MODELS FOR PROPORTIONS IN A TWO-WAY CROSS

CLASSIFICATION WITHOUT

INTERACTION

By

ANDRONIKOS MAUROMOUSTAKOS

Bachelor of Science
Oral Roberts University
Tulsa, Oklahoma
1982

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1984

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
May, 1989

MODELS FOR PROPORTIONS IN A TWO-WAY CROSS

CLASSIFICATION WITHOUT

INTERACTION

Thesis Approved:

Ronald W. McNew

Thesis Adviser

U M Jarts

David F. Weeks

Michael Smith

Norman N. Durham

Dean of the Graduate College

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Statement of Problem

In many research studies the response variable has the form of a binomial random variable as a result of summing independent binary observations. In such situations an analysis can involve procedures based on the binomial distribution. In most of these cases, the observed proportions display greater variability than that predicted by the binomial model. In these instances, which are described as exhibiting "overdispersion", the additional variation makes the binomial maximum likelihood procedure inappropriate. There are several causes of overdispersion. In survey situations, sampled populations tend to be clustered. In biological experiments utilizing families, related individuals will tend to have more similar responses than unrelated individuals. In clinical trials, repeated observations on the same individual through time tend to be correlated. In general, the overdispersion could be attributed to any unmeasured source of experimental error that affects a group of binary observations as a whole.

In this study, we will consider two-way cross-classified data that arise in the form of proportions (based on unequal numbers of binary observations per cell) that exhibit overdispersion. In particular we are going to study the following two cases: both factors are random, and one factor is random and the other fixed. The models developed here will be extensions of models developed for one-way classifications and nested classifications.

The literature is reviewed in Chapter II, beginning with early

1

methodology and parametric models that attempted to account for the extra binomial variation, such as the beta-binomial model. The remainder deals with heteroscedastic models, random logistic models, quasi-likelihood models and empirically weighted least squares.

Williams (1982) proposed a model to account for extra-binomial variation. In Chapter III models are considered which include overdispersion but extend his approach to incorporate the variation associated with random effects. These models will be referred to as *Type I models*.

In Chapter IV procedures such as fitting random logistic models are studied using the approach taken by Hinde (1982) in which random effects for overdispersion enter the model on the same scale of additivity as the fixed effects. He used a mixture distribution that he referred to as Poisson-Normal for analyzing count data by a log-linear model that exhibited extra-Poisson variation. The extensions investigated in this study will accommodate random effects in addition to an overdispersion parameter. In all of the above models, only those transformations in which additivity is achieved in the logit scale are considered. Using generalized linear model terminology, we study only the canonical link function for the binomial distribution. These models will be referred to as *Type II models*.

Finally, in Chapter V an attempt is made to compare the models developed in Chapters III and IV. This will involve comparisons of the estimates, why and how the models differ, properties of the estimates, and the advantages and disadvantages of each model.

In addition, the proposed models are applied to real data based on a genetic study of survival of Virginia Pine seedlings, given in Appendix A. Results based on simulations are reported. The calculations required by both types of models will be compared. A listing of the SAS programs for

fitting the models discussed in Chapters III and IV are provided in Appendices B, C and D. Suggestions will be made concerning further research.

CHAPTER II

LITERATURE REVIEW

Early Methodology

Several authors have discussed the analysis of proportions from many different viewpoints. See, Cox (1970) and Bishop, Fienberg, and Holland (1975) for extensive discussions. The oldest approach probably goes back to Fisher's maximum likelihood method based on the binomial distribution. Cochran (1954), Anscombe (1954) and Bartlett (1954) suggested (in a discussion of a paper by Fisher (1954)) the transformations which have become standard for analyzing binomial data. Their approaches use analysis of variance methods after a transformation. The validity of these approaches relies on the robustness of the ANOVA methodology, particularly for "balanced situations". However, transforming the response alone may not result in an appropriate model because the transformation simultaneously tries to achieve normality, constant variance, and additivity of the effects of the explanatory variables.

As was mentioned previously, data in the form of counts usually display greater variability than what a binomial or Poisson model would predict. Thus, the maximum likelihood method is inappropriate. Cox (1983) stated that even if overdispersion is not a component of the parametric model, maximum likelihood estimates still retain high efficiency for modest amounts of overdispersion. Most of the literature on extra-binomial variation investigates the cases of the one way classification (in the form of comparisons of treatments) or nested classifications (in the form of

4

two-stage sampling), rather than cross classifications.

Cochran (1943), was the first to discuss the problem of analyzing proportions in an unbalanced case, and noted the possibility that the true probabilities of observations may vary. Such variation will increase the variance of observed proportions which are independent. He then investigated the efficiencies of equal, binomial, and partial weighting, the latter being a combination of the other two. He concluded that partial weighting is more efficient when the extra variation is between .30 and .80. Finney (1971) realized that mistakenly treating proportions as being binomially distributed would ordinarily lead to unbiased parameter estimates, but the associated standard errors would be underestimated, perhaps thereby leading an investigator into unwarranted conclusions. Thus, he proposed a simple way of correcting the problem by scaling the covariance matrix of the parameter estimates by the factor $\frac{X^2}{(n-k)}$ where n is the sample size, k is the number of parameters and X^2 is the Pearson Chi-Square statistic.

