GENERALIZED LINEAR MIXED MODELS - AN OVERVIEW

W. W. Stroup

S. D. Kachman

GENERALIZED LINEAR MIXED MODELS - AN OVERVIEW

by W.W. Stroup and S.D. Kachman

Department of Biometry, University of Nebraska-Lincoln 68583-0712

Abstract: Generalized linear models provide a methodology for doing regression and ANOVA-type analysis with data whose errors are not necessarily normally-distributed. Common applications in agriculture include categorical data, survival analysis, bioassay, etc. Most of the literature and most of the available computing software for generalized linear models applies to cases in which all model effects are fixed. However, many agricultural research applications lead to mixed or random effects models: split-plot experiments, animal- and plant-breeding studies, multi-location studies, etc. Recently, through a variety of efforts in a number of contexts, a general framework for generalized linear models with random effects, the "generalized linear mixed model," has been developed.

The purpose of this presentation is to present an overview of the methodology for generalized mixed linear models. Relevant background, estimating equations, and general approaches to interval estimation and hypothesis testing will be presented. Methods will be illustrated via a small data set involving binary data.

Key Words: Generalized Linear Model, Mixed Model

1. INTRODUCTION

It is often of interest to agricultural researchers to conduct experiments involving random model effects and response variables whose distributions are not normal. For example, consider an experiment to compare 2 different treatments, conducted at several randomly selected locations. At each location, subjects are assigned at random to treatment 1 or treatment 2. Subjects are subsequently evaluated to determine whether their response to the treatment is favorable or unfavorable. Letting n_{ij} be the number of subjects assigned to the i^{th} treatment at the j^{th} location and y_{ij} be the number of subjects having favorable outcomes for the i^{th} treatment at the j^{th} location, the response variable of interest would be $p_{ij} = y_{ij} / n_{ij}$, the proportion of favorable outcomes. A model for this experiment is

$$p_{ij} = \mu + \tau_i + L_j + TL_{ij},$$

where μ is the intercept, τ_i is the i^{th} treatment effect, L_j is the j^{th} location effect, and TL_{ij} is the treatment-location interaction. The location and location-by-treatment effects are random, because locations are randomly sampled.

If the response variable in the above model were normally distributed, the experiment could be analyzed using the following analysis of variance.

Source	EMS
TRT	$\sigma^2 + k_1\sigma_{TL}^2 + \phi_T$
LOC	$\sigma^2 + k_2\sigma_{TL}^2 + k_3\sigma_L^2$
TRT*LOC	$\sigma^2 + k_1\sigma_{TL}^2$
error	σ^2

The constants k_1, k_2, and k_3 are determined by the n_{ij}. One could test for treatment effect using F = MS(TRT)/MS(LOC*TRT).

On the other hand, if treatment and locations were both fixed effects and the response variable was p_{ij}, as defined above, then the experiment could be analyzed using standard linear model methods for categorical - in this case binomial - data (see, for example, Agresti, 1990). Linear models for categorical data are special cases of *generalized linear models*. These models allow χ^2 statistics for TRT and TRT*LOC to be computed, but do not allow TRT*LOC variation to be used in the construction of a test for TRT if locations are random. The test for TRT in the standard categorical linear model corresponds to using the F-ratio MS(trt)/MS(error) in the above ANOVA - and is equally inappropriate.

The model given above illustrates a problem agricultural researchers often face. That is, the experiment is most naturally described by a mixed model, whose variance structure must be taken into account in order to obtain appropriate test statistics and standard errors, but the response variable is not normal, meaning that standard mixed model methods are not applicable. Other examples in which mixed models and non-normal response variables occur are animal- and plant-breeding experiments with random sire, dam, or entry effects and response variables such as calving difficulty or disease-resistance ratings, split-plot experiments with response variables such as insect count or botanical composition, etc.

Nelder and Wedderburn (1972) introduced the generalized linear model as a generalization of standard linear models. They showed that regression and analysis of variance methods could be applied to any response variable whose distribution belongs to the exponential family. Comprehensive presentations of the generalized linear model are given in such texts as McCullagh and Nelder (1989), Dobson (1990), and Aitkin, et. al. (1989). Despite its versatility, the usefulness of the generalized linear model as presented in these texts is limited by the fact that it is strictly a fixed-effects model.

