

Chapter 4

Generalized Linear Mixed Models for Non-normal Responses



4.1 Introduction

Generalized linear mixed models (GLMMs) have been recognized as one of the major methodological developments in recent years, which is evidenced by the increased use of such sophisticated statistical tools with broader applicability and flexibility. This family of models can be applied to a wide range of different data types (continuous, categorical (nominal or ordinal), percentages, and counts), and each is appropriate for a specific type of data. This modern methodology allows data to be described through a distribution of the exponential family that best fits the response variable. These complex models were not computationally possible up until recently when advances in statistical software have allowed users to apply GLMMs (Zuur et al. 2009; Stroup 2012; Zuur et al. 2013). Researchers in fields other than statistical science are also interested in modeling the structure of data. For example, in the social sciences there have been applications in the field of education when several tests are applied to students; in longitudinal personality studies when the occurrence of an emotion is repeatedly observed over time over a set of people; and in surveys to investigate the political preference of a population, among others.

Likewise, agriculture and life sciences are other major areas, where the measurement of response variables depart from the conventionally used classical methodology based on “normality” to model or describe the data set, i.e., data that generally fall within the nominal, ordinal or interval (continuous) scales of measurement. In a GLMM, the data response does not undergo any transformation, but, instead, the response is modeled as a function of the expected value through a linear relationship with the explanatory variables. GLMMs, a powerful tool, allow proper modeling of variations between groups and between space and time, leading to accuracy in the modeling of the observed data as well as in the estimation of variance components.

4.2 A Brief Description of Linear Mixed Models (LMMs)

Before addressing GLMMs, we present a brief overview of linear mixed models (LMMs). An LMM is a model whose response variable is normal and assumes: (1) that the relationship between the mean of the dependent variable (y) and fixed and random effects can be modeled as a linear function; (2) that the variance is not a function of the mean; and (3) that random effects follow a normal distribution.

The classic representation of an LMM in the matrix form, is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon} \quad (4.1)$$

where \mathbf{y} is the vector ($n \times 1$) of the response variable; \mathbf{X} is the design matrix ($n \times (p + 1)$) of fixed effects with rank k ; $\boldsymbol{\beta}$ is the vector of unknown parameters ($(p + 1) \times 1$); \mathbf{Z} is the design matrix ($n \times q$) of random effects; and \mathbf{b} is the vector of unknown parameters of random effects ($q \times 1$), assuming that the vector of random effects \mathbf{b} follows a normal distribution with mean $\mathbf{0}$ and variance matrix \mathbf{G} , that is, $\mathbf{b} \sim N(\mathbf{0}, \mathbf{G})$. Finally, $\boldsymbol{\varepsilon}$ is the error vector with a normal distribution with mean $\mathbf{0}$ and a variance–covariance matrix ($\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{R})$); both vectors \mathbf{b} and $\boldsymbol{\varepsilon}$ are assumed to be independent of each other.

Model 4.1, as previously mentioned, can be described in terms of a probability distribution in two ways: the first is the marginal model $\mathbf{y} \sim N(E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}, \text{Var}[\mathbf{y}] = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R})$, where the mean is based solely on the fixed effects, and the parameters describing the random effects are contained in the variance and covariance matrix \mathbf{V} (Littell et al. 2006), while the second form is the conditional model $\mathbf{y} \mid \mathbf{b} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \mathbf{R})$. Under normality assumptions, both models are exactly the same and hence produce the same solution, whereas when normality is not satisfied, the models produce different solutions (Stroup 2012).

4.3 Generalized Linear Mixed Models

Most datasets in agricultural, biological, and social sciences often fall outside the scope of the traditional methods taught in introductory statistics and statistical methods. Often, these data (response variables) are: (a) binary (the presence or absence of a trait of interest, success or failure, the infection status of an individual, or the expression of a genetic disorder); (b) proportional (the ratio of females to males, infection or mortality rates within a group of individuals); or (c) counts (the number of emerging seedlings, the number of sprouts, etc.), where basic statistical methods attempt to quantify the effects of each predictor variable. However, often, studies of these experiments involve random effects, the purpose of which is to quantify variation among individuals or units. The most common random effects are blocks in experimental or observational studies that are replicated across sites (locations or environments) or over time. Random effects also encompass variations

among individuals (when measuring multiple responses per individual such as survival of multiple offspring or sex ratios of multiple offspring), genotypes, species, and regions or periods over time.

