Notes on distributions, links, and estimation methods

Du Toit, S.H.C. & Du Toit, M. (2022). Chapel Hill, NC: Scientific Software International, Inc.

1 Introduction

This document contains theoretical information on the models fitted by the program. In Section 2, the general model is defined. This model describes the various levels of the hierarchy and assumptions. As it stands, it is immediately applicable to two- and three-level models with continuous outcomes.

Section 3 contains information on how normally distributed continuous outcomes are handled in the program, while Section 4 deals with outcomes that are not normally distributed. For these models, link functions are used to map a non-linear relationship to a linear one, thus allowing the fitting of a linear model to the data. The link function transforms the probabilities of the levels of a categorical outcome variable to an unbounded continuous scale. Adaptive Gaussian quadrature estimation is employed in the fitting of these models.

This is followed by a discussion of the estimation methods employed in the program. Fit measures, scale and dispersion parameters are discussed in Section 6, and lastly notes on the calculation of intracluster correlations are given in Section 7.

2 Definition of a hierarchical model

2.1 Three-level model

It is assumed that y_{ijk} is an outcome variable, where $i = 1, 2, ..., N$ denotes level-1 units and $j = 1, 2, ..., n_i$ denotes level-2 units, nested within each level-3 unit k. There are i = 1, ..., n_{jk} level-1 units nested within each of j = 1, ..., J_k level-2 units, which in turn are nested within each of k = 1, ..., K level-3 units.

2.1.1 Level-1 model

The level-1 model can be expressed in matrix notation as $y_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\pi} + e_{ijk}$, where $\mathbf{x}'_{ijk}$ represent the predictors and $\boldsymbol{\pi}$ the vector of coefficients to be estimated. The level-1 random effect is denoted by e_{ijk}.

A more familiar way to express the level-1 model is

$$y_{ijk} = \pi_{0jk} + \pi_{1jk}x_{1jk} + \pi_{2jk}x_{2jk} + \ldots + \pi_{pjk}x_{pjk} + e_{ijk}$$

where the π_{ijk}s are the level-1 coefficients and the x_{ijk}s the p predictors at this level of the model. It is assumed that $e_{ijk} \sim N\left(0, \sigma^2\right)$.

2.1.2 Level-2 model

At level-2 each π_{pjk} coefficient becomes and outcome variable:

$$\pi_{pjk} = \beta_{p0k} + \beta_{p1k}x_{1jk} + \beta_{p2k}x_{2jk} + \cdots + \beta_{pq_pk}x_{qjk} + r_{pjk}$$

with β_{pqk} a level-2 coefficient, x_{qjk} a level-2 predictor, and r_{pjk} a level-2 random effect. The vector $\left(r_{0jk}, r_{1jk}, \ldots, r_{Pjk}\right)'$ is assumed to be distributed as multivariate normal where each element has a mean of zero and the variance of r_{pjk} is

$$\text{Var}(r_{pjk}) = \tau_{\pi pp}$$

with

$$\mathrm{Cov}(r_{pjk}, r_{p'jk}) = \tau_{\pi pp'}$$

Representing the covariance between any pair of random effects p and p'. Variances and covariances form the symmetrical level-2 variance-covariance matrix τ, with maximum dimension $(p+1)\times(p+1)$. The dimension of this matrix is a function of the number of level-1 coefficients specified as random. We also note that a different set of level-2 predictors may be used in each of the $P+1$ equations that form the level-2 model.

2.1.3 Level-3 model

At level-3 each β_{pqk} coefficient in the level-2 model in turn becomes an outcome variable:

$$\beta_{pqk} = \gamma_{pq0} + \gamma_{pq1}S_{1k} + \gamma_{pq2}S_{2k} + \cdots + \gamma_{pqS_{pq}}S_{S_{pq}k} + u_{pqk}$$

where γ_{pqs} $(s = 0,1,\ldots,S_{pq})$ is a typical level-3 coefficient associated with a level-3 predictor S_{sk}. The associated level-3 random effect is denoted by u_{pqk}. We again assume that, for each level-3 unit, the vector of level-3 random effects has a multivariate normal distribution with each effect having a mean of zero and a variance $\mathrm{Var}(u_{pqk}) = \tau_{\pi qq}$. The dimension of the symmetric level-3 variance-covariance matrix depends on the number of level-2 coefficients specified as random. Again, different level-3 predictors may be used in each of the level-3 equations.

2.2 Two-level model

For two-level models, the subscript k is omitted and y_{ij} denotes level-1 unit i nested within level-2 unit j. The notation used for a two-level model is slightly different, as shown below.

The level-1 model is expressed as

$$Y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \cdots + \beta_{pj}x_{pij} + r_{ij}$$

with the β_{pj}s representing the level-1 coefficients, x_{pij}s the level-1 predictors, and r_{ij} the level-1 random effect.

At level-2 the β_{pj}s again become the outcome variables:

$$\beta_{pj} = \gamma_{p0} + \gamma_{p1}x_{1j} + \gamma_{p2}Sx_{2j} + \cdots + \gamma_{pS_q}x_{S_q j} + u_{pj}$$

where γ_{ps} denote the level-2 coefficient to be estimated, $x_{S\,j}$ a level-2 predictor, and u_{pj} a level-2 random effect. Assumptions regarding the distribution of the random effects are the same as noted for the three-level model.

A multilevel model with a non-normal outcome variable is transformed to a linear model by using a *link* function which defines the relationship between the dependent variable η_{ijk} of the linear model and the mean μ_{ijk} of the *distribution* selected. Detailed information on non-normally distributed outcomes and the link functions available for each are given in the Section 4.

3 Continuous outcomes

3.1 General model for normally distributed outcomes

We assume the outcome (response) variable to be normally distributed. In general, the probability density for the models considered here can be written as

$$f\left(y_{ijk},\mu_{ijk},\sigma\right)=\frac{1}{\sqrt{2\pi\sigma^{2}}}\exp(-\frac{1}{2}\left(\frac{y_{ijk}-\mu_{ijk}}{\sigma}\right)^{2}$$

Let μ_{ijk} represent the mean of the response variable y_{ijk} :

$$\mu_{ijk}=\frac{1}{N}\sum_{i}\sum_{j}\sum_{k}E\left(y_{ijk}\right)$$

The variance can be expressed as a function of the mean as

$$\operatorname{var}(y_{ijk})=V(\mu_{ijk}).$$

where V is the variance function. These models may be fitted by selecting the Normal (HLM) option from the list of available distributions in the program.