Parametric Models

Skellam (1948), in dealing with genetic data, used the beta-binomial as a model and he derived method of moments and maximum likelihood estimates. Several subsequent papers focused on the beta-binomial distribution. This is the distribution when one assumes the beta distribution as the sampling distribution of the unknown binomial parameters. There are several advantages to using such a model. First, the beta distribution is one of the richest distributions on the zero-one interval in terms of different possible shapes depending on the choices of the parameters. Second, the beta distribution is the conjugate prior for the binomial distribution which means that the resulting marginal distribution (beta-binomial) remains in the exponential family. Several techniques exist that give rise to

estimates of parameters via the method of moments or maximum likelihood for the exponential family.

Mosimann (1962), Griffiths (1973), and Brier (1980) investigated some of the properties of the above distribution. They showed that under beta-binomial sampling, the usual Pearson and likelihood ratio statistics are asymptotically multiples of chi-square random variables. Williams (1975) analyzed binary response data from a completely randomized experiment in which the experimental units were animal litters. He described a parametric approach based on the beta-binomial model in which the parameters of the beta distribution were estimated using maximum likelihood and treatment differences were tested using asymptotic likelihood ratio tests. Crowder (1978) considered a regression of proportions where the beta-binomial distribution was used as the error distribution. His linearized beta-binomial model generalized logistic regression to account for random effects.

The beta-binomial model has been applied to analyzing a variety of data from many disciplines. In teratological studies (Williams (1975), Paul (1982), Shirley and Hickling (1981)), found that abnormalities of animals in the same litter closely follow beta-binomial distributions instead of the simple binomial distribution. The implication is that animals from the same litter were found to have closer associations with regard to their responses to treatment. Griffiths (1973) applied the beta-binomial model in studying the incidence of non-infectious diseases in households. It was also considered by Chatfield and Goodhardt (1970), when investigating consumer purchasing behavior. Edwards (1958) applied the model to data on the distributions of boys and girls in families in Saxony for the period 1876-1885. Otake and Prentice (1984) analyzed data based

on studies of atomic bomb survivors. They generalized previous work by a allowing an extra binomial type variation in the aberrant cell counts corresponding to within subject correlations in respect to cell aberrations. If the distributional assumptions made by these parametric models are correct then the resulting estimates are maximum likelihood estimates, and from classical theory are known to be consistent, asymptotically normal and efficient.

Smith (1983) documented a FORTRAN subroutine that calculates maximum likelihood estimates of the parameters of the beta-binomial distribution and their standard errors. The procedure is an iterative Newton-Raphson technique that uses moment estimates as initial values.

Paul (1982) compared the beta-binomial model to other competing models (parametric and nonparametric). Some of these models were: the correlated binomial model of Kupper and Haseman (1978), the multiplicative binomial model of Altham (1978), and Gladen's (1979) model. The latter model used jackknife estimates of means and standard deviations. Paul applied all these models to data from toxicology, and also compared them via simulations based on all of the above models. He concluded from his simulation results that the beta-binomial model is more sensitive to departures from the binomial models than the other models considered and therefore superior for his data and data in similar fields. If one were to draw analogies to the usual analysis of variance models, the beta binomial model would be analogous to the random effects model with uncorrelated errors.

Quasi-Likelihood and Generalized Linear Models

Let R_{ij} be independent binomial random variables, where $i=1,...,I$, and $j=1,...,J$, with mean $n_{ij}\pi_{ij}$ and variance $n_{ij}\pi_{ij}(1-\pi_{ij})$. The method that allows estimation of the mean and also accommodates the existence and the

estimation of a variance parameter in a simplest form was introduced as a quasi-likelihood by Wedderburn (1974). This method formalized Finney's (1971) suggestion and extended the original theory of Generalized Linear Models by Nelder and Wedderburn (1972). The extension to the original theory was that knowledge of the full likelihood is no longer necessary; rather knowledge of the first two moments is sufficient, if the variance is a known function of the mean. In its simplest form this quasi-likelihood methodology models the variance of $Y_{ij} = \dfrac{R_{ij}}{n_{ij}}$ as

$$\phi \left[\frac{\pi_{ij}(1-\pi_{ij})}{n_{ij}} \right].$$

This variance does not correspond to a specific distribution although its form looks very similar to the variance function of the beta-binomial distribution. Nevertheless, a method very similar to maximum likelihood can be applied to obtain estimates of π_{ij} and ϕ. If $\phi = 1$ then the estimation procedure becomes the ordinary maximum likelihood estimation. McCullagh (1983) further developed this quasi-likelihood theory and proved some asymptotic optimality properties for the so called quasi-likelihood estimators. Fahrmeir and Kaufmann (1985) showed that the general conditions required for consistency and asymptotic normality of the maximum likelihood estimators in the generalized linear models reduce to weak requirements for special exponential families. The quasi-likelihood method for overdispersion may also be considered as one of the methods that requires only information of the first two moments and the ability to express the variance in terms of the mean. These models are fitted by iterative least squares which is reviewed in the subsequent sections. These methods provided researchers