GLMMs are a powerful class of statistical tools that combine the concepts and ideas of generalized linear models (GLMs) with linear mixed models (LMMs). That is, a GLMM is an extension of the GLM, in which the linear predictor contains random effects in addition to fixed effects. These models handle a wide range of both response distributions and scenarios in which observations are sampled. GLMMs extend the theory of LMMs to response variables that have a non-normal distribution. In GLMMs, the response data are not transformed; instead, the explanatory variables are expressed as a linear relationship through a function g of the expectation of $\mathbf{y} \mid \mathbf{b}$; that is, the response is conditional on random effects. This performs the link function that relates the response to the explanatory variables in a linear manner, thus allowing the use of standard LMM techniques for estimation and hypothesis testing.

A conditional model is used to describe a GLMM with non-Gaussian errors (Model 4.1), given a link function (g), as shown below:

$$g(\boldsymbol{\eta}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b},$$

which is a function of the conditional expectation given by

$$E[\mathbf{y} \mid \mathbf{b}] = g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}) = g^{-1}(\boldsymbol{\eta}) = \boldsymbol{\mu} \quad (4.2)$$

where $g^{-1}(\cdot)$ is the inverse link and the other terms have already been mentioned earlier. The fixed and random effects are combined to form the conditional linear predictor

$$\boldsymbol{\eta} = g(E[\mathbf{y} \mid \mathbf{b}]) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} \quad (4.3)$$

The relationship between the linear predictor and the vector of observations is modeled as follows:

$$\mathbf{y} \mid \mathbf{b} \sim (g^{-1}(\boldsymbol{\eta}), \mathbf{R}) \quad (4.4)$$

The above notation (4.4) expresses the conditional distribution of \mathbf{y} , given \mathbf{b} has a mean $g^{-1}(\boldsymbol{\eta})$ and variance \mathbf{R} . Note that instead of specifying the distribution for \mathbf{y} as in the case of a GLM, we specify a distribution for the conditional response $\mathbf{y} \mid \mathbf{b}$.

The variance and covariance matrix for the observations is given by:

$$V(\mathbf{y}) = E[V(\mathbf{y} \mid \mathbf{b})] + V(E[\mathbf{y} \mid \mathbf{b}]) = \mathbf{A}^{1/2} \mathbf{R} \mathbf{A}^{1/2} + \mathbf{Z} \mathbf{G} \mathbf{Z}' \quad (4.5)$$

where matrix \mathbf{A} is a diagonal matrix containing the variance functions of the model. GLMMs cover an important group of statistical models, such as:

- (a) Linear models (LMs): absence of random effects, identity link function and the assumption of a normal distribution.
- (b) Generalized linear models (GLMs): random effects are absent, link function is different from the identity function, and the response variables are non-normally distributed.
- (c) Linear mixed models (LMMs): presence of random effects, identity link function and normal distribution assumed for the response variable.

GLMMs have been formulated to correct the shortcomings of LMMs, as there are many cases where the assumptions made in linear mixed models are inadequate. First, an LMM assumes that the relationship between the mean of the dependent variable (y) and fixed and random effects (β, b) can be modeled through a linear function. This assumption is questionable, like when a researcher wishes to model the incidence of a disease or the success or failure of an event.

The second assumption of an LMM is that variance is not a function of the mean and that the random effects follow a normal distribution. The assumption of constant variance is not met when the response variable is binary (1, 0). In this case, the variance is $\pi(1 - \pi)$, which is a function of the mean. The result is a random variable, which can take two values (0, 1); in contrast, the normal distribution can take any real number. Finally, the predictions for an LMM can take any real value, whereas the predictions for a binary variable are bounded in the interval (0, 1), since it is a probability and this prediction cannot support negative values.