4 Non-continuous outcomes

4.1 General model for non-normally distributed outcomes

We assume the response to possess a probability distribution of exponential form. In general, the probability density for the models considered here can be written as

$$f\left(y_{ijk}\right) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y_{ijk}, \phi)\right)$$

where the functions *a*, *b*, and *c* determine the specific distribution. Let w_{ijk} denote the known weight for an observation. Then we can express the functions *a* and *c* of the form

$$a(\phi) = \phi / w_{ijk}$$

and

$$c = c(y, \phi / w_{ijk})$$

Let μ_{ijk} represent the mean of the response variable y_{ijk}. The variance can be expressed as a function of the mean as

$$\operatorname{var}(y_{ijk}) = \frac{V(\mu_{ijk})\phi}{w_{ijk}}$$

where *V* is the variance function.

In generalized linear models such as these considered here, it is convenient to parameterize the models in terms of the mean μ and the dispersion parameter ϕ instead of the natural parameter θ.

4.2 Link function and derivatives

The link functions available are the log, logistic, complimentary log-log, log-log, probit, power, and identity. Table 1 contains a summary of these link functions by distribution.

Table 1: Distributions and link functions for estimation with adaptive quadrature

	LOG	LOGIT	LOG-LOG	Comp. LOG-LOG	PROBIT	POWER	IDENTITY	LOG+LOGIT
Poisson	x							
Bernoulli		x	x	x	x			
Binomial		x	x	x	x			
Ordinal		x	x	x	x			
Nominal		x						
Gamma	x					x		
Inverse Gaussian	x					x		
Negative binomial	x							
Zero-inflated Poisson								x
Zero-inflated negative binomial								x
Normal							x	
Survival analysis*		x	x	x	x			

* Survival analysis, like the ordinal and nominal models, is part of the multinomial family. In the current program, we have opted to let the user specifically select a survival model during analysis.

4.3 The Bernoulli distribution

The Bernoulli distribution is a discrete distribution. Variables that have a Bernoulli distribution can take one of two values. An example of a variable with a Bernoulli distribution is a coin toss, where the outcome is either heads (success) or tails (failure). The probability of a success is *p*, where $0 < p < 1$.

Sampling distribution

$$f(y_{ijk}) = \mu_{ijk}^{y_{ijk}}(1-\mu_{ijk})^{1-y_{ijk}}$$

Variance

$$\sigma^2(y_{ijk}) = \mu_{ijk}(1-\mu_{ijk})$$

Log-likelihood function

$$l_{ijk} = \prod p^{y_{ijk}} (1 - p_{ijk})^{1-y_{ijk}}$$

Four link functions are available for use with the Bernoulli distribution.

4.3.1 Logit link

$$\eta_{ijk} = \text{logit}(\mu_{ijk}) = \ln\left\{\frac{\mu_{ijk}}{1-\mu_{ijk}}\right\}$$

Model for the means

$$\mu_{ijk} = \frac{1}{1+\exp\left(-\eta_{ijk}\right)}$$

Notes

If the model is appropriate for the data, it is assumed that the residual variance $\sigma^2 = \pi^2 / 3$.

4.3.2 Complimentary log-log link

$$\eta_{ijk} = \ln(-\ln(1-\mu_{ijk}))$$

Model for the means

$$\mu_{ijk} = 1 - \exp\{-\exp(\eta_{ijk})\}$$

Notes

If the model is appropriate for the data, it is assumed that the residual variance $\sigma^2 = \pi^2 / 6$.

4.3.3 Probit link

$$\eta_{ijk} = \Phi^{-1}(\mu_{ijk})$$

Model for the means

$$\mu_{ijk} = \Phi(\eta_{ijk})$$

Notes

If the model is appropriate for the data, it is assumed that the residual variance $\sigma^2 = 1$.

4.3.4 Log-log link

$$\eta_{ijk} = -\ln\{-\ln(\mu_{ijk})\}$$

Model for the means

$$\mu_{ijk} = \exp\{-\exp(\eta_{ijk})\}$$

Notes

If the model is appropriate for the data, it is assumed that the residual variance $\sigma^2 = \pi^2 / 6$.

4.4 The Binomial distribution

The Binomial distribution is a discrete distribution in which the outcome is binary. While the Bernoulli distribution is used to describe the outcome of a single trial of an event, the Binomial distribution is used when the outcome of an event is observed multiple times.

Sampling distribution

Let y_{ijk} denote the proportion of successes in n_{ijk} independent trials:

$$f(y_{ijk}) = \begin{pmatrix} n_{ijk} \\ n_{ijk} y_{ijk} \end{pmatrix} \mu_{ijk}^{n_{ijk} y_{ijk}} \left(1 - \mu_{ijk}\right)^{n_{ijk}\left(1 - y_{ijk}\right)}$$

Variance

$$\sigma^2(y_{ijk}) = \frac{\mu_{ijk}\left(1-\mu_{ijk}\right)}{n_{ijk}}$$

Notes

The binomial trials parameter is added to the model by regarding it as weight so that the model for the means becomes $\mu_{ij} * trials$.

Log-likelihood function

$$l_{ijk} = \left[n_{ijk} y_{ijk} \log(p_{ijk}) + n_{ijk}(1-y_{ijk})\log(1-p_{ijk})\right]$$

Four link functions are available for use with the Bernoulli distribution.

4.4.1 Logit link

$$\eta_{ijk} = \text{logit}(\mu_{ijk}) = \ln\left\{\frac{\mu_{ijk}}{1-\mu_{ijk}}\right\}$$

Model for the means

$$\mu_{ijk} = \frac{1}{1+\exp\left(-\eta_{ijk}\right)}$$

Notes

If the model is appropriate for the data, it is assumed that the residual variance $\sigma^2 = \pi^2 / 3$.

4.4.2 Complimentary log-log link

$$\eta_{ijk} = \ln(-\ln(1-\mu_{ijk}))$$

Model for the means

$$\mu_{ijk} = 1 - \exp\{-\exp(\eta_{ijk})\}$$

Notes

If the model is appropriate for the data, it is assumed that the residual variance $\sigma^2 = \pi^2 / 6$.

4.4.3 Probit link

$$\eta_{ijk} = \Phi^{-1}(\mu_{ijk})$$

Model for the means

$$\mu_{ijk} = \Phi(\eta_{ijk})$$

Notes

If the model is appropriate for the data, it is assumed that the residual variance $\sigma^2 = 1$.

4.4.4 Log-log link

$$\eta_{ijk} = -\ln\{-\ln(\mu_{ijk})\}$$

Model for the means

$$\mu_{ijk} = \exp\{-\exp(\eta_{ijk})\}$$

Notes

If the model is appropriate for the data, it is assumed that the residual variance $\sigma^2 = \pi^2 / 6$.

4.5 The Gamma distribution

The gamma distribution is a two-parameter continuous probability distribution. It occurs when the waiting times between Poisson distributed events are relevant.