with alternative methods of dealing with overdispersed data. Cox (1983) reports that the order of efficiency of the quasi-likelihood method was high for moderate amounts of overdispersion. Nelder (1985) and Nelder and Pregibon (1987) extended the quasi-likelihood definition to allow the comparisons of variance functions as well as those of linear predictors and link functions. Jorgensen (1983) extended the class of generalized linear models to allow for correlated observations. Firth (1987) suggested that quasi-likelihood estimates retain fairly high efficiency under "moderate" departures from the corresponding natural exponential family. One class of models that he investigated was overdispersion models relative to some exponential family. Crowder (1987) gives three examples that highlight particular pitfalls in maximum quasi-likelihood estimation. Hill and Tsai (1988) suggested that maximum quasi-likelihood estimation is a strong alternative to fully efficient maximum likelihood estimation. They showed that in fact transformation methods were less efficient than the general quasi-likelihood methods when models are mis-specified, and that maximum quasi-likelihood estimates and their standard errors were much easier to calculate than those for maximum likelihood estimates.

<center>Heteroscedastic and Other Related Models</center>

When a logistic regression model fits the data poorly there are several alternatives to consider. Using the generalized linear models formulation, these alternatives can be divided into either link function modifications or frequency distribution modifications. Since our study involves the second type, we are not going to review research methodology concerning link function modifications. It is worth while to mention the work of Aitkin (1987), and Smyth (1986), who in an attempt to model possible variance heterogeneity, describe a method of fitting overdispersed data for which the

amount of overdispersion depends on the same covariates as the means. Their approach considers mean and dispersion structure separately. Mean and dispersion submodels are formulated and both are essentially generalized linear models themselves. The deviance components of the mean sub-models are used as the dependent variable for the dispersion sub-model. Their results can be extended to discrete models by using quasi-likelihood models such as those of Nelder and Pregibon (1987).

Within the context of the maximum likelihood methods that are available, Laird (1978) gave an approach to the problem of estimating fixed effects in a mixed effects model. An empirical Bayes approach is used to combine the log-linear model with normal prior distributions, thereby obtaining estimates of the contingency table cell probabilities. The model proposed by Laird is given by

$$\log\left[\pi_{ij}\right] = u_0 + u_{1(i)} + u_{2(j)} + u_{12(ij)}$$

where $u_{1(i)}$ is the main effect of the ith level of the row variable, $u_{2(j)}$ is the main effect of the jth level of the column variable, and $u_{12(ij)}$ is the interaction. The estimate of σ^2 (the variance of $u_{12(ij)}$) is obtained by integrating out $\underset{\sim}{u}$ and using the marginal likelihood of σ^2 given $\underset{\sim}{x}$ (the observed cell frequencies). The derivation of the likelihood equation and second derivative is accomplished by viewing $\underset{\sim}{u}$ as missing data and treating the estimation of σ^2 as involving incomplete data. An iterative algorithm, called the EM algorithm, is used. Each iteration of this algorithm involves an expectation step (E) and a maximization step (M). Laird notes that due to the rich variety of distributional assumptions one can make concerning the

observed data, variance components models for categorical data are more complex than their quantitative counterparts. Leonard (1975) employed a two-stage Bayesian approach for the two-way table where, instead of assuming flat priors for the rows, columns, and their interaction in Laird's model, he assumed normal priors for those effects. Laird and Ware (1982) used two-stage random effects models for the analysis of longitudinal data where one must account for the relationship between serial observations on the same unit. Stiratelli, Laird and Ware (1984) presented a general mixed model for the analysis of serial dichotomous responses provided by a study of participants. Each subject's serial responses are assumed to arise from a logistic model but with regression coefficients that vary between subjects. The logistic regression parameters are assumed to be normally distributed in the population. Inference is based upon maximum likelihood estimation of fixed effects and variance components, and empirical Bayes estimation of random effects. Approximate solutions based on the mode of the posterior distribution of the random parameters is proposed and is implemented by means of the EM algorithm.

Moser (1985) considered analysis procedures for the balanced randomized complete block design with binary observations. He considered the block effects as random effects. He discussed a procedure for getting approximate maximum likelihood estimates after integrating out the random effects and provided an approximate method for obtaining standard errors of these estimates using a resampling type methodology. He also studied the asymptotic distribution of the linear form of the logit model and provided an alternate procedure for approximating and testing the unknown parameters using ordinary least squares.

Haseman and Kupper (1979) reviewed the literature and catalogued the

models, estimators, and approximations that had been used up to that time to analyze data with extra-binomial variation. They concluded that further research is needed.

Kempthorne and Koch (1983), investigated partially structured variations of a complete two-way random cross classification design. This methodology which is based on infinite sampling arguments, allows the estimation of the mean response, among-row correlation coefficient, among-column correlation coefficient, and the within all cell correlation coefficient as well as their standard errors.

Harville and Mee (1984) developed a mixed model procedure for predicting the value of an ordered categorical response from knowledge of various predictor variables. Their approach is based on the mixed model version of the threshold model in which it is assumed that the observed category is determined by the value of an underlying unobserved continuous response that follows a mixed linear model.