Historically, a number of options have been used to address and solve some LMM problems, even though their use is not the most appropriate. These include applying logarithmic transformations ($\log(y)$), transformations using the square root (\sqrt{y}), arcsine transformations ($\text{seno}^{-1}(y)$), and so on. However, many of these transformations use linear mixed models by ignoring the fact that these models are not the most accurate, despite being aware that the response variable does not satisfy the assumption of normality. These options are attractive because they are relatively simple and easy to implement using the LMM machinery. However, they circumvent the problem that a linear mixed model is not the best model for analyzing data.

4.4 The Inverse Link Function

In a GLMM, the canonical link function maps the original data to the linear predictor of the model $g(\eta) = X\beta + Zb$. This linear predictor can be transformed to an observed data scale through an inverse link function. In other words, the inverse link function is used to map the value of the linear predictor for the i th observation to the conditional mean at the data scale η_i . For example, suppose that we are conducting an experiment in which we are assessing the number of undesirable weeds observed in a crop of interest after the application of a certain number of treatments; the response variable is assumed to have a Poisson distribution with a mean λ_{ij} , the linear predictor of which is given by

$$\eta_{ij} = \eta + \tau_i + b_j$$

where η is the intercept, τ_i is the fixed effect due to treatments, and b_j is the random effect assuming $b_j \sim N(0, \sigma_b^2)$.

To obtain the inverse function of the following predictor

$$\log(\lambda_{ij}) = g(\eta_{ij}) = \eta + \tau_i + b_j,$$

we proceed by exponentiating both sides of the previous equation, with which we obtain the inverse function of the link shown below:

$$\lambda_{ij} = e^{\eta + \tau_i + b_j},$$

which is denoted as $g^{-1}(g(\eta_{ij})) = g^{-1}(\eta + \tau_i + b_j)$.

Therefore, for this example, λ_{ij} depends on the linear predictor through the inverse link function and the variance σ_{ij}^2 depends on λ_{ij} through the variance function.

4.5 The Variance Function

The variance function is used to model the inconsistent variability of the phenomenon under study. With GLMMs, the residual variability arises from two sources, namely, the variability of the distribution of sampling units in an experimental arrangement (blocks, plots, locations, etc.) and the variability due to overdispersion. Overdispersion can be modeled in several ways. When dealing with a GLMM, the scale parameter or the dispersion parameter ϕ is extremely important since it can either increase or decrease the variance in the model for each observation.

$$\text{Var}(y_{ij}|b_j) = \phi \text{Var}(\eta_{ij})$$

If overdispersion exists, one way to remove it is to add the random effects (in SAS `_residual_`) of each observation to the linear predictor. Another alternative is to use another distribution to model the dataset; for example, the two-parameter negative binomial (NB) distribution (η_{ij}, ϕ) instead of the single-parameter Poisson distribution (λ_{ij}) in the case of count data.

4.6 Specification of a GLMM

A GLMM is composed of three parts: (1) fixed effects that convey systematic and structural differences in responses; (2) random effects that convey stochastic differences between blocks or other random factors, as these effects allow generalizations

Table 4.1 Common distributions with their respective link functions

Distribution	Link function	Distribution syntax	Syntax of the link function
Binomial	Logit or probit	dist = binomiallbinlb	link = logit or probit
Poisson	Log	dist = PoissonlPoi	link = log
Beta	Logit	dist = beta	link = logit
Normal	Identity	dist = normallgaussian	link = identitylid
Negative binomial	Log	dist = negbinomiallnegbinl nb	link = log
Multinomial	Cumulative logit	dist = multinomialmulti	link = cumlogitlclogit

of the population from which the sampling units have been (randomly) sampled; and (3) distribution of errors. Thus, a complete definition of a GLMM is as follows:

$$\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\phi} \sim f(\boldsymbol{\mu}, \boldsymbol{\phi}) \text{ (conditional distribution)}$$

$$\mathbf{b} \sim N(0, \mathbf{G}) \text{ (random effects)}$$

$$g(\boldsymbol{\mu}) = \boldsymbol{\eta} \text{ (link function)}$$

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} \text{ (linear predictor)}$$

where the distribution function $f(\cdot)$ is a member of the exponential family, $g(\boldsymbol{\mu})$ is the linear function, \mathbf{X} and \mathbf{Z} are the design matrices, and $\boldsymbol{\beta}$ and \mathbf{b} are the unknown parameters for fixed and random effects, respectively.