Sampling distribution

$$f(y_{ijk}) = \frac{1}{\Gamma\left(\frac{1}{\psi}\right) y_{ijk}} \left(\frac{y_{ijk}}{\mu_{ijk}\psi}\right)^{\frac{1}{\psi}} \exp\left(-\frac{y_{ijk}}{\mu_{ijk}\psi}\right)$$

Variance

$$\sigma^2\left(y_{ijk}\right) = \psi\mu_{ijk}^2$$

For the Gamma model, the scale parameter is represented by $1/\psi$ in the distribution function above. The dispersion parameter $\phi = \psi$. It should be noted that the meaning of the scale parameter in the case of the Gamma distribution is different from those for the other distributions discussed in this document. For the normal, inverse Gaussian, Binomial and Poisson distribution the scale parameter is equal to $\sqrt{\phi}$. When a weight w_{ijk} is present, ϕ is replaced by ϕ / w_{ijk} .

Log-likelihood function

$$l_{ijk} = \frac{w_{ijk}}{\phi} \log\left(\frac{w_{ijk} y_{ijk}}{\mu_{ijk}^2 \phi}\right) - \frac{w_{ijk} y_{ijk}}{\mu_{ijk}\phi} - \log\left(y_{ijk}\right) - \log\left(\Gamma\left(\frac{w_{ijk}}{\phi}\right)\right)$$

Two link functions are available for the Gamma distribution.

4.5.1 Log link

$$\eta_{ijk} = \log[\Phi(\mu_{ijk})]$$

Model for the means

$$\mu_{ijk} = \exp[\Phi(\eta_{ijk})]$$

where $\Phi(\cdot)$ denotes the cumulative function of the Gamma distribution.

Notes

If the model is appropriate for the data, it is assumed that the residual variance $\sigma^2\left(y_{ijk}\right)=\psi\mu_{ijk}^2$. A description of the estimation of σ^2 is given elsewhere.

4.5.2 Power link

$$\eta_{ijk}{}^{\alpha}=\boldsymbol{\Phi}^{-1}(\mu_{ijk})$$

Model for the means

$$\mu_{ijk}=\boldsymbol{\Phi}(\eta_{ijk})^{\alpha}$$

where $\boldsymbol{\Phi}(\cdot)$ denotes the cumulative function of the Gamma distribution.

4.6 The inverse Gaussian distribution

The inverse Gaussian distribution is a two-parameter family of continuous probability distributions, first studied in relation to Brownian motion. This distribution is one of a family of distributions that have been called the Tweedie distributions, named after M.C.K. Tweedie who first used the name Inverse Gaussian as there is an inverse relationship between the time to cover a unit distance and distance covered in unit time.

Sampling distribution

$$f(y_{ijk})=\frac{1}{\sqrt{2\pi y^3{}_{ijk}\psi}}\exp\left(-\frac{1}{2y_{ijk}}\left(\frac{y_{ijk}-\mu_{ijk}}{\mu_{ijk}\psi}\right)^2/\psi\right)$$

Variance

$$\sigma^2\left(y_{ijk}\right)=\psi\mu_{ijk}^3$$

For the inverse Gaussian model, the dispersion parameter $\phi=\psi^2$ and the scale parameter is equal to ψ. When a weight w_{ijk} is present, ϕ is replaced by ϕ/w_{ijk} .

Log-likelihood function

$$l_{ijk} = -\frac{1}{2}\left[\frac{w_{ijk}(y_{ijk} - \mu_{ijk})^2}{y_{ijk}\mu_{ijk}^2\phi} + \log\left(\frac{\phi y_{ijk}^3}{w_{ijk}}\right) + \log(2\pi)\right]$$

Two link functions are available for the inverse Gaussian distribution.

4.6.1 Log link

$$\eta_{ijk} = \log[\boldsymbol{\Phi}(\mu_{ijk})]$$

Model for the means

$$\mu_{ijk} = \exp[\boldsymbol{\Phi}(\eta_{ijk})]$$

where $\boldsymbol{\Phi}(\cdot)$ denotes the cumulative function of the inverse Gaussian distribution.

Notes

If the model is appropriate for the data, it is assumed that the residual variance $\sigma^2\left(y_{ijk}\right) = \psi\mu_{ijk}^3$. A description of the estimation of σ^2 is given here.

4.6.2 Power link

$$\eta_{ijk}{}^{\alpha} = \boldsymbol{\Phi}^{-1}(\mu_{ijk})$$

Model for the means

$$\mu_{ijk} = \boldsymbol{\Phi}(\eta_{ijk})^{\alpha}$$

where $\boldsymbol{\Phi}(\cdot)$ denotes the cumulative function of the inverse Gaussian distribution.

4.7 The Negative Binomial distribution

The negative binomial distribution is a discrete probability distribution. It is used to model the number of successes in a sequence of independent and identically distributed Bernoulli trials before a specified, nor random, number of failures occurs. The negative binomial model is an extension of the Poisson model, in the sense that it adds a normally distributed overdispersion effect.

Sampling distribution

$$f(y_{ijk}) = \frac{\Gamma\left(y_{ijk} + \frac{1}{\psi}\right)}{\Gamma\left(y_{ijk}+1\right)\Gamma\left(\frac{1}{\psi}\right)} \frac{\left(\psi\mu_{ijk}\right)^{y_{ijk}}}{\left(1+\psi\mu_{ijk}\right)^{y_{ijk}+\frac{1}{\psi}}}$$

Variance

$$\sigma^2\left(y_{ijk}\right) = \mu_{ijk} + \psi\mu_{ijk}^2$$

For the negative binomial distribution, the dispersion is equal to ψ. Note that this parameter, the so-called negative binomial dispersion parameter, is different from the dispersion parameter for the other distributions but is an additional distribution parameter. As such, it must either be estimated or set to a fixed value. When a weight w_{ijk} is present, ψ is replaced by ψ / w_{ijk} .

Log-likelihood function

$$l_{ijk} = y_{ijk}\log(\psi\mu_{ijk}) - (y_{ijk} + 1/\psi)\log(1+\psi\mu_{ijk}) + \log\left(\frac{\Gamma(y_{ijk}+1/\psi)}{\Gamma(y_{ijk}+1)\Gamma(1/\psi)}\right)$$

For this model, the log link function is used.

4.7.1 Log link

$$\eta_{ijk} = \log[\Phi(\mu_{ijk})]$$

Model for the means

$$\mu_{ijk} = \exp[\Phi(\eta_{ijk})]$$

where $\Phi(\cdot)$ denotes the cumulative function of the negative Binomial distribution.

Notes

- For the negative binomial, the option to specify an offset or overdispersion parameter is available.
- If the model is appropriate for the data, it is assumed that the residual variance $\sigma^2 = \mu + \psi\mu^2$, where ψ denotes the dispersion parameter.