A number of articles presenting methods for specific mixed models, mostly with binary data, appeared in the 1980's - Harville and Mee (1984), Gilmour (1985), Beitler and Landis (1985). Zeger, et. al. (1988), Breslow and Clayton (1993), and Vonesh and Carter (1993) presented more comprehensive approaches to extending the generalized linear model to the random effects case. However, these articles are oriented toward biomedical applications, limiting their applicability to specific issues of interest to agricultural researchers. The purpose of this paper is to provide an overview of the generalized linear model with random effects in a framework that addresses statisticians who work with agricultural problems.

This paper will be organized in five sections, the first being the introduction. The second section will contain a brief introduction to the generalized linear model. The third section will contain a brief review of the mixed model with normal errors. The fourth section will present the "marriage," that is, the basic elements of the generalized linear mixed model, the estimation procedure, and the primary tools for statistical inference. The fifth section will contain an example using the two-way model with binary data described above. The data used in this example appear in Beitler and Landis (1985).

2. THE GENERALIZED LINEAR MODEL

At the risk of creating confusion, we will use the acronym "GLM" to refer to the *generalized linear model*. GLM is also used as an acronym for the more restrictive, normal errors "general linear model," and as the name of a well-known statistical computing procedure. However, we agree with an anonymous comment to the effect that it is time to begin to refer to the normal errors model as the "*specific* linear model." GLM is the standard jargon among people who work with generalized linear models.

The central idea of the generalized linear model (GLM) can be understood by first considering the traditional "general linear model." This model is

$$y = X\beta + e,$$

where y is the vector of observations, X is a matrix of known constants, β is the parameter vector, and e is the error vector, distributed $N(0,I\sigma^2)$. The main role of the linear model is to characterize E(y) by the linear combination of parameters $X\beta$.

The *GLM* also characterizes E(y) by $X\beta$. However, because of the probability distribution of the errors, it often is more reasonable to model a function of E(y). Letting $\mu = E(y)$, the function $\eta = g(\mu)$ is modeled by $X\beta$ in the GLM. Following Nelder and Wedderburn's (1972) terminology, $g(\mu)$ is called the *link function* - it "links" μ to the linear model $X\beta$.

Generalized linear models can be applied to any data whose error distribution belongs to the exponential family. The basic form of the log-likelihood function for members of the exponential family is

$$L(\alpha, \phi, y) = \frac{y'\alpha - b(\alpha)}{a(\phi)} + c(y, \phi)$$

Where α is the canonical, or natural, parameter, ϕ is a scale parameter, and $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ are specific functions whose form depends on the particular distribution.

Common examples include the normal, poisson, and binomial distributions, whose log-likelihoods, expressed in the above form, are given with their expected value and variance in Table 1.

Because it is linear with respect to the observation vector, α is often a desirable link function, i.e. an appropriate function of E(y) to model by $X\beta$. While the natural parameter is not the only useful form of the link function, it does suggest a starting point. We will discuss this point further in the example in section 5.

For members of the exponential family, the GLM is thus

$$\eta = X\beta,$$

where $\eta = g(\mu)$ is the link function and $\mu = E(y)$. The function $h(\eta) = \mu$, where $h(\cdot) = g^{-1}(\cdot)$, is called the *inverse link*. In many applications, including the generalized linear *mixed* model, it is more useful, and often necessary, to define the GLM in terms of the inverse link rather than the link function.

The parameter vector, β, can be estimated using maximum likelihood. Nelder and Wedderburn (1972) showed that the maximum likelihood estimate can be obtained by solving the generalized least squares equations

$$(X'H'R^{-1}HX)\,\beta = XH'R^{-1}y^*$$

where $R = Var(y)$,

$$H = diag\left(\frac{\partial h(\eta)}{\partial \eta}\right)$$

and

$$y = y - \mu + H\eta$$

Inference for the GLM uses three basic tools.

1. For linear combinations of k'ß, where k' is a vector, $k'\hat{\beta}$ is distributed approximately $N(k'ß, k'(XH'R^{-1}HX)^{-1}k)$. This is used to obtain confidence intervals for $k'\beta$ and to test hypotheses of the form H_0: $k'\beta = k'\beta_0$. Typical $k'\beta$ of interest are treatment means, specific regression parameters, treatment differences, and contrasts.