When fitting a GLMM, the data remain on the original measurement scale (data scale). However, when means are estimated from a linear function of the explanatory variables (the predictor), these means are on the model scale. A link function is used to link the model scale back to the original data scale. This is not the same as transforming the original measurements to a different measurement scale. For example, applying the log transformation for counts followed by an analysis of variance (ANOVA) under a normal distribution is not the same as fitting a generalized linear model, assuming a Poisson distribution and using a log link (Gbur et al. 2012). In the first case, the least squares means would normally be equal to the arithmetic means, whereas in the second case, the means are inversely linked to the data scale, which may not be equal to the arithmetic means of the original sample.

The distribution specifications in “proc GLIMMIX” have default link functions, but it is always highly recommended to explicitly code the link function, since for some type of response variable, more than one alternative exists. This way, there is no doubt that an appropriate function was used. Using the wrong link function will lead to totally meaningless and incorrect results. Table 4.1 shows some common distributions, the appropriate link function, and the proper syntax for each.

For a complete list, see the online Statistical Analysis Software (SAS/STAT) documentation for PROC GLIMMIX.

4.7 Estimation of the Dispersion Parameter

The overall measures of fit compare the observed values of the response variable with fitted (predicted) values. The dispersion parameter is unknown and therefore must be estimated. There are two methods for estimating the overdispersion parameter. McCullagh (1983) proposed estimating overdispersion as follows:

$$\phi = \frac{(\mathbf{y} - \boldsymbol{\mu})' \mathbf{V}_{\boldsymbol{\mu}}^{-1} (\mathbf{y} - \boldsymbol{\mu})}{N - p} = \frac{\text{Pearson's } \chi^2}{N - p}$$

where $\mathbf{V}_{\boldsymbol{\mu}}^{-1}$ is the diagonal matrix of the variance functions and $N - p$ is the degree of freedom for lack of fit. Later, McCullagh and Nelder (1989) suggested using deviance

$$\hat{\phi} = \frac{\text{Deviance}}{N - p} = \frac{-2[\ln(LM_1) - \ln(L(M_2))]}{N - p}$$

Deviance is a global fit statistic that also compares fitted and observed values; however, its exact function depends on the likelihood function of the random component of the model. Deviance compares the maximum value of the likelihood function of a model, like M_1 , with the maximum possible value of the likelihood function that is calculated using data. When data are used in the likelihood function, the model is saturated and has as many parameters as possible. Thus, M_2 is saturated and has as many parameters as the data. Model M_2 tries to fit the data and gives the highest possible value for the likelihood.

If the overdispersion parameter is significantly greater than one, this indicates that overdispersion exists; in other words, it indicates that the variance is greater than the mean. Therefore, the parameter should be used to adjust the variance. If overdispersion is not taken into account, inflated test statistics may be generated. However, when the dispersion parameter is less than 1, the test statistics are more conservative, which is not considered a big problem.

The following example is intended to show how GLIMMIX in SAS estimates the dispersion parameter in a GLMM.

Example An agronomist wants to test the effectiveness of a new herbicide offered on the market (we will denote this as herb_N) and compare it with the herbicide that has been used for several cycles (herb_C). The experimental arrangement used was a randomized complete block design as shown below (Table 4.2).