4.8 The Nominal model

The nominal model is part of a family of models based on the multinomial distribution. The multinomial distribution is a generalization of the Binomial distribution. It is commonly used in to describe the probability of the outcome of *n* independent trials each of which leads to a success for one of *c* categories, with each category having a given fixed probability of success. A nominal variable has categories that cannot be ordered.

Sampling distribution

$$f\left(y_{ijk,1}, y_{ijk,2}, \ldots, y_{ijk,C-1}\right) = \frac{n_{ijk}!}{\left(\prod_{l=1}^{C-1} y_{ijk,l}!\right)\left(n_{ijk} - \sum_{l=1}^{C-1} y_{ijk,l}\right)!}\left(\prod_{l=1}^{C-1} \mu_{ijk,l}^{y_{ijk,l}}\right)\mu_{ijk,C}^{n_{ijk} - \sum_{l=1}^{C-1} y_{ijk,l}}$$

where *C* represents the number of categories of the outcome variable. The nominal model uses the generalized logistic form of the logit link functions: category 1 vs category 4 , … category 3 vs category 4, where 4 is the last category.

Covariance matrix

$$\mathbf{\Sigma}(\mathbf{y}^*_{ijk}) = \mathbf{D}_{\mu_{ijk}} - \boldsymbol{\mu}_{ijk}\boldsymbol{\mu}'_{ijk}$$

where $\mathbf{y}_{ijk}^{*} = \begin{bmatrix} y_{ijk,1} & y_{ijk,2} \cdots & y_{ijk,C-1} \end{bmatrix}'$ and $\mathbf{D}_{\mu_{ijk}}$ denotes a $(C-1)\times(C-1)$ diagonal matrix with the elements of $\boldsymbol{\mu}_{ijk} = \begin{bmatrix} \mu_{ijk,1} & \mu_{ijk,2} \cdots & \mu_{ijk,C-1} \end{bmatrix}'$ on the diagonal.

Log-likelihood function

$$l_{ijk} = \sum_{jk} \log\left(\mu_{ijk}\right)$$

For the nominal model, the logistic link function is specified.

4.8.1 Logit link

$$\eta_{ijk,l} = \operatorname{logit}(\mu_{ijk,l}) = \ln\left\{\frac{\mu_{ijk,l}}{\mu_{ijk,C}}\right\} \quad \forall \quad l = 1, 2, \ldots, C-1$$

Model for means

$$\mu_{ijk,l}^{*} = \sum_{r=1}^{l} \mu_{ijk,r} = \frac{\exp\left\{\tau_l - \eta_{ijk,l}\right\}}{1+\exp\left\{\tau_l - \eta_{ijk,l}\right\}} \quad \forall \quad l = \ldots, C-1$$
$$= \frac{1}{1+\exp\left\{-\eta_{ijk}^{*}\right\}}$$

Notes

If the model is appropriate for the data, it is assumed that the residual variance $\sigma^2 = \pi^2/3$.

4.9 The Normal distribution with identity link

Sampling distribution

$$f\left(y_{ijk}\right) = \frac{1}{\sqrt{2\pi\psi^2}} \exp\left(-\frac{1}{2}\left(\frac{y_{ijk}-\mu_{ijk}}{\psi}\right)^2\right)$$

Variance

$$\sigma^2\left(y_{ijk}\right)=\psi^2$$

The scale parameter is equal to ψ and the dispersion parameter $\phi=\psi^2$. When a weight w_{ijk} is present, ϕ is replaced by ϕ/w_{ijk}.

Log-likelihood function

$$l_{ijk}=-\frac{1}{2}\left[\frac{w_{ijk}(y_{ijk}-\mu_{ijk})^2}{\phi}+\log\left(\frac{\phi}{w_{ijk}}\right)+\log(2\pi)\right]$$

For the Normal model, the identity link function is used.

4.9.1 Identity link

$$\eta_{ijk}=\mu_{ijk}$$

Model for the means

$$\mu_{ijk}=\Phi(\eta_{ijk})$$

where $\Phi(\cdot)$ denotes the cumulative function of the standard normal distribution.

Notes

If the model is appropriate for the data, results should correspond closely with those obtained running a Normal (HLM) model. If results differ, the Gamma and Inverse Gaussian distributions offer alternative models that may be better suited.

4.10 The Ordinal model

The ordinal model is also part of a family of models based on the multinomial distribution. The multinomial distribution is a generalization of the Binomial distribution. It is commonly used in to describe the probability of the outcome of *n* independent trials each of which leads to a success for one of c categories, with each category having a given fixed probability of success. An ordinal outcome is an outcome whose levels can be ordered.

Sampling distribution

$$f\left(y_{ijk,1}, y_{ijk,2}, \ldots, y_{ijk,C-1}\right) = \frac{n_{ijk}\,!}{\left(\prod_{l=1}^{C-1} y_{ijk,l}\,!\right)\left(n_{ijk} - \sum_{l=1}^{C-1} y_{ijk,l}\right)!}\left(\prod_{l=1}^{C-1} \mu_{ijk,l}^{y_{ijk,l}}\right)\mu_{ijk,C}^{\,n_{ijk}-\sum_{l=1}^{C-1} y_{ijk,l}}$$

where C represents the number of categories of the outcome variable.

Covariance matrix

$$\boldsymbol{\Sigma}(\mathbf{y}_{ijk}^{*}) = \mathbf{D}_{\mu_{ijk}} - \boldsymbol{\mu}_{ijk}\boldsymbol{\mu}_{ijk}'$$

where $\mathbf{y}_{ijk}^{*} = \left[y_{ijk,1} \;\; y_{ijk,2} \cdots \;\; y_{ijk,C-1}\right]'$ and $\mathbf{D}_{\mu_{ijk}}$ denotes a $(C-1)\times(C-1)$ diagonal matrix with the elements of $\boldsymbol{\mu}_{ijk} = \left[\mu_{ijk,1} \;\; \mu_{ijk,2} \cdots \;\; \mu_{ijk,C-1}\right]'$ on the diagonal.

Log-likelihood function

$$l_{ijk} = \sum_{jk} \log\left(\mu_{ijk}\right)$$

The ordinal model differs from the nominal model in that cumulative link functions are used in the ordinal model, in contrast with the nominal model where a generalized logistic model is used.

For the ordinal model four link functions are available.