2. The Wald statistic, $(K'\hat{ß})'[K'(XH'R^{-1}HX)^{-1}K]^{-1}(K'\hat{ß})$, where K' is not necessarily a vector, is distributed approximately $\chi^2_{rank(K)}$. This result is used to test hypotheses of the form H_0: K'ß = 0. Typical $K'\beta$ of interest are the hypothesis of no overall treatment effect, no overall main effect or interaction, where a treatment or factor has more than one degree of freedom.

3. The deviance, $2\{\ln[L(\alpha_{max},y)] - \ln[L(\alpha_{Xß},y)]\}$, is a generalization of the MS(error) in standard ANOVA and the likelihood ratio lack of fit statistic in a contingency table. It is distributed approximately $\chi^2_{(N-rank(X))}$. It is used to evaluate the lack of fit of the generalized linear model. The difference between deviances for various models may be used to construct likelihood ratio tests. This method is often used as an alternative to the Wald statistic.

Methods for evaluating model fit, influence, adequacy of assumptions, appropriateness of link function, etc., for GLM's are discussed in detail in texts such as McCullagh and Nelder (1989), Dobson (1990), and Aitkin, et. al. (1989). These texts discuss a variety of applications, including normal errors models - i.e. standard regression and ANOVA - log-linear, logistic, and probit models for categorical data, proportional hazards models for survival analysis, etc. Recent developments include quasi-likelihood and pseudo-likelihood methods, discussed in McCullagh and Nelder (1989), which permit working with certain distributions that are not members of the exponential family. The generalized linear *mixed* model draws on some of these methods.

3. THE MIXED MODEL

The mixed model is well-known in statistics. Eisenhart (1947) identified three types of linear models: the fixed effects, random effects, and mixed model. Most standard statistical texts, e.g. Snedecor and Cochran (1990) and Steel and Torrie (1980), introduce aspects of the mixed model. Henderson developed much of what is now commonly considered mixed model methodology, most importantly the mixed model equations and their properties and the method of "best linear unbiased prediction," or "BLUP." Although much of the work was done in the 1940's and 1950's, the most complete reference is Henderson (1984a). A number of researchers, most notably Harville (1976, 1977) and Miller (1977), have provided theoretical support for mixed model methods. Recent mixed model articles include McLean, Sanders and Stroup (1991), and Robinson (1991). Also of interest is the Southern Regional cooperative Series Bulletin, *Mixed Models in Agriculture and Related Disciplines* (1989), which provides several examples of applications in agriculture.

We will present only a brief introduction to the mixed model, focusing on details relevant to understanding the generalized linear mixed model.

The general form of the mixed model is

$$y = X\beta + Zu + e,$$

where y is the vector of observations,
X is a matrix of known constants associated with the fixed effects,
β is a vector of fixed effects,
Z is a matrix of known constants associated with the random effects,
u is a vector of random model effects, and
e is a vector of random errors.

The joint distribution of the random models effects and errors is

$$\begin{bmatrix} u \\ e \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}\right)$$

Thus E(y) = Xß and Var(y) = ZGZ' + R. For the purposes of connecting the mixed model and the generalized linear model, it is useful to note that the conditional expectation of y, E(y|u) = Xß + Zu, and the conditional variance is Var(y|u) = R.

The model effects are estimated by solving the mixed model equations

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix}\begin{bmatrix} \beta \\ u \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

When G and R are known, the resulting solution for β is BLUE and the resulting solution for u is BLUP. In typical applications, the components of G and R must be estimated. There are many methods to estimate variance and covariance components. See Henderson (1984b) or Searle, et. al. (1992) for a detailed treatment of this subject. Most mixed model statistical software permits estimating the components of G and R by method of moments or, more commonly, restricted maximum likelihood (REML).

Inference for the mixed model is based on *estimable functions*, i.e. functions of the form K'β such that estimability criteria (see Searle, 1971) are satisfied, or *predictable functions*, i.e. functions of the form K'β + M'u, such that K'β is estimable. Examples of estimable functions include treatment means, differences and contrasts. Examples of predictable functions, or "best linear unbiased predictors," include animal breeding values, predicted patient outcomes in clinical trials, and predicted performance at a specific location in multi-location trials. The basic tools for inference are analogous to those for generalized linear models. Let L' = [K' M'], and θ = [β, u]. Then

1. When L is a vector, L'($\hat{\theta}$ - θ) is approximately distributed N(L'θ, L'CL). This result in useful for confidence intervals and simple tests of hypotheses.