The components of a GLMM with a Poisson response variable are listed below:

$$\begin{aligned} \text{Distribution: } y_{ij} \mid b_j &\sim \text{Poisson}(\lambda_{ij}) \\ b_j &\sim N\left(0, \sigma_{\text{bloque}}^2\right) \end{aligned}$$

Table 4.2 Number of undesirable weeds per plot

Block	Herb_C	Herb_N
1	1	36
2	5	109
3	21	30
4	7	48
5	2	3
6	6	0
7	0	5
8	19	26

Linear predictor: $\eta_{ij} = \eta + \text{herbicide}_i + b_j$

Link function: $\log(\lambda_{ij}) = \eta_{ij}$

This model assumes that the slopes are the same for each herbicide. The following SAS code is used for the proposed model:

```
proc glimmix nobound method=laplace;
class block trt;
model count = trt/dist=poisson link=log;
random block;
lsmeans trt/ilink lines;
run;
```

Explanation The “method = ”option is used to specify the method used to optimize the logarithm of the likelihood function. In “proc GLIMMIX,” there are two popular methods: adaptive quadrature (quad) or Laplace (laplace), which are the preferred methods for categorical response variables. Both of these methods fit a conditional model. When the quadrature method is used (method = quad), subjects (individuals) must be declared in the random effects (e.g., for the above program, “random intercept/subject=block”). In addition, processing random effects by subject is more efficient than using the syntax “random block” random effects in blocks. The “dist” option is where you specify the probability distribution that is appropriate for the type of response; in this case, it is the Poisson distribution. The “link” option is for specifying the link function of the distribution. The “ddfm” option is omitted so that GLIMMIX uses – by default – the method for calculating the denominator degrees of freedom for the fixed effects tests that result from the model. The “ilink” option converts the estimates of the treatment means (lsmeans) on the model scale to the data scale. Finally, “proc GLIMMIX” supports the “lines” option, which adds letter groups to the mean differences resulting from using “lsmeans.”

The most relevant parts of the SAS output, for the purposes of what we want to show, are shown in Tables 4.3 and 4.4. The fit statistics of the fitted model are shown in part (a) and part (b) of Table 4.3. The $-2 \log$ likelihood statistic is extremely

Table 4.3 Fit statistics and variance components

(a) Fit statistics (Akaike’s information criterion (AIC), a small sample bias corrected Akaike’s information criterion (AICC), Bozdogan Akaike’s information criterion (CAIC), Schwarz’s Bayesian information criterion (BIC), Hannan and Quinn information criterion (HQIC))		
–2 Log likelihood	175.35	
AIC (smaller is better)	181.35	
AICC (smaller is better)	183.35	
BIC (smaller is better)	181.59	
CAIC (smaller is better)	184.59	
HQIC (smaller is better)	179.74	
(b) Fit statistics for conditional distribution		
–2 Log L (count r. effects)	139.03	
Pearson’s chi-square	77.56	
Pearson’s chi-square/degree of freedom (DF)	4.85	
(c) Covariance parameter estimates		
Cov Parm	Estimate	Standard error
Block	1.5590	0.8690

Table 4.4 Type III fixed effects tests and estimated least squares means

(a) Type III tests of fixed effects							
Effect	Num DF	Den DF	F-value	Pr > F			
Herbicide	1	7	101.34	<0.0001			
(b) Trts least squares means							
Trts	Estimate	Standard error	DF	t-value	Pr > t	Mean	Standard error mean
Herb_C	1.4604	0.4696	7	3.11	0.0171	4.3076	2.0227
Herb_N	2.8947	0.4561	7	6.35	0.0004	18.0778	8.2447

useful for comparing nested models, whereas the different versions of information criteria that exist, such as Akaike information criterion (AIC), Akaike’s information criteria with small sample bias correction (AICC), Bayesian information criterion (BIC), Bozdogan Akaike’s information criteria (CAIC), and Hannan and Quinn information criteria (HQIC), are useful when comparing models that are not necessarily nested (subsection (a)). The table of fit statistics for the conditional distribution shows the sum of the independent contributions to the conditional (part (b)) –2 log likelihood, the value of which is 139.03, whereas the value of Pearson’s statistic divided by the degrees of freedom for the conditional distribution (Pearson’s chi – square/DF) is 4.85.

The estimated dispersion parameter ($\phi = \text{Pearson’s chi-square/DF}$) has a value far from 1; in this case, it is $\hat{\phi} = 4.85$, which indicates that there is a strong overdispersion. This may be because the specified distribution of the data is not appropriate, the counts are too small, or the variance function was not correctly

specified. The estimate of the variance component due to a block is tabulated in part (c) of Table 4.2, the estimated value of which is $\hat{\sigma}_{\text{block}}^2 = 1.559$.