Beitler and Landis (1985) proposed a mixed model for categorical data from unbalanced designs which is directly analogous to a two-way ANOVA model for quantitative data. An extension of the fitting constants method is developed to estimate variance components based on appropriate reductions in sums of squares. The resulting variance component estimators are incorporated into Wald statistic to test for treatment differences. The weighted least squares methodology of Grizzle, Stamner and Koch (1969) is employed to calculate the overall estimates of treatment means and Wald statistics. Kempthorne (1982), investigated the analysis of clustered attribute data from nested and classification designs using random effects models.

Empirically Iterative Reweighted Least Squares
(EIRWLS)

Unlike the parametric methods mentioned above, EIRWLS methods do not specify the form of the distribution. All the methods in this category involve some kind of approximation and their differences and similarities are not very clear. Overdispersion as discussed above can be attributed to the existence of two levels of randomness. In the first level, the observations conditioned on unknown parameters are regarded as random variables. In the second level, these unknown parameters, p_{ij}, are regarded as random variables from some distribution. Usually, at the first level, these models assume independent Bernoulli random variables conditional on the success probability for the specific cell. That is,

$$R_{ij} | P_{ij} = p_{ij} \sim \text{Binomial}\left[n_{ij}, p_{ij} \right]$$

where $i = 1, \ldots, a$ indexes the rows, $j = 1, \ldots, b$ indexes the columns, and P_{ij} is itself a random variable with

$$E\left[P_{ij} \right] = \pi_{ij}$$

and

$$V\left[P_{ij} \right] = \phi \, \pi_{ij} \left[1 - \pi_{ij} \right]. \tag{2.1}$$

where $0 \le \phi \le 1$. Then the unconditional mean and variance of R_{ij} are given by

$$E\left[R_{ij} \right] = n_{ij} \pi_{ij}$$

and

$$V\left[R_{ij}\right] = n_{ij}\,\pi_{ij}\left[1-\pi_{ij}\right]\left[1+\left[n_{ij}-1\right]\phi\right].$$

(2.2)

Thus, the marginal variance of R_{ij} is made up of the two components, the binomial variance and the extra-binomial component. If $\phi=0$, then the variance reduces to the binomial variance. The variance (2.2) could also result from a model in which ϕ is the pair-wise correlation of the binary variables whose sum is R_{ij}. In such a formulation, negative values of ϕ could be permissible, which would lead to "underdispersion." In the present formulation, ϕ is introduced in the variance of P_{ij} and, therefore, is necessarily non-negative.

Kleinman (1973) was the first since Cochran (1943), to consider the above formulation in the single sample and one-way classification cases using empirically weighted least squares. He used method-of-moments estimation for obtaining estimates of the extraneous variance and used them as weights in a least squares analysis. He also provided results based on Monte Carlo simulations in which he studied the performance of the empirical weighting scheme in moderate and small samples. He concluded that it performed adequately for practical application purposes. Furthermore, he found that his empirical weights had high efficiency relative to exact least squares estimates (similar to the later findings of Cox (1983)). His method is equivalent to Finney's for the equal sample size case. Kleinman (1975) extended the model to a two-way classification for the comparison of two drugs in a random sample of clinics from an assumed infinite population of hospitals. Within hospital i, n_{ij} patients are randomly assigned to drug j. After a specific period, the number of patients, r_{ij}, in hospital i cured after receiving drug j is observed. Thus, the hospital effects are random and the drug effects are fixed. Again, he used the same empirical weighting

procedure that he proposed in 1973. Crowder (1979) proposed likelihood inference procedures for the ϕ in (2.2) above that he refers to as intraclass correlation in a one-way classification setup.

Pierce and Sands (1975) translated the suggestion by Finney (1971) to allow for extra binomial variance in the logit setting in which the overdispersion is additive in the logit scale. Also Pierce (1976) discussed his random effects model with normal perturbations on the logit scale as it applied to matched pairs of binomial data. Williams (1982) further developed Kleinman's approach for logistic linear models. He gave an algorithm in GLIM that produced a method of moments estimate of the overdispersion parameter ϕ using iterative reweighted least squares for the one-way classification. GLIM is a well known statistical package that allows one to interactively fit and test linear models with various error distributions. In GLIM a regression type analysis is easily performed on discrete data as long as the data is assumed to be free of random effects. Williams also discussed how his methods compared with the approach of Pierce and Sands. Breslow (1984), following the suggestion of Williams, implemented a similar algorithm for log-linear models with extra-Poisson variation.

Crowder (1985) discussed further the mean/variance structure of correlated binomial data and suggested using Whittle's (1961) Gaussian estimation as a useful general method. He also compared his method to many other existing method-of-moment estimation methods for the same setting. Moore (1987) discussed extensions of both the beta-binomial and method-of-moments methods to allow modeling the extra binomial component of the variance as a function of the mean. He also applied these methods to data on chromosomal aberrations in survivors of the atomic bombing in Hiroshima. Moore (1986) discussed consistency and asymptotic normality of moment

estimates of the regression parameters and the overdispersion parameter for enumerative data with extra variation. He determined the asymptotic variance of the overdispersion parameter.