4.10.1 Cumulative logit link

$$\eta_{ijk,l}^{*} = \text{clogit}(\mu_{ijk,l}^{*}) = \ln\left\{\frac{\mu_{ijk,l}^{*}}{1-\mu_{ijk,l}^{*}}\right\} \quad \forall \quad l = 1, 2, \ldots, C-1$$

Model for means

$$\mu_{ijk,l} = \frac{\exp\left\{\eta_{ijk,l}\right\}}{1+\sum_{l=1}^{C-1}\exp\left\{\eta_{ijk,l}\right\}} \quad \forall \quad l = 1, 2, \ldots, C-1$$

where

$$\eta^{*}_{ijk,l} = \tau_l - \eta_{ijk,l}$$

the elements of $\tau_1, \quad \tau_{2,} \quad \cdots \quad , \tau_{C-1}$ denote threshold parameters.

Notes

If the model is appropriate for the data, it is assumed that the residual variance $\sigma^2 = \pi^2 / 3$.

4.10.2 Cumulative complimentary log-log link

$$\eta^{*}_{ijk,l} = \text{cloglog}(\mu^{*}_{ijk,l}) = \ln\left(-\ln\left(1-\mu^{*}_{ijk,l}\right)\right)$$

Model for means

$$\mu^{*}_{ijk,l} = \sum_{r=1}^{l}\mu_{ijk,r} = 1-\exp\left(-\exp\left\{\eta^{*}_{ijk,l}\right\}\right) \quad \forall \quad l = 1, 2, \ldots, C-1$$

where C represents the number of categories of the outcome variable.

Notes

If the model is appropriate for the data, it is assumed that the residual variance $\sigma^2 = \pi^2 / 6$.

4.10.3 Cumulative probit link

$$\eta^*_{ijk,l} = \Phi^{-1}\left(\mu^*_{ijk,l}\right)$$

Model for means

$$\mu^*_{ijk,l} = \sum_{r=1}^{l} \mu_{ijk,r} = \Phi\left(\eta^*_{ijk,l}\right) \quad \forall \quad l = 1, 2, \ldots, C-1$$

where $\Phi\left(\cdot\right)$ denotes the cumulative distribution function of the standard normal distribution and C represents the number of categories of the outcome variable.

Notes

If the model is appropriate for the data, it is assumed that the residual variance $\sigma^2 = 1$.

4.10.4 Cumulative log-log link

$$\eta^*_{ijk,l} = \text{loglog}(\mu^*_{ijk,l}) = -\ln\left(-\ln\left(\mu^*_{ijk,l}\right)\right)$$

Model for means

$$\mu^*_{ijk,l} = \sum_{r=1}^{l} \mu_{ijk,r} = \exp\left(-\exp\left\{-\eta^*_{ijk,l}\right\}\right) \quad \forall \quad l = 1, 2, \cdots, C-1$$

where C represents the number of categories of the outcome variable.

Notes

If the model is appropriate for the data, it is assumed that the residual variance $\sigma^2 = \pi^2 / 6$.

4.11 The Poisson distribution

The Poisson distribution is a discrete frequency distribution that gives the probability of several independent events occurring in a fixed time, given the average number of times the event occurs over that time period.

Sampling distribution

Assume y_{ijk} follows a Poisson distribution with mean μ_{ijk}. In other words, the probability density function of y_{ijk} is given by

$$f\left(y_{ijk},\mu_{ijk}\right)=\frac{e^{-\mu_{ijk}}\mu_{ijk}^{y_{ijk}}}{y_{ijk}!}\Rightarrow \ln f\left(y_{ijk},\mu_{ijk}\right)=y_{ijk}\ln\left\{\mu_{ijk}\right\}-\mu_{ijk}-\ln\left\{y_{ijk}!\right\}$$

Variance

$$\sigma^2\left(y_{ijk}\right)=\mu_{ijk}$$

In the case of the Poisson distribution, $\phi=1$. The offset parameter is estimated by replacing the model for the means by $\mu_{ij}*\phi$, where ϕ is the offset variable.

Log-likelihood function

$$l_{ijk}=y_{ijk}\log(\mu_{ijk})-\mu_{ijk}$$

For the Poisson distribution, the model is transformed to a linear model by using the log link function.

4.11.1 Log link

$$\eta_{ijk}=\log\boldsymbol{\Phi}^{-1}(\mu_{ijk})$$

Model for the means

$$\mu_{ijk}=\log\boldsymbol{\Phi}(\eta_{ijk})$$

where $\mathbf{\Phi}(\cdot)$ denotes the cumulative function of the Poisson distribution.

Notes

- For the Poisson model, the option to specify an offset or overdispersion parameter is available.
- If the model is appropriate for the data, it is assumed that the residual variance is equal to the mean, i.e., it is assumed that $\sigma^2 = \mu_{ijk}$. A description of the estimation of σ^2 is given here.

4.12 The Survival analysis model

The survival analysis model is used to describe the expected duration of time until one or more events occur. Observations are censored, in that for some units the event of interest did not occur during the entire time period studied. In addition, there may be predictors whose effects on the waiting time need to be controlled or assessed.

Sampling distribution

$$f\left(y_{ijk,1}, y_{ijk,2}, \ldots, y_{ijk,C-1}\right) = \frac{n_{ijk}\,!}{\left(\prod_{l=1}^{C-1} y_{ijk,l}\,!\right)\left(n_{ijk} - \sum_{l=1}^{C-1} y_{ijk,l}\right)!}\left(\prod_{l=1}^{C-1} \mu_{ijk,l}^{y_{ijk,l}}\right)\mu_{ijk,C}^{\,n_{ijk} - \sum_{l=1}^{C-1} y_{ijk,l}}$$

where C stands for the number of categories of the outcome variable.

Covariance matrix

$$\mathbf{\Sigma}(\mathbf{y}_{ijk}^{*}) = \mathbf{D}_{\mu_{ijk}} - \boldsymbol{\mu}_{ijk}\boldsymbol{\mu}_{ijk}'$$

where $\mathbf{y}_{ijk}^{*} = \left[y_{ijk,1} \quad y_{ijk,2} \cdots \quad y_{ijk,C-1}\right]'$ and $\mathbf{D}_{\mu_{ijk}}$ denotes a $(C-1)\times(C-1)$ diagonal matrix with the elements of $\boldsymbol{\mu}_{ijk} = \left[\mu_{ijk,1} \quad \mu_{ijk,2} \cdots \quad \mu_{ijk,C-1}\right]'$ on the diagonal.

Log-likelihood function

$$l_{ijk} = \sum_{jk} \log\left(\mu_{ijk}\right)$$

Under the survival analysis model, the number of thresholds estimated are equal to the number of categories. For the other models from the multinomial distribution, the probability for the last category could be calculated as $1-\sum_{l=1}^{C-1} p(y_{ijk}=l)$ which does not apply in the survival analysis model where we are no longer modeling probabilities adding up to 1.

Let y_{ijk} denote the outcome c , where c = 1, 2, … ,C for individual *ijk*. For each level-1 unit, observation continues until time y_{ijk} , at which point either an event occurs, indicated by $d_{ijk}=1$ or the observation is censored, indicated by $d_{ijk}=0$. Censoring indicates being observed at c but not at $c+1$.

