma}_k$ for a location-only model.

Linear Hypothesis Testing

For a general model, let $\mathbf{L}$ be a matrix of coefficients for the linear hypotheses

$$H_0\!:\mathbf{LB} = \mathbf{c}$$

where $\mathbf{c}$ is a $k\times 1$ vector of constants. The Wald statistic for H_0 is

$$\text{Wald}(\mathbf{L},\mathbf{c}) = (\mathbf{L}\hat{\mathbf{B}} - \mathbf{c})^{\text{T}}\{\mathbf{L}\text{Cov}(\hat{\mathbf{B}})\mathbf{L}^{\text{T}}\}^{-1}(\mathbf{L}\hat{\mathbf{B}} - \mathbf{c}).$$

Under the null hypothesis, $\text{Wald}(\mathbf{L},\mathbf{c})$ is asymptotically chi-squared distributed with l degrees of freedom, where l is the rank of $\mathbf{L}$.

Replace $\hat{\mathbf{B}}$ by $\breve{\mathbf{B}}$ for a location-only model.

References

Cox, D.R.(1972). Regression Models and Life Tables (with discussion). *Journal of Royal Statistical Society*. B 46, 1-30.

Goodman, L.A. (1979). Simple Models for the Analysis of Association in Cross-Classifications having Ordinal Categories. *Journal of American Statistical Association*. 74, 537-52.

Goodman, L.A. (1981). Association Models and Canonical Correlation in the Analysis of Cross-Classifications having Ordered Categories. *Journal of American Statistical Association*. 76, 320-34.

Greenland, S. (1994). Alternative Models for Ordinal Logistic Regression. *Statistics in Medicine* 13, 1665-77.

Hosmer. D.W.J. and Lemeshow, S. (1981). Applied Logistic Regression Models. *Biometrics*, 34, 318-327.

Magidson, J. (1995). Introducing a New Graphical Model for the Analysis of an Ordinal Categorical Response – Part I. Journal of Targeting, *Measurement and Analysis for Marketing* (UK), Vol. IV, 2, 133-48.

McCullagh, P. (1980). Regression Models for Ordinal Data. *Journal of Royal Statistical Society*. B, 42, No. 2, 109-142.

Pregibon, D. (1981). Logistic Regression Diagnostics. *Annals of Statistics*, 9, 705-24.

Williams, D.A. (1982). Extra-Binomial Variation in Logistic Linear Models. *Applied Statistics*, 31, 144-48.