Stirling (1984) extended the Nelder and Wedderburn (1972) method of maximum likelihood estimation of the parameters in an exponential family of regression models to models that included cases such as the negative binomial and beta-binomial distributions. He showed that Iterative Re-weighted Least Squares converges considerably faster than the EM algorithm of Dempster, Laird and Rubin (1977). Green (1984) has written the most extensive paper on EIRWLS. He included many references on the scope of applications of iterative reweighted least squares. He also extended beyond the exponential-type family of generalized linear models (GLM's) to other distributions, to non-linear parameterizations and to dependent observations. The algorithms for fitting such models are shown to be numerically stable, highly suited for interactive computations, and easily programmed.

Random Logistic Models

In random logistic models, the extra variation is modeled as an additive random effect on the same scale as the fixed effects or covariates. The scale most often used with proportions is the logit scale. This scale of additivity, or form of the link function as known in the context of GLM's, will not be discussed much further. Our studies are going to be centered around the logit link since as pointed out by Tornqvist, Vartia, and Vartia (1985) it is the only symmetric, additive, normed indicator of relative change. Further discussions on the logit link function can be found in Berkson (1951).

Hinde (1982), assuming normally distributed errors on the log scale for

independent counts, used GLIM to maximize a compound Poisson likelihood. His technique seems powerful and the computations could be used in random log-linear and random logistic models if the link function used is the canonical link for the distribution considered. Brillinger and Preisler (1983) considered a similar model for count data. As in Hinde's model latent variates which cannot be measured directly play an essential role in the description of the observed quantities. Maximum likelihood estimation of the parameters and their properties are discussed and shown to be a viable approach to a broad class of latent variable models. They show that GLIM provides an effective tool for carrying out the needed computations. The methodology is discussed in the context of analyzing radioactive counts from a nuclear medicine experiment.

Gilmour, Anderson and Rae (1985) considered estimating the variance of the random effects for a set of independent binomial proportions after conditioning on the random effects. Additivity of the random effects in their case is achieved on the probit scale. Im and Giannola (1988) discussed a derivative free maximization algorithm as an alternative to the EM algorithm in computing maximum likelihood estimates in mixed probit and logit models with binomial data. They investigated predicting the random effects and estimating the fixed effects and variance components in a two-way nested layout. The methodology developed is used in analyzing mortality data.

Anderson and Aitkin (1985) in the context of survey design studied interviewer variability in a binary response as another example of variance component estimation in a non-normal family. Stokes (1988), showed in the same setting that the loss of precision in the estimates of means due to the variability among interviewers can be substantial. She also showed how parameters from a model of variance components in binary variables are

related to the increased variance in the population estimates.

CHAPTER III

TYPE I MODELS

Introduction

Consider data from a two-way classification with a and b levels of the two factors (Figure 1). Let Y_{ijk} denote the binary observation on the kth unit, $k=1,...n_{ij}$ in row i and column j. The observed frequency for the ij^{th} cell is $R_{ij} = \sum_{k=1}^{n_{ij}} Y_{ijk}$.

	1	2	....	b
1	R_{11}	R_{12}		R_{1b}
2	R_{21}	R_{22}		R_{2b}
⋮				
a	R_{a1}	R_{a2}		R_{ab}

Figure 1. Two-Way Setup

Our objective is to extend Williams' approach for dealing with overdispersion in regression to the two-way classification in which one or both factors are random. The variance of R_{ij} using the binomial distribution is given by $V(R_{ij}) = n_{ij}p_{ij}(1-p_{ij})$. To account for the fact that the variance of R_{ij} is very often greater than $n_{ij}p_{ij}(1-p_{ij})$ investigators have considered approaches that fall into two different categories as far as generalizing the logistic regression model.

The first of these two approaches to be considered is that of Williams (1982) who regards the logistic probability functions as unobservable random variables. The other approach is the "random" logistic regression first studied by Pierce and Sands (1975) in which random effects are additive on the logit scale. Williams' approach is a generalization of both the beta-binomial approach and McCullagh and Nelder's GLM approach to modeling discrete data.

For the remainder of this chapter, we will review Williams' approach. We will then extend his model to the case where both factors are random and to the case where one factor is fixed and the other is random. For both cases, we propose method-of-moments estimators for the unknown parameters. We will also provide SAS code that implements the proposed methodology for both cases (mixed and random). The variances of the estimates are addressed using the results of extensive computer simulations. They seem to suggest that the technique could be implemented in any of the standard statistical packages such as SAS, GLIM and probably others.