The probability of failure up to and including time interval c can be expressed as

$$P_{ijk,c} = P(y_{ijk} \le c) \quad .$$

The probability of survival beyond time interval c is then $1-P_{ijk,c}$.

Although the logit, probit and log-log function can be used, it is standard practice to use the complimentary log-log link described below.

4.12.1 Complimentary log-log link (proportional hazards model)

$$\begin{aligned}\eta^*_{ijk,c} &= \text{cloglog}(\mu^*_{ijk,c}) = \ln\left(-\ln\left(1-\mu^*_{ijk,c}\right)\right) \\ &= \gamma_c + \eta_{ijk,c}\end{aligned}$$

where the threshold terms γ_c represent the natural logarithm of the integrated baseline hazard and $\eta_{ijk,c} = \mathbf{x}'\boldsymbol{\beta}$, with c = 1, 2, …, C.

Model for means

$$\mu^{*}_{ijk,c} = 1 - \exp\left(-\exp\left\{\eta^{*}_{ijk,c}\right\}\right)$$

where C stands for the number of categories of the outcome variable.

Notes

For the survival analysis model, the complimentary log-log link function is usually specified. However, the logit, probit and log-log link functions are also available. Assumptions regarding the residual variance are the same for each link function as for the ordinal model.

4.13 The Zero-Inflated Poisson model

The zero-inflated Poisson model is a mixture model used to model count data that has an excess of zero counts. It is assumed that for non-zero counts the counts are generated according to a Poisson model.

Let π_{ijk} denotes the probability of recording a zero count.

Sampling distribution

$$\begin{aligned}\Pr(y_{ijk} = n) &= (1-\pi_{ijk})\frac{\mu_{ijk(1)}^{y_{ijk}}\exp(-\mu_{ijk(1)})}{y_{ijk}\,!} \quad \text{when} \quad n > 0 \\ &= \pi_{ijk} + (1-\pi_{ijk})\exp(-\mu_{ijk(2)}) \quad \text{when} \quad n = 0\end{aligned}$$

where $\mu_{ijk(1)}$ is the mean for the first distribution, and $\mu_{ijk(2)}$ the mean of the second distribution in the mixture model.

Variance

$$\sigma^2_{(1)}\left(y_{ijk}\right)=\mu_{ijk(1)}, \quad n>0$$

$$\sigma^2_{(2)}\left(y_{ijk}\right)=(1-\pi_{ijk})\mu_{ijk(2)}(1+\mu_{ijk(2)}\pi_{ijk}), \quad n=0$$

For the zero-inflated Poisson model, the logit link function must be specified. For the log (Poisson) model, modelling the non-zero responses, the link function is given by

$$\eta_{ijk(1)}=\log\Phi^{-1}(\mu_{ijk(1)})$$

For the zero-inflated part of the model, modeling the zero responses, the logit link function is used:

$$\eta_{ijk(2)}=\operatorname{logit}(\mu_{ijk(2)})=\ln\left\{\frac{\mu_{ijk(2)}}{1-\mu_{ijk(2)}}\right\}$$

For the combined models

$$\eta_{ijk}=\eta_{ijk(1)}+\eta_{ijk(2)}=\log\Phi^{-1}(\mu_{ijk(1)})+\operatorname{logit}(\mu_{ijk(2)})$$

Model for the means

$$\mu_{ijk(1)}=\log\Phi(\eta_{ijk(1)})$$

where Φ denotes the cumulative function of the Poisson distribution, and

$$\mu_{ijk(2)}=\frac{1}{1+\exp\left(-\eta_{ijk(2)}\right)}$$

Likelihood function:

$$l\left(\pi_{ijk},\lambda\right)=n_{ijk}r_0\ln\left(\pi_{ijk}+(1-\pi_{ijk}\right)\exp(-\lambda)+n_{ijk}(1-r_0)(\ln\left(1-\pi_{ijk}\right)-\lambda)+n_{ijk}\,\bar{y}_{ijk}\ln\lambda$$

where r_0 represents the proportion of observed zeros in the data.

Notes

- In this model, the definition of the means differs for the two cases, as the means for the two distributions may be modelled using different predictors. For $j = 0$, the predictors used are known as “zero-inflated” predictors. For the zero-inflated Poisson, the option to specify an offset or overdispersion parameter is available.
- If the model is appropriate for the data, it is assumed that the residual variance of the zero-inflated component can be expressed as $\sigma^2 = (1 - p_{ijk})\mu_{ijk}(1 + p_{ijk}\mu_{ijk})$ where p_{ijk} denotes the inflation factor (i.e., the proportion of zeros to the total number of counts).
- For a sample of N observations, you may observe n_1 zero values, of which only n_2 $(n_2 < n_1)$ can be explained by the regular Poisson distribution. In that case, $n_1 - n_2$ zero values are considered excess. The inflation factor p_{ijk} is defined as $\frac{n_1 - n_2}{N}$. In the current program, the approximation $\frac{n_1}{N}$ is employed as estimate of the inflation factor.
- A description of the estimation of $\sigma^2_{(1)}$ is given here.

4.14 The Zero-Inflated negative Binomial model

The zero-inflated negative Binomial model is a mixture model used to model count data that has an excess of zero counts. It is assumed that the count in the not-always-zero group has a negative binomial distribution.

Sampling distribution

$$f(y_{ijk}) = \frac{\Gamma(y_{ijk} + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_{ijk} + 1)} \left(\frac{1}{1 + \alpha\mu_{ijk}} \right)^{\alpha^{-1}} \left(\frac{\alpha\mu_{ijk}}{1 + \alpha\mu_{ijk}} \right)^{y_i}$$

$$\begin{aligned} \Pr(y_{ijk} = n) &= \pi_{ijk} + (1 - \pi_{ijk})(1 + \alpha\mu_{ijk(2)})^{-1/\alpha} \quad \text{when} \quad n = 0 \\ &= (1 - \pi_{ijk}) \frac{\Gamma(y_{ijk} + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_{ijk} + 1)} \left(\frac{1}{1 + \alpha\mu_{ijk(1)}} \right)^{\alpha^{-1}} \left(\frac{\alpha\mu_{ijk(1)}}{1 + \alpha\mu_{ijk(1)}} \right)^{y_i} \quad n > 0 \end{aligned}$$

where $\mu_{ijk(1)}$ the mean for the first distribution modeling the non-zero counts, and $\mu_{ijk(2)}$ the mean of the second distribution (modeling the zero counts) in the mixture model. The parameter α^{-1} denotes the dispersion parameter and π_{ijk} the probability of recording a zero count.