2. When L is a matrix and G and R are known, the Wald statistic,

$$(L'\hat{\theta})'(L'CL)^{-1}(L'\hat{\theta})$$

may be used to test H_0: $L'\theta = 0$. The Wald statistic is distributed approximately $\chi^2_{rank(L)}$.

3. For estimated G and R, the F statistic

$$(L'\hat{\theta})'(L'CL)^{-1}(L'\hat{\theta}) \,/\, rank(L)$$

has an approximate F distribution and is used to test H_0: $L'\theta = 0$. In many mixed model applications, the random effects lead to expected mean squares which imply appropriate ratios of mean squares (or linear combinations of mean squares) for testing various hypotheses. The split-plot ANOVA is a common example. For balanced data and tests defined on $K'\beta$ only (i.e. M=0), the above F-statistic is identical to the ratio determined from the expected mean squares. The numerator degrees of freedom correspond to rank(L) and the denominator degrees of freedom equal the degrees of freedom of the corresponding denominator mean square or can be obtained using Satterthwaite's approximation. Jeske and Harville (1988) discuss this topic in detail. McLean and Sanders (1988) examined the small sample properties of the Satterthwaite approximation.

The mixed model may be used in a variety of applications. Split-plot models, models for multi-location experiments, quantitative genetics models, etc., are common application of mixed models. Since G and R are general, models with correlated errors, e.g. auto regression models and models with spatial variability, are special cases of the mixed model. The main limitation of the mixed model is the requirement of normally distributed errors. We turn our attention now to the generalized linear mixed model, which drops this requirement.

4. THE GENERALIZED LINEAR MIXED MODEL

The generalized linear mixed model (GLMM) involves fixed effects and normally distributed random model effects as in the "traditional" mixed model, but the error distribution is more general. Errors may have any distribution belonging to the exponential family.

To describe the basic features of the generalized mixed linear model, let

$\mu = E(y|u)$,
$R = Var(y|u)$,
$u \sim N(0,G)$, and
$\eta = g(\mu)$,

where $g(\mu)$ is the link function. Then the generalized linear mixed model has the form

$$\eta = X\beta + Zu,$$

where X, β, Z, and u are defined as before. As shown in Breslow and Clayton (1993), the model effects β and u can be estimated by solving a generalized form of the mixed model equations:

$$\begin{bmatrix} X'HR^{-1}HX & X'HR^{-1}HZ \\ Z'HR^{-1}HX & Z'HR^{-1}HZ + G^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ u \end{bmatrix} = \begin{bmatrix} X'HR^{-1}y^{*} \\ Z'HR^{-1}y^{*} \end{bmatrix}$$

where

$$H = diag\left[\frac{\partial h(\eta)}{\partial \eta}\right]$$

and

$$y^{*} = y - \mu + H\eta$$

The generalized mixed model equation is obtained from solving the joint quasi-likelihood function of the observations and random model effects. From Breslow and Clayton (1993), the quasi-likelihood has the form

$$L(y, u) = -\frac{1}{2}[\ln|G| + u'G^{-1}u] + \frac{y'\theta - b(\theta)}{a(\phi)} - \frac{q}{2}\ln(2\pi)$$

The reader is referred to Breslow and Clayton for more detail. For a more general treatment of the method of quasi-likelihood, see McCullagh and Nelder (1989).

It is worth noting that the generalized mixed model equations provide an overall framework for linear models. That is, solutions for other linear models are special cases of the generalized mixed model equations. For example, if the identity link, $\eta = \mu$, is used, then H=I, yielding the "traditional" mixed model equations given in section 3. If η is general, but the model has only fixed effects, then we have the weighted least squares solution equations for the GLM given in section 2. The identity link *and* the fixed effects model yields the solution equations for generalized least squares, $(X'R^{-1}X)\beta = X'R^{-1}Y$. Finally, the identity link, fixed effects model with $R=I\sigma^2$ yields the normal equations for ordinary least squares, $X'X\beta = X'y$.

The components of G and R are usually unknown in practical applications. The components of R are often functions of μ, as in the binomial and Poisson cases. The components of G and components of R that are not functions of μ may be estimated using REML-like procedures, that is, using the same computing formula that would be used in REML estimation of variance components for normal errors mixed models. Thus the procedure for obtaining estimates of ß and u is iterative, beginning with starting values for G and R, estimating ß and u, using these to estimate the components of G and R, etc.