The fixed effects test and least squares means are shown in Table 4.4. The type III fixed effects tests indicate that there is a highly significant difference (part (a)) in the effectiveness of herbicides in weed suppression; the estimated means with their respective standard errors are tabulated under the “Mean” column (part (b)). The “Estimate” column containing the estimates of the means of lsmeans is on the model scale. They are derived from the log likelihood function. SAS always lists the means obtained with lsmeans from the model scale when creating least squares means test tables. The “Mean” column has been converted back to the data scale using the “ilink” inverse link function. These values are estimates of the average counts for each treatment level (in this case, the herbicide type on the data scale). When we report the results, we must replace the corresponding model’s least squares values in the test tables with these estimates (means on the data scale corresponding to the values in the “Mean” column).

Since there is a strong overdispersion ($\hat{\phi} > 1$), assuming that the data have a Poisson distribution is risky because this implies that the mean and variance are equal, which is an assumption implying that the data have a Poisson distribution, i.e., that the mean and variance are the same. A useful alternative distribution might be a negative binomial distribution; this distribution has a mean λ and variance $\lambda + \lambda\phi^2$ with $\phi > 0$ commonly known as the scale parameter.

The following is the specification of the components of a GLMM with a negative binomial (NB) response variable:

$$\text{Distribution : } y_{ij} \mid b_j \sim \text{Negative binomial}(\lambda_{ij}, \phi)$$

$$b_j \sim N(0, \sigma_{\text{block}}^2)$$

$$\text{Linear predictor: } \eta_{ij} = \eta + \text{herbicide}_i + b_j$$

$$\text{Link function: } \log(\lambda_{ij}) = \eta_{ij}$$

The GLIMMIX procedure also allows modeling a GLMM with a negative binomial response variable:

```
proc glimmix data=itam nobound method=laplace;
class block trts;
model count = trts/dist=negbin;
random block;
lsmeans trts/ilink;
run;
```

Part of the output is shown in Table 4.5. The fit statistics for the model comparison (part (a)) and that for the conditional distribution (part (b)) are both provided by the GLIMMIX procedure when a conditional distribution is specified. Since in the previous analysis, it was observed that overdispersion exists when assuming a Poisson distribution, the results – under a negative binomial distribution – indicate that this overdispersion problem no longer exists; i.e., the binomial distribution is no

Table 4.5 Fit statistics under a negative binomial distribution

(a) Fit statistics		
-2 Log likelihood		120.50
AIC (smaller is better)		128.50
AICC (smaller is better)		132.13
BIC (smaller is better)		128.81
CAIC (smaller is better)		132.81
HQIC (smaller is better)		126.35
(b) Fit statistics for conditional distribution		
-2 Log L (count r. effects)		120.50
Pearson's chi-square		9.21
Pearson's chi-square/DF		0.58
(c) Fit statistics and Pearson's chi-square/DF		
	Poisson	Negative binomial
-2 Log likelihood	175.35	120.50
AIC (smaller is better)	181.35	128.50
AICC (smaller is better)	183.35	132.13
BIC (smaller is better)	181.59	128.81
CAIC (smaller is better)	184.59	132.81
HQIC (smaller is better)	179.74	126.35
$\hat{\phi}$ (Pearson's chi-square/DF)	4.85	0.58

longer overdispersed ($\hat{\phi} = 0.58$). In other words, the negative binomial distribution does a better job than the Poisson distribution in fitting these data, since it effectively controls the overdispersion.

Comparing the fit statistics tabulated in Table 4.3 subsection (c) under both distributions, we can observe that when the data are modeled under a negative binomial distribution, the values of the fit statistics are lower than those under a Poisson distribution, since the dispersion parameter $\hat{\phi} < 1$. This indicates that the negative binomial models this dataset better.