Variance

$$\sigma^2_{(1)}\left(y_{ijk}\right) = \mu_{ijk(1)} + \alpha\mu^2_{ijk(1)}, \quad n > 0$$
$$\sigma^2_{(2)}\left(y_{ijk}\right) = \mu_{ijk(2)}, \quad n = 0$$

For the log (negative binomial) model, modelling the non-zero responses, the link function is given by

$$\eta_{ijk(1)} = \log\Phi^{-1}(\mu_{ijk(1)})$$

For the zero-inflated part of the model, modeling the zero responses, the logit link function is used:

$$\eta_{ijk(2)} = \text{logit}(\mu_{ijk(2)}) = \ln\left\{\frac{\mu_{ijk(2)}}{1-\mu_{ijk(2)}}\right\}$$

For the combined models

$$\eta_{ijk} = \eta_{ijk(1)} + \eta_{ijk(2)} = \log\Phi^{-1}(\mu_{ijk(1)}) + \text{logit}(\mu_{ijk(2)})$$

Variance

$$\ln(l) = \sum_{i:y_i=0} \ln\{\lambda_i + (1+\alpha\mu_i)^{-\alpha^{-1}}\} + \sum_{i:y_i=0}\sum_{j=0}^{y_i-1} \ln\{j+\alpha^{-1}\} +$$
$$\sum_{i:y_i>0}\{-\ln(y_i!) - (y_i+\alpha^{-1})\ln(1+\alpha\mu_i) + y_i\ln(\alpha) + y_i\ln(\mu_i)\} + \sum_{i=1}^{n}\ln\{1+\lambda_i\}$$

Notes

- In this model, the definition of the means differs for the two cases, as the means for the two distributions may be modelled using different predictors. For $j = 0$, the predictors used are known as “zero-inflated” predictors.
- For the zero-inflated negative binomial, the option to specify an offset or overdispersion parameter is available.
- If the model is appropriate for the data, the assumption is made that the residual variance $\sigma^2 = (1-p_{ijk})\mu_{ijk}(1+\mu_{ijk}(p_{ijk} + dispersion))$ where p_{ijk} denotes the inflation factor (i.e., the proportion of zeros to the total number of counts). The inflation factor is approximated

by $\frac{n_1}{N}$, where n_1 denotes the number of zero observations and N the total number of observations. See the notes for the zero-inflated Poisson model for more details.

- A description of the estimation of $\sigma_{(1)}^2$ is given here.

5 Methods of estimation

Models with normally distributed outcomes are estimated by full Maximum Likelihood. For starting values, the solution obtained when all random effects are set to identity is used.

For models with binary, ordinal, count, and nominal outcomes, or when a normal distribution with an identity link function is specified, two methods of estimation are available: maximization of the posterior distribution (MAP) and numerical integration (adaptive and non-adaptive quadrature) to obtain parameter and standard error estimates.

The MAP method of estimation can be used to obtain a point estimate of an unobserved quantity based on empirical data. It is closely related to Fisher's method of maximum likelihood (ML) but employs an augmented optimization objective which incorporates a prior distribution over the quantity one wants to estimate.

The Maximum Likelihood estimate, $\hat{\psi}_{ML}$, of ψ is obtained by using a three-stage estimation procedure. In stage one, the Fisher Scoring algorithm is used to obtain Maximum Likelihood estimates of the elements of $\boldsymbol{\theta}$. These estimates are then used as fixed values for the elements of $\boldsymbol{\theta}$ in a Newton-Raphson algorithm to obtain an estimate, $\hat{\psi}_{ML}$, of the dispersion parameter ψ. In this algorithm, the method of moments estimates of ψ is used as the starting values for $\hat{\psi}_{ML}$. In stage three of the procedure, the Fisher scoring algorithm is extended to include the dispersion parameter to yield Maximum Likelihood estimates of the dispersion parameter and the elements of $\boldsymbol{\theta}$.

This three-stage procedure is used in the case of the Negative Binomial, Inverse Gaussian, and Gamma sampling distributions. In the case of the Gamma and Inverse Gaussian sampling distributions, the Maximum Likelihood estimate of the scale parameter, $\hat{\phi}_{ML}$, is computed from $\hat{\psi}_{ML}$ and the Delta method (Bishop, Feinberg & Holland, 1988) is used to compute the corresponding standard error estimate. In the case of the Normal sampling distribution, $\hat{\psi}_{ML}$ is computed as

$$\hat{\psi}_{ML} = \frac{\sum_{h=1}^{H}\sum_{i=1}^{n_h}\sum_{j=1}^{n_{h_i}} w_{h_{ij}} f_{h_{ij}} \left(y_{h_{ij}} - \hat{\mu}_{h_{ij}}\right)^2}{\sum_{h=1}^{H}\sum_{i=1}^{n_h}\sum_{j=1}^{n_{h_i}} w_{h_{ij}} f_{h_{ij}}}$$

The standard error estimates are multiplied with the scale parameter estimate to correct them with respect to scale.

Adaptive quadrature estimation is a numeric method for evaluating multi-dimensional integrals. For mixed effect models with count and categorical outcomes, the log-likelihood function is expressed as the sum of the logarithm of integrals, where the summation is over higher-level units, and the dimensionality of the integrals equals the number of random effects. Typically, MAP estimates are used as starting values for the quadrature procedure. When the number of random effects is large, the quadrature procedures can become computationally intensive. In such cases, MAP estimation is usually selected as the final method of estimation.

Numerical quadrature, as implemented here, offers users a choice between adaptive and non-adaptive quadrature. Quadrature uses a quadrature rule, i.e., an approximation of the definite integral of a function, usually stated as a weighted sum of function values at specified points within the domain of integration. Adaptive quadrature generally requires fewer points and weights to yield estimates of the model parameters and standard errors that are as accurate as would be obtained with more points and weights in non-adaptive quadrature. The reason for that is that the adaptive quadrature procedure uses the empirical Bayes means and covariances, updated at each iteration to essentially shift and scale the quadrature locations of each higher-level unit to place them under the peak of the corresponding integral. The algorithm used is based on the maximization of the posterior distribution (MAP) with respect to the random effects.

6 Fit measures, dispersion, and scale parameters

In addition to the likelihood function value at convergence, several related statistical measures for assessing model adequacy are available.

Pearson or deviance scale parameters may be requested. The likelihood ratio test statistic is based on the ratio of observed and expected frequencies, while the Pearson χ^2 involves the squared difference between the observed and expected frequencies. Many statisticians think the likelihood ratio chi square test to be preferable to the conventional Pearson chi-square test for the simultaneous analysis of several overlapping associations in a multiple-classification table, because under certain conditions it has the property of additivity of effects.