One way to conceptualize the GLMM is to imagine a set of fixed parameters, β, and random variables, u which are NID(0,G). These yield a *linear predictor*, $\eta = X\beta + Zu$. The expected value of the observations, $\mu = E(y \mid u)$, is related to the linear predictor by the *inverse link*, $h(\eta) = \mu$. The observations actually obtained are the sum of the conditional expectation and *noise*, i.e. $y = \mu +$ noise. The noise has some distribution belonging to the exponential family, and has variance $V(y|u) = R$. Because the inverse link, $h(\eta)$ is not necessarily linear, it makes more sense to fit a linear model to η than to fit a model directly to y. Once the model is fit to η, predictions can be made about y through the

inverse link, or model effects can be tested or compared.

Inference on the GLMM uses the same basic ideas as the traditional mixed model. That is, inference is based on estimable functions, $K'\beta$, or predictable functions, $K'\beta + M'u$. The rationale for specific estimable or predictable functions is identical to that used in section 3 for the mixed model. Basic results are given as follows. As in section 3, let $L' = [K' \; M']$, and $\theta' = [\beta' \; u']$.

1. When L is a vector, $L'\hat{\theta}$ is distributed approximately $N(L'\theta, L'CL)$, where C is the generalized inverse of the right hand side of the generalized mixed model equations.

2. The Wald statistic for $H_0: L'\theta = 0$ is

$$(L'\hat{\theta})'(L'CL)^{-1}(L'\hat{\theta}).$$

Its distribution is approximately $\chi^2_{\text{rank}(L)}$.

While considerable further work needs to be done to evaluate the small-sample properties of these approximations, the same general guidelines as presented in section 3 for the normal errors mixed model appear to apply. That is, they are reasonable when M=0, but less so otherwise. We now present an example of the GLMM.

5. AN EXAMPLE

This example uses data which appear in Beitler and Landis (1985). This was an early article on mixed models for binary data. The method they present can be thought of as a precursor to the generalized linear mixed model.

The data, given in Table 2, involved two treatments and eight randomly selected clinics. At each clinic, patients were assigned to receive either treatment 1 or treatment 2. Patients were classified as having favorable or unfavorable response to the treatment they received. Thus the response variable was p_{ij}, as described in section 1. Beitler and Landis used the model

$$p_{ij} = \mu + \tau_i + L_j + TL_{ij},$$

as described in section 1. Thus, we have a mixed model with a non-normal response variable. In generalized linear model terms, Beitler-Landis model used the *identity link*, since they predict the conditional expectation of p_{ij} given the random effects L_j and TL_{ij} directly. This point will be discussed in more detail below.

To fit a generalized linear mixed model to these data, the following items must be identified:

- The error distribution
- The link function, or inverse link
- The form of G

The error covariance matrix, R, is a consequence of specifying the error distribution. The link function determines the form of the H matrix and y^* vector used in the generalized mixed model equations.

For this example, the response variable is the sample proportion, $p_{ij} = f_{ij} / n_{ij}$, the number of favorable outcomes divided by the number of subjects for the ijth treatment-clinic combination. Thus the error distribution is binomial divided by n_{ij}. The conditional expectation of p_{ij} given the random effects is the probability of a success for treatment i and clinic j. Denote the conditional expectation as π_{ij} = P{success| trt i, clinic j} = $E(p_{ij} \mid u)$. The variance-covariance matrix $R = Var(p_{ij} \mid u)$ is thus diagonal and its elements are $\pi_{ij}(1-\pi_{ij})/n_{ij}$.

The link function may be chosen from a number of alternatives. Typical alternatives include

1. The identity link, $\eta_{ij} = \pi_{ij}$

   The advantage of using the identity link is its simplicity. Agricultural researchers often find the identity link attractive because it appears to be easier to understand. However, there is no guarantee that the estimated π_{ij} will be between 0 and 1, making interpretation problematic. Also, because the binomial distribution is not linear with respect to p_{ij}, the identity link typically yields a poorer fitting model than the alternatives.