Review of Williams' Approach

Williams assumed that the ith response $(1 \leq i \leq N)$ is a count of R_i successes and $n_i - R_i$ failures. Associated with this response are the values $x_{i1}, x_{i2}, \ldots, x_{ip}$ of p explanatory variables, where the $N \times p$ matrix X is of rank

p. The ordinary logistic linear model assumes that the R_i are independently distributed Binomial(n_i, θ_i), where

$$\theta_i = \frac{e^{\sum_s x_{is}\beta_s}}{1 + e^{\sum_s x_{is}\beta_s}}. \tag{3.1}$$

To allow for extra binomial variation he introduced unobserved continuous variables P_i, $i=1,\ldots,N$, independently distributed on $(0,1)$ with $E(P_i)=\theta_i$, and $V(P_i) = \phi\,\theta_i(1-\theta_i)$, $(0 \le \phi \le 1)$ and assumed that R_i conditional on $P_i=p_i$ is distributed Binomial(n_i, p_i). This means that unconditionally $E(R_i)=n_i\theta_i$ and $V(R_i)=\dfrac{v_i}{w_i}$ where $v_i=n_i\theta_i(1-\theta_i)$, and $\dfrac{1}{w_i}=1+\phi(n_i-1)$. He proceeded by saying that, although maximum likelihood estimation cannot be used since the distribution of the R_i is not fully specified, the relationship between the expectation and the variance of R_i allows the definition of a quasi-likelihood which is maximized with respect to the parameters $\underset{\sim}{\beta}$ via iterative use of the weighted least squares equations

$$X'W\,V^*X\,\hat{\underset{\sim}{\beta}} = X'W\,V^*Y^*. \tag{3.2}$$

Here θ_i^*, v_i^* are the values of θ_i, v_i corresponding to $\underset{\sim}{\beta}=\underset{\sim}{\beta}^*$, $V^*=\mathrm{diag}(v_i^*)$, $W=\mathrm{diag}(w_i)$ and $y_i^*=\sum_s x_{is}\beta_s^* + \dfrac{R_i - m_i\theta_i^*}{v_i^*}$.

The weights w_i needed in (3.2) depend on ϕ which is usually unknown. If the weights w_i are calculated from an initial estimate of ϕ and $\underset{\sim}{\beta}$ is estimated iteratively from (3.2) then the goodness of fit statistic

$$X^2 = \sum_{i=1}^{N} w_i \frac{(R_i - m_i \hat{\theta}_i)^2}{n_i \hat{\theta}_i (1 - \hat{\theta}_i)} \tag{3.3}$$

is approximately the weighted sum of squares of residuals $(Y^* - X\hat{\beta})'WV^*(Y^* - X\hat{\beta})$. Williams proposed using a method-of-moments estimator for ϕ based on the following

$$E(X^2) = \sum_{i=1}^{N} \left[w_i(1 - w_i v_i^* q_i)\left[1 + \phi(n_i - 1)\right] \right] \tag{3.4}$$

(when $W \neq I$), where q_i is the ith diagonal element of the variance of $X\hat{\beta}$ which is given by $X(X'WV^*X)^{-1}X'$. The proof of (3.4) is sketched below and depends on the following two results from linear models.

Let Y be a random vector with $E(Y) = X\beta$ and $Cov(Y) = \Sigma$. Let W be a constant symmetric weight matrix not necessarily equal to Σ^{-1}. Then if $\hat{Y} = X(X'WX)^{-1}X'WY$ the following holds

$$E\left[(Y - \hat{Y})'W(Y - \hat{Y})\right] = tr\left[I - \left\{W^{\frac{1}{2}}X\left[X'WX\right]^{-1}X'W^{\frac{1}{2}}\right\}W^{\frac{1}{2}}\Sigma W^{\frac{1}{2}}\right] \tag{3.5}$$

Thus if W and Σ in (3.5) above are diagonal with $W = diag(w_i)$ and $\Sigma = diag(\sigma_i)$ then

$$E\left[(Y - \hat{Y})'W(Y - \hat{Y})\right] = \sum_i w_i \sigma_i \left[1 - q_i w_i\right]$$

where q_i is the ith diagonal element of $X\left[X'WX\right]^{-1}X'$. This result leads to

the following.

Let $\underset{\sim}{Y}^*$ be a random vector with $E(\underset{\sim}{Y}^*)=X\underset{\sim}{\beta}$ and $Cov(\underset{\sim}{Y}^*)=(D'V^{-1}D)^{-1}$ where $D=diag(d_i)$ and $V=diag(v_i)$ are diagonal matrices. Then

$$E\left[(\underset{\sim}{Y}-\hat{\underset{\sim}{Y}}^*)'W\,(\underset{\sim}{Y}-\hat{\underset{\sim}{Y}}^*)\right] = \sum_i w_i\frac{v_i}{d_i^2}\left[1-q_iw_i\right].\qquad(3.6)$$

This completes the sketch of the proof of (3.4). Now, by equating (3.3) and (3.4), and solving for ϕ we obtain

$$\hat{\phi} = \frac{X^2-\sum_i\left[w_i(1-w_iv_i^*q_i)\right]}{\sum_i\left[w_i(n_i-1)(1-w_iv_i^*q_i)\right]}.\qquad(3.7)$$

Williams concluded by proposing the following procedure:

(1) Assume that $\phi=0$; estimate $\hat{\underset{\sim}{\beta}}$ and calculate X^2 by using ordinary logistic regression.

(2) Compare X^2 with the Chi square distribution with (N-p) d.f. If X^2 is unacceptably large, conclude that $\phi>0$ and calculate the estimate of ϕ by (3.7).

(3) Using new weights $w_i=\dfrac{1}{1+\hat{\phi}(n_i-1)}$ re-estimate $\hat{\underset{\sim}{\beta}}$ iteratively using (3.2), and re-calculate X^2.