The deviance χ^2 estimate

$$\hat{\phi}_D = \sqrt{\frac{\chi_D^2}{d}}$$

$$\chi_D^2 = 2\ln L(\mathbf{y} \mid \mathbf{y}) - 2\ln L(\hat{\boldsymbol{\mu}} \mid \mathbf{y})$$

$$d = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} w_{ijk} - q$$

The Pearson χ^2 estimate

$$\hat{\phi}_P = \sqrt{\frac{\chi_P^2}{d}}$$

$$\chi_P^2 = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} \frac{w_{ijk}\left(y_{ijk} - \hat{\mu}_{ijk}\right)^2}{\hat{\sigma}^2\left(y_{ijk}\right)}$$

Estimates of these two deviances are available for the Binomial, Poisson, Gamma, Inverse Gaussian, ordinal, normal-identity, and zero-inflated Poisson models.

The program will calculate the following scaled Pearson and Likelihood Ratio factors:

- For Poisson, Negative Binomial and Zero-Inflated models, setting SCALE to PEARSON produces a scale factor equal to $\sqrt{Pearson\ \chi^2 / df}$. The scale factor thus obtained is also known as the overdispersion parameter. If set to DEVIANCE, the scale factor is equal to $\sqrt{Likelihood\ Ratio\ \chi^2 / df}$. When the overdispersion parameter is equal to 1, a Poisson model is appropriate; if it is larger than one, a negative binomial model is recommended.
- For the Normal distribution, SCALE to PEARSON produces a scale factor of $\sqrt{Pearson\ \chi^2 / df}$ and dispersion of $scale^2$. And for DEVIANCE $\sqrt{Likelihood\ Ratio\ \chi^2 / df}$ and dispersion $scale^2$.
- For the Gamma and Inverse Gaussian distributions, setting SCALE to PEARSON produces a scale factor equal to $\frac{1}{Pearson\ \chi^2 / df}$ and a dispersion of $Pearson\ \chi^2 / df$. If DEVIANCE is specified, these are equal to $\frac{1}{Likelihood\ Ratio\ \chi^2 / df}$ and $Likelihood\ Ratio\ \chi^2 / df$.
- For the Binomial model, setting SCALE to PEARSON produces a scale factor of $\sqrt{Pearson\ \chi^2 / df}$ and for DEVIANCE $\sqrt{Likelihood\ Ratio\ \chi^2 / df}$.

The expected value of a χ^2 distribution is equal to its degree of freedom, so a well-fitting model's scale factor should be close to 1.

Both the Akaike information criterion (AIC) and the Schwarz Bayesian criterion (SBC) are functions of the number of estimated parameters, and therefore "penalize" models with large numbers of parameters.

- The AIC was originally proposed for time-series models but is also used in regression. It is defined as $-2\ln L + 2r$, where r denotes the number of parameters estimated in the model. The model with minimum AIC, in a set of nested models, will be the most parsimonious according to this criterion.
- The SBC is defined as $-2\ln L + r\log n$, where n denotes the number of units at the highest level of the hierarchy. A smaller value of this criterion would indicate the most parsimonious of the models being compared.

For some of the sampling distributions discussed in the previous sections a dispersion parameter may also be estimated.

There is some confusion as to whether the additional parameter should be called a scale or a dispersion parameter. In the original GLIM software package released by the Royal Statistical Society, the parameter ϕ was defined using a SCALE command. In the original book by McCullagh and Nelder (1989), the parameters were called the dispersion parameter.

They defined the PDF as

$$f\left(y;\theta;\phi\right) \propto \exp\left[\frac{w}{a(\phi)}\left(y\theta - \kappa\left(\theta\right)\right)\right]$$

and assumed that

$$a(\phi) = \phi / \psi$$

with w a known prior weight. The variance of y was then expressed as

$$\operatorname{var}(y) = \frac{\phi}{w}V(\mu).$$

Chambers and Hastie (1991) also used this terminology. Over time, the term dispersion parameter has become the standard.

Although the estimate of the dispersion parameter is often used to indicate overdispersion or underdispersion, this estimate may also indicate other problems such as an incorrectly specified model or outliers in the data. You should carefully assess whether this type of model is appropriate for your data.

7 Intracluster correlation coefficients

The program automatically calculates intracluster coefficients when no more than one random effect is specified at higher level(s). For a two-level model, the intracluster correlation coefficient can be expressed as

$$ICC = \frac{Var(level-2)}{\left[Var(level-1)+Var(level-2)\right]}$$

For a three-level model it similarly follows as

$$ICC = \frac{Var(level-2)+Var(level-3)}{\left[Var(level-1)+Var(level-2)+Var(level-3)\right]}$$

In the case of the non-continuous outcomes, the residual variance at level-1 is no longer assumed to be σ^2. Instead, it is a function of link function and distribution. The table below provides the assumptions regarding and formulae used to calculate residual variance. The estimation of residual variance for some of the models is discussed here.

Distribution	**link**	**Assumed or estimated residual variance**
Bernoulli	Logit	$\pi^2/3$
	CLL	$\pi^2/6$
	Probit	1
	Log-log	$\pi^2/6$
Binomial	Logit	$\pi^2/3$
	CLL	$\pi^2/6$
	Probit	1
	Log-log	$\pi^2/6$
Gamma	Log	$\sigma^2 = \alpha\mu_{ijk}^2$ (Estimated)
Inverse Gaussian	Log	$\sigma^2 = \alpha\mu_{ijk}^3$ (Estimated)
Negative binomial	Log	$\sigma^2 = \mu + \psi\mu^2$ (Estimated)
Nominal	Logit	$\pi^2/3$
Normal	Identity	$\mathrm{var}(y_{ijk})$
Ordinal	Logit	$\pi^2/3$
	CLL	$\pi^2/6$

	Probit	1
	Log-log	$\pi^2/6$
Poisson	Log	μ_{ijk} (Estimated)
Survival	Logit	$\pi^2/3$
	CLL	$\pi^2/6$
	Probit	1
	Log-log	$\pi^2/6$
Zero-inflated Poisson	Log	μ_{ijk} (Estimated)
	Logit	$(1-p_{ijk})\mu_{ijk}(1+p_{ijk}\mu_{ijk})$
Zero-inflated neg. binomial	Log	μ_{ijk} (Estimated)
	Logit	$(1-p_{ijk})\mu_{ijk}(1+\mu_{ijk}(p_{ijk}+dispersion))$

9 References

Bishop, Y. M. M, Feinberg, S. E., & Holland, P. W. (1988). *Discrete multivariate analysis: Theory and practice.* Cambridge, MA: MIT Press.

Chambers, J. & Hastie, T. (1991). *Statistical Models in S*. Wadsworth & Brooks.

MNcCullagh, P. & & Nelder, J.A. (1989). *Generalized Linear Models*. Second edition, Chapman and Hall.