2. The logit link, $\eta_{ij} = \ln[\pi_{ij}/(1-\pi_{ij})]$, or $\pi_{ij} = \exp(\eta_{ij})/[1 + \exp(\eta_{ij})]$

   This is the *canonical link* for the binomial distribution. That is, the link function is the natural parameter of the binomial distribution; hence, the binomial is linear with respect to the logit. Estimated π_{ij} are bounded between 0 and 1. The main disadvantage of the logit link results from the conventions for journal article reporting preferred by most agricultural researchers. They wish to report results in terms of treatment means or differences among π_{ij}. This requires converting model estimates and their standard errors using the inverse link. Because the inverse link is not linear, estimable or predictable functions defining differences and other contrasts cannot be converted directly.

3. The probit link, $\eta_{ij} = \Phi^{-1}(\pi_{ij})$, or $\pi_{ij} = \Phi(\eta_{ij})$, where $\Phi(.)$ is the normal c.d.f.

   This model assumes there is some underlying, unobservable, quantitative, normally distributed process η_{ij}. When this process is below some *threshold value* (see Figure 1), the outward, observable response is a "failure," or an unfavorable outcome. When η_{ij} is above the *threshold* the observed response is a "success" or favorable outcome. Like the logit link, the resulting estimated π_{ij} are bounded between 0 and 1. Also, reporting results in terms of estimated π_{ij} requires considerations analogous to the logit link. The concept of an underlying normal process gives the probit link an advantage in certain applications. For example, in animal breeding, much of the theory in quantitative genetics is based on the normal distribution. Using the probit link, this theory can be applied to binary data without modification.

The form of the G matrix follows from the assumption that clinic effects, the c_j are i.i.d. normal with variance σ_C^2 and treatment-by-clinic effects are i.i.d. normal with variance σ_{TC}^2. The G is as given in Table 3.

Table 2 gives the best linear unbiased predictors (BLUP's) of all treatment-clinic combinations using the logit link function. These are computed using the predictable function $\eta_{ij} = m + \tau_i + c_j + \tau c_{ij}$. Estimated η_{ij} are computed using the solution to the generalized mixed model equation and the REML-like variance component estimator described above. The estimated π_{ij} are then computed using the inverse link $\exp(\eta_{ij})/[1 + \exp(\eta_{ij})]$. The main advantage of the BLUP is that it utilizes information about clinic and

treatment-clinic variance to obtained more refined predictions of the likelihood of favorable outcomes than would be obtained from naive point estimates. For example, for clinic 5, treatment 2, 0 favorable outcomes are observed out of 10 subjects. However, because of the relatively low number of subjects, 0 is not a reasonable $\hat{\pi}_{ij}$; the BLUP of 0.14, however, is reasonable.

Table 4 gives the estimated variance components, the estimated $\pi_{i.}$ (the likelihood of a favorable outcome for the i^{th} treatment over the entire population of clinics) and the Wald statistic to test the equality of $\pi_{i.}$ for the 2 treatments. These were computed for the identity as well as the probit link functions. The results obtained by Beitler and Landis (1985) are given as well for comparison purposes. Their results are based on modelling π_{ij} directly, i.e. $\pi_{ij} = m + \tau_i + c_j + \tau c_{ij}$, and is equivalent to the GLMM with an identity link. However, they used a method of moments estimate of the variance components and make somewhat different assumptions about the error variance. In essence, they assume that error variance is constant across treatment, whereas the GLMM assumes that the conditional error variance is unique for each treatment-clinic combination. The results are similar except that the Wald statistic for the GLMM is much higher. A more systematic investigation would be needed to see if this pattern holds or is just an isolated occurrence. The GLMM with the logit link provides a much lower Wald statistic and different estimates of the $\pi_{i.}$. Model diagnostics as described by McCullagh and Nelder indicate that the logit link is a better model than the identity link and the results are thus more believable.

6. SUMMARY AND CONCLUSIONS

The generalized linear mixed model provides a unifying framework for linear models. Depending on one's perspective, it allows the extension of generalized linear models to accommodate random effects, or it allows the extension of mixed model methods to accommodate non-normal errors. The generalized linear model, the mixed model, and the traditional "general" linear model are all special cases of the GLMM.

Inference on the GLMM involves straightforward extension of methods used for generalized and mixed linear models. These methods can be used for models with correlated errors, non-scaler link functions, and, in principle, can be extended to more general distributions for random model effects.