(4) If X^2 is close to N-p, the estimate of ϕ is satisfactory. If not, then re-estimate ϕ as in (3.7) and return to Step (3).

He also gave an algorithm in GLIM for the procedure described above and applied it to some data given by Crowder (1978).

Extension of Williams' Methodology for the

Random Case

Consider the data layout shown in Figure 2 (a two-way random cross classification) assumed to have arisen as a result of a random sample of a and b levels of the two factors. Let $\theta_{1i} \overset{\mathrm{i\ i\ d}}{\sim} (\eta, \phi_1)$ for $i=1,...,a$, and $\theta_{2j} \overset{\mathrm{i\ i\ d}}{\sim} (\eta, \phi_2)$ for $j=1,...,b$, by which we mean that

$$E(\theta_{1i}) = E(\theta_{2j}) = \eta$$

and the variances are given by

$$V(\theta_{1i}) = \phi_1 \eta(1-\eta) \text{ and } V(\theta_{2j}) = \phi_2 \eta(1-\eta) \tag{3.8}$$

where ϕ_1 and $\phi_2 \in [0,1]$.

If we let $\theta_{ij}^* = \dfrac{\theta_{1i} + \theta_{2j}}{2}$, then θ_{ij}^* has expectation and variance given by

$$E(\theta_{ij}^*) = \eta \text{ and } V(\theta_{ij}^*) = \left[\frac{1}{4}(\phi_1 + \phi_2)\right]\eta(1-\eta).$$

The P_{ij} are independently distributed random variables on $(0,1)$ such that, given θ_{1i} and θ_{2j}, their conditional expectations and conditional variances are given by

$$E(P_{ij}|\theta_{1i}, \theta_{2j}) = \theta_{ij}^*,$$

and

$$V(P_{ij}|\theta_{1i}, \theta_{2j}) = \phi_3 \theta_{ij}^*(1-\theta_{ij}^*).$$

Let $R_{ij} = \overset{n_{ij}}{\underset{k=1}{\sum}} Y_{ijk}$ where Y_{ijk}'s, $k=1,...n_{ij}$, are conditionally independent binary observations given P_{ij}. Then $R_{ij}|P_{ij} = p_{ij} \sim \text{Binomial}(n_{ij}, p_{ij})$.

$$(\ \theta_2\ \text{represents the column random effect})$$

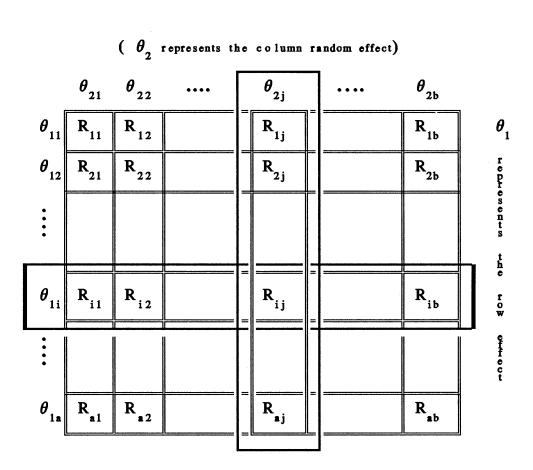|	θ_{21}	θ_{22}	....	θ_{2j}	....	θ_{2b}	
θ_{11}	R_{11}	R_{12}		R_{1j}		R_{1b}	θ_1
θ_{12}	R_{21}	R_{22}		R_{2j}		R_{2b}	
$\vdots$							
θ_{1i}	R_{i1}	R_{i2}		R_{ij}		R_{ib}	
$\vdots$							
θ_{1a}	R_{a1}	R_{a2}		R_{aj}		R_{ab}	

θ_1 represents the row effect

Figure 2. Two-Way Setup With Rows and Columns Random

Unconditionally, P_{ij} has mean η and variance

$$V(P_{ij}) = E(P_{ij}^2) - \left[E(P_{ij})\right]^2$$

$$= E_{\theta_{1i},\theta_{2j}}\left[E_{P_{ij}|\theta_{1i},\theta_{2j}}(P_{ij}^2|\theta_{1i},\theta_{2j})\right] - \eta^2$$

$$= E_{\theta_{1i},\theta_{2j}}\left[V_{P_{ij}|\theta_{1i},\theta_{2j}}(P_{ij}|\theta_{1i},\theta_{2j}) + \frac{1}{4}(\theta_{1i} + \theta_{2j})^2\right] - \eta^2$$

$$= E_{\theta_{1i},\theta_{2j}}\left[\phi_3\frac{1}{2}(\theta_{1i} + \theta_{2j}) - \phi_3\frac{1}{4}(\theta_{1i} + \theta_{2j})^2 + \frac{1}{4}(\theta_{1i} + \theta_{2j})^2\right] - \eta^2$$

$$= \phi_3\eta + \frac{1}{4}(1-\phi_3)\left[\eta^2 + \phi_1\eta(1-\eta) + 2\eta^2 + \eta^2 + \phi_2\eta(1-\eta)\right] - \eta^2$$

$$= \phi_3\eta + (1-\phi_3)\eta^2 + \left[\frac{1}{4}(1-\phi_3)(\phi_1+\phi_2)\right]\eta(1-\eta) - \eta^2$$

$$= \left[\phi_3 + \frac{1}{4}(\phi_1 + \phi_2) - \phi_3\frac{1}{4}(\phi_1 + \phi_2)\right]\eta(1-\eta). \tag{3.10}$$