4.8 Estimation and Inference in Generalized Linear Mixed Models

4.8.1 Estimation

In GLMMs, inference involves the estimation and testing of the hypotheses of unknown parameters in β , G , and R as well as the best linear unbiased predictions (BLUPs) of random effects, b . In most modern statistical tools, including GLMMs, parameter fitting is performed via maximum likelihood (ML) or methods derived from this method. For simple analyses, in which the response variables are normal, classical ANOVA methods are based on calculating the differences of the sums of

the squares that produce the same results as an ML estimation. However, this equivalence is not obtained in models with more complex structures such as LMMs or GLMMs. To find the ML estimators, in GLMMs, one must integrate over all possible values of the random effects. For GLMMs, this computation is at best slow and at worst (a large number of random effects) computationally infeasible.

Statisticians have proposed several ways to approximate the parameter estimates of a GLMM, including penalized quasi-likelihood (PQL) and pseudo-likelihood methods (Schall 1991; Wolfinger and O'Connell 1993; Breslow and Clayton 1993), Laplace approximations (Raudenbush et al. 2000) and Gauss–Hermite quadrature (Pinheiro and Chao (2006), and Bayesian methods based on Markov chain Monte Carlo (Gilks et al. 1996). In all these approaches, researchers must distinguish between a standard ML estimation, which estimates the standard deviations of the random effects assuming that the fixed effects estimates are precisely correct, and restricted maximum likelihood (REML), a variant that averages over the uncertainty in the fixed effects parameters (Pinheiro and Bates 2000; Littell et al. 2006).

The ML method underestimates the standard deviations of random effects, except in extremely large datasets, but it is most useful for comparing models with different fixed effects. Pseudo- and quasi-likelihood methods are the simplest and the most widely used in approximating a GLMM. They are widely implemented in statistical packages that promote the use of GLMMs in many areas of ecology, biology, and quantitative and evolutionary genetics (Breslow 2004). Unfortunately, pseudo- and quasi-likelihood methods produce biases in parameter estimation if the standard deviations of the random effects are large, especially when using binary data (Rodriguez and Goldman 2001; Goldstein and Rasbash 1996). Lee and Nelder (2001) have implemented several improvements to the PQL version, but these are not available in most common statistical software packages. As a rule of thumb, PQL performs poorly for Poisson data when the average number of counts per treatment combination is less than five or for binomial data when the expected numbers of successes and failures for each observation are less than five (Breslow 2004). Another disadvantage of PQL is that it calculates a quasi-likelihood rather than the true likelihood. Because of this, many statisticians believe that PQL-based methods should not be used for inference.

There are two more accurate approximations available, which also reduce bias. One is the Laplace approximation (Raudenbush et al. 2000), which approximates the true likelihood of a GLMM instead of a quasi-likelihood, allowing the maximum likelihood method in the GLMM inference process. The other approach is called Gauss–Hermite quadrature (Pinheiro and Chao 2006), which is more accurate than the Laplace approximation but is slower (requires more computational resources). Therefore, the procedures for parameter estimation of a GLMM that are approximations are as follows:

The penalized quasi-likelihood method performs the estimation process by alternating between (1) estimating the fixed parameters by fitting a GLM with a variance–covariance matrix based on an LMM fit and (2) estimating the variances and

covariances by fitting a GLM with unequal variances calculated from the previous GLM fit. Pseudo-likelihood, a close cousin of the ML method, estimates variances differently and estimates a scale parameter to account for overdispersion (some authors use these terms interchangeably). In summary, GLMMs require an iterative process in parameter estimation. Two categories of iterative procedures are used by SAS: linearization and integral approximation. The GLIMMIX procedure uses the pseudo-likelihood method in linearization, and integral approximation uses the Laplace approximation or adaptive methods such as Gauss–Hermite quadrature. These methods maximize the log likelihood of the exponential distribution family, i.e., non-normal distributions. The pseudo-likelihood method is the default procedure in the GLIMMIX procedure (Proc GLIMMIX). The Laplace method and quadrature are an approximation for maximum likelihood, but the Laplace method is computationally simpler than quadrature and also provides excellent estimates.