However, the inference methods used depend on asymptotic properties whose small sample behavior is poorly understood at this time. This is especially true of inference involving *predictable* functions; the behavior of *estimable* functions is better understood. Much more study in this area is needed.

Finally, although the GLMM clearly extends the applicability of linear model methods far beyond traditional statistical practice, little work exists studying the implications for the design of experiments. Optimal theory of design is based almost exclusively on fixed effects, i.i.d. normal errors models. While it may be that many standard designs will prove to be optimal - or nearly so - for experiments for which the GLMM is appropriate, there is obviously no guarantee. This is clearly an area in need of a great deal more investigation.

References

Agresti, A. (1990) *Categorical Data Analysis*. New York: John Wiley and Sons.

Aitkin, M., Anderson, D., Francis, B. and Hinde, J. (1989) *Statistical modeling in GLIM*. New York: Oxford University Press.

Beitler, P.J. and Landis, J.R. (1985) A mixed effects model for categorical data. *Biometrics* 41:9.

Breslow, N.E. and Clayton, D.G. (1993) Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.* 88:9

Dobson, A.J. (1990) *An introduction to generalized linear models*. New York: Chapman and Hall.

Eisenhart, C. (1947) The assumptions underlying the analysis of variance. *Biometrics* 3:1.

Gilmour, A.R. (1985) The analysis of binomial data by a generalized linear mixed model. *Biometrika* 72:593.

Harville, D.A. (1976) Extension of the Gauss-Markov theorem to include the estimation of random effects. *Ann. Statist.* 2:384.

Harville, D.A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.* 72:320.

Harville, D.A. and Mee, R.W. (1984) A mixed-model procedure for analyzing ordered categorical data. *Biometrics* 31:423.

Henderson, C.R.(1984a) *Applications of mixed models for animal breeding.* University of Guelph, Canada.

Henderson, C.R. (1984b) ANOVA, MIVQUE, REML, and ML algorithms for estimation of variances and covariances. in *Statistics: an appraisal. Proceedings 50th anniversary conference, Iowa State Statistical Laboratory*, H.A. David and H.T. David, ed. Ames: Iowa State University Press.

Jeske, D.R. and Harville, D.A. (1988) Prediction-interval procedures and (fixed effects) confidence-interval procedures for mixed linear models. *Commun. Statist. - Theory Meth.* 17(4): 1053.

Kacker, R.N. and Harville, D.A. (1984). Approximations for standard errors of estimators for fixed and random effects in mixed linear models. *J. Amer. Statist. Assoc.* 79:853.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models, 2nd ed.* New York: Chapman and Hall.

McLean, R.A. and Sanders, W.L. (1988). Approximating degrees of freedom and standard errors in mixed linear models. *Proceedings of the Statistical Computing Section*, American Statistical Association, New Orleans, 50.

McLean, R.A., Sanders W.L., and Stroup, W.W. (1991) A Unified approach to mixed linear models. *Amer. Statist.* 45:54.

Miller, J.J. (1977) Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Ann. Statist.* 5:746.

Nelder, J.A. and Wedderburn, R.W.M. (1972) Generalised linear models. *J. R. Statist. Soc. A.* 135:370.

Robinson, G.K. (1991) That BLUP is a good thing: the estimation of random effects (with discussion). *Statist. Sci.* 6:15.

Searle, S.R. (1971) *Linear models*. New York: Wiley.

Searle S.R., Casella, G. and McCulloch, C.E. (1992) *Variance components*. New York: Wiley.

Snedecor, G.W. and Cochran, W.G. (1980) *Statistical methods, 7th ed.* Ames: Iowa State University Press.

Southern Regional Cooperative Series Bulletin 343. (1989) *Applications of mixed models to agriculture and related disciplines*. Louisiana State Experiment Station, Baton Rouge.

Steel, R.G.D. and Torrie, J.H. (1980) *Principle and procedures of statistics: a biometrical approach, 2nd ed.* New York: McGraw-Hill.

Vonesh, E.F. and Carter, R.L. (1992) Mixed-effects nonlinear regression for unbalanced repeated measures. *Biometrics* 48:1.

Zeger, S.L., Liang, K.Y. and Albert, P.S. (1988) Models for longitudinal data: a generalized estimating approach. *Biometrics* 44:1049.

Figure 1. Probit LInk Function

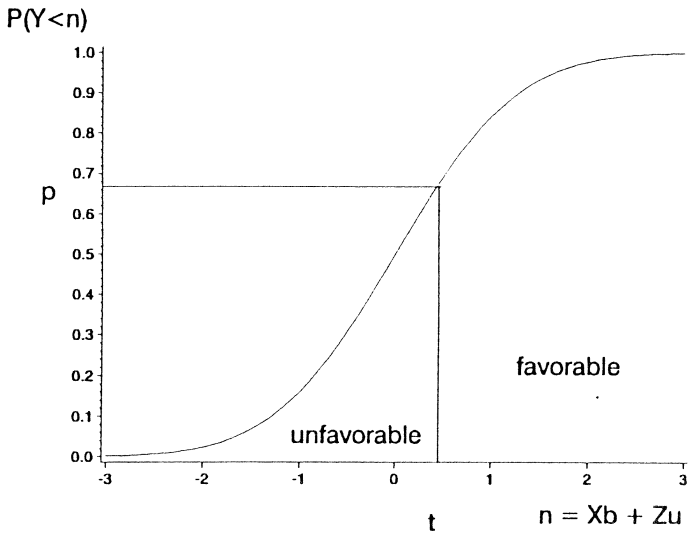

TABLE 1. Examples of common probability distributions from the exponential family

DISTRIBUTION	LOG - LIKELIHOOD	E(y)	Var(y)	α
Normal	$\frac{y'\mu - \mu'\mu}{\sigma^2} + \frac{n}{2}[\ln(2\pi\sigma^2)] - \frac{y'y}{2\sigma^2}$	μ	$I\sigma^2$	μ
Poisson	$y'[\ln(\lambda)] - 1'y - 1'[\ln(y!)]$	λ	diag (λ)	$\ln(\lambda)$
Binomial	$y'\left[\ln\left(\frac{p}{1-p}\right)\right] + n'[\ln(1-p)] + 1'\left[\ln\binom{n}{p}\right]$	p	diag[p(1 - p)]	$\ln\left(\frac{p}{1-p}\right)$

Table 2. Data and BLUP's using GLMM with probit link for Beitler & Landis example

Clinic	Treatment	N	Y (# fav)	GLMM - BLUP
1	1	36	11	.35
	2	37	10	.23
2	1	20	16	.80
	2	32	22	.66
3	1	19	14	.65
	2	19	7	.43
4	1	16	2	.15
	2	17	1	.08
5	1	17	6	.28
	2	12	0	.14
6	1	11	1	.12
	2	10	0	.06
7	1	5	1	.24
	2	9	1	.13
8	1	6	4	.77
	2	7	6	.64

Table 3. Element of the generalized linear mixed model for Beitler & Landis data, logit link

Distribution: Binomial

$$\textit{Link} \qquad \textit{logit} = \ln\left(\frac{\pi}{1-\pi}\right)$$

$$\textit{InverseLink}: h(\eta) = \frac{\exp(\eta)}{1+\exp(\eta)}$$

$$\textit{Model} \qquad \eta = \ln\left(\frac{\pi}{1-\pi}\right) = m + \tau_i + c_j + \tau c_{ij}$$

where $c_j \sim N(0,\sigma_C^2)$; and $\tau c_{ij} \sim N(0,\sigma_{TC}^2)$

$$R = diag\left[\frac{\pi(1-\pi)}{n}\right]$$

$$H = diag\left[\frac{\partial h(\eta)}{\partial \eta}\right] = diag\left[\pi(1-\pi)\right]$$

$$G = \begin{bmatrix} I_C\sigma_C^2 & 0 \\ 0 & I_{TC}\sigma_{TC}^2 \end{bmatrix}$$

Table 4. Estimates and test results for Beitler & Landis example

	Beitler & Landis Method	GLMM Identity Link	GLMM Logit Link
BLUE - π_1	0.42	0.41	0.40
BLUE - π_2	0.29	0.28	0.24
Wald χ^2 (H_0: $\pi_1 = \pi_2$)	5.35	7.06	5.14
σ_C^2	0.0700[1]	0.0745[2]	1.9932[3]
σ_{TC}^2	0.0019[1]	0.0004[2]	0.0491[3]

[1] Identity link, method of moments estimate

[2] Identity link, REML-like estimate

[3] Logit link, REML-like estimate