4.8.2 Inference

After estimating the parameter values in a GLMM, the next step is to extract information and draw statistical conclusions from a given dataset through careful analysis of the parameter estimates (confidence intervals, hypothesis testing) and select a model that best describes or explains the most variability in the dataset. Inference can generally be based on three types: (a) hypothesis testing, (b) model comparison, and (c) Bayesian approaches. Hypothesis testing compares test statistics (F -test in ANOVA) to verify their expected distributions under the null hypothesis (H_0), estimating the value of P (P -value) to determine whether H_0 can be rejected. On the other hand, model selection compares candidate model fits. These can be selected using hypothesis testing; that is, testing nested versus more complex models (Stephens et al. 2005) or using information theory approaches such as Wald tests (Z , χ^2 , t , and F). In model selection, likelihood ratio (LR) tests can ensure the significance of factors or choose the best of a pair of candidate models. On the other hand, information criteria allow multiple comparisons and selections of non-nested models. Among these criteria are the Akaike information criterion (AIC) and related information criteria that use deviance as a measure of fit, adding a term to penalize more complex models. Information criteria can provide better estimates. Variations of AIC are highly common when sample sizes are not large (AICC), when there is overdispersion in the data (quasi-AIC, QAIC), or when one wishes to identify/determine the number of parameters in a model (Bayesian information criterion, BIC).

4.9 Fitting the Model

The mathematics behind a GLMM is quite complex. It is difficult to conceptualize the use of constructs such as distributions, link functions, log likelihood, and quasi-likelihood when fitting a model. Perhaps the following points will help explain the modeling process.

- (a) An analysis of variance model is a vector of linear predictors (equation) with unknown parameter estimates.
- (b) Each distribution has a corresponding probability function.
- (c) The vector of linear predictors is substituted into the likelihood function.
- (d) Solutions to the parameter estimates are found by minimizing the negative of the log likelihood function ($-\log$ likelihood).
- (e) The means (least squares means – lsmeans) are derived from the parameter estimates and are on the model scale.
- (f) The link function converts the mean estimates at the model scale to the original data scale.

The key concepts of proc GLIMMIX are (1) it uses a distribution to estimate the model parameters; it does not fit the data to a distribution, and (2) the data values are not transformed by the link function; the link function converts the means (least squares means) to the data scale after estimation at the model scale.

4.10 Exercises

1. As a simple example of these types of data, consider the following results of an experiment on wheat germination, carried out in pots under glass. The experiment consisted of four blocks of six treatments (Table 4.6).
 - (a) According to the response variable, what type(s) of probability distribution do you suggest for the variable?
 - (b) Construct a GLMM to study the effect of treatments on seed germination.
 - (c) Analyze the dataset according to the model proposed in (a). Is the probability distribution proposed in (a) adequate?
 - (d) Is there a significant difference in the proportion of germinated seeds between treatments?

Table 4.6 Number of seeds not germinating (out of 50)

	Trt1	Trt2	Trt3	Trt4	Trt5	Trt6
A	10	11	8	9	7	6
B	8	10	3	7	9	3
C	5	11	2	8	10	7
D	1	6	4	13	7	10

Table 4.7 Control of cockchafer larvae

	A		B		C		D		E		F		G		H	
	a	B	a	b	a	b	a	b	a	b	a	b	a	b	a	b
Trt1	3	7	7	15	5	7	5	14	0	3	1	7	1	10	4	13
Trt2	4	3	3	12	4	2	12	5	2	3	1	6	3	5	4	11
Trt3	3	10	6	12	4	4	1	14	2	2	1	7	1	8	7	10
Trt4	5	8	4	11	1	5	5	9	2	7	3	7	0	3	3	12
Trt5	4	6	4	11	2	2	3	8	0	1	5	4	1	6	1	8

2. Table 4.7 shows the counts per sample area of a variety type of cockchafer larva (two age groups a and b). The experiment consisted of five treatments in eight randomized blocks and two age groups to study the differential effects of treatments on insect age.
- (a) Considering the type of answer of this exercise; what type(s) of probability distribution(s) do you suggest for this type of response?
 - (b) Construct a GLMM to study the effect of treatments and the age of Cockchafer larvae.
 - (c) Analyze the dataset according to the model proposed in (a).
 - (d) Is the model used in (a) sufficient? If so, discuss your findings.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