In what follows we see that for a given i level (i.e., if we were to fix the row level) the variance of P_{ij} with respect to θ_{2j} given θ_{1i} can be expressed as some function of the ϕ_i's times its conditional mean times one minus its conditional mean. First the conditional mean of P_{ij}, given θ_{1i} with respect to θ_{2j}, is

$$E_{P_{ij}|\theta_{1i}}\left(P_{ij}|\theta_{1i}\right) = E_{\theta_{2j}}\left[E_{P_{ij}|\theta_{1i}\theta_{2j}}\left(P_{ij}|\theta_{1i},\theta_{2j}\right)\right]$$

$$=E_{\theta_{2j}}\left[\frac{1}{2}(\theta_{1i}+\theta_{2j})\right]$$

$$=\frac{1}{2}(\theta_{1i}+\eta). \tag{3.11}$$

If we let $\eta_{1i}=\frac{1}{2}(\theta_{1i}+\eta)$ then the conditional variance is given by

$$V\left[P_{ij}|\theta_{1i}\right] = E_{\theta_{2j}}\left[E_{P_{ij}|\theta_{1i}\theta_{2j}}\left[P_{ij}^2|\theta_{1i},\theta_{2j}\right]\right] - \eta_{1i}^2$$

$$=E_{\theta_{2j}}\left[\phi_3\theta_{ij}^*(1-\theta_{ij}^*)+(\theta_{ij}^*)^2\right] - \eta_{1i}^2, \qquad \left[\text{Note: } \theta_{ij}^*=\frac{\theta_{1i}+\theta_{2j}}{2}\right]$$

$$= E_{\theta_{2j}}\left[\phi_3\theta_{ij}^*\right] + (1-\phi_3)E_{\theta_{2j}}\left[\theta_{ij}^{*2}\right] - \eta_{1i}^2$$

$$= \phi_3\eta_{1i} + (1-\phi_3)\left[V_{\theta_{2j}}(\theta_{ij}^*) + \left[E_{\theta_{2j}}(\theta_{ij}^*)\right]^2\right] - \eta_{1i}^2$$

$$= \phi_3\eta_{1i} + (1-\phi_3)\left[\frac{1}{4}\phi_2\eta(1-\eta) + \frac{1}{4}(\theta_{1i}+\eta)^2\right] - \eta_{1i}^2$$

$$= \phi_3\eta_{1i} - \phi_3\eta_{1i}^2 + (1-\phi_3)\left[\frac{1}{4}\phi_2\eta(1-\eta)\right]$$

$$= \phi_3\eta_{1i}(1-\eta_{1i}) + (1-\phi_3)\left[\frac{1}{4}\phi_2\eta(1-\eta)\right]$$

$$= \left[\phi_3 + \frac{1}{4}(1-\phi_3)\phi_2 \frac{\eta(1-\eta)}{\eta_{1i}(1-\eta_{1i})}\right]\eta_{1i}(1-\eta_{1i}) \tag{3.12}$$

Thus, the average conditional variance is given by (recall that $\eta_{1i} = \theta_{1i} + \eta$)

$$E\left[V\left[P_{ij}|\theta_{1i}\right]\right] = \phi_3 \frac{1}{2} E_{\theta_{1i}}\left[\theta_{1i} + \eta - \frac{1}{2}(\theta_{1i}^2 + 2\theta_{1i}\eta + \eta^2)\right] + \frac{1}{4}\phi_2(1-\phi_3)\eta(1-\eta)$$

$$= \phi_3 \frac{1}{2}\left[2\eta - \frac{1}{2}\left[\phi_1\eta(1-\eta) + \eta^2 + 2\eta^2 + \eta^2\right]\right] + \frac{1}{4}\phi_2(1-\phi_3)\eta(1-\eta)$$

$$= \phi_3 \frac{1}{2}\left[2\eta(1-\eta) - \frac{1}{2}\phi_1\eta(1-\eta)\right] + \frac{1}{4}\phi_2(1-\phi_3)\eta(1-\eta)$$

$$= \left[\phi_3 - \frac{1}{4}\phi_3\phi_1 + \frac{1}{4}\phi_2 - \frac{1}{4}\phi_2\phi_3\right]\eta(1-\eta). \tag{3.13}$$

Therefore, we have shown that the average conditional expectation of P_{ij} for a fixed row divided by $\eta(1-\eta)$ equals $\left[\phi_3 - \frac{1}{4}\phi_3\phi_1 + \frac{1}{4}\phi_2 - \frac{1}{4}\phi_2\phi_3\right]$.

The corresponding result when considering the column is

$$\frac{E\left[V\left[P_{ij}|\theta_{2j}\right]\right]}{\eta(1-\eta)} = \left[\phi_3 - \frac{1}{4}\phi_3\phi_1 + \frac{1}{4}\phi_1 - \frac{1}{4}\phi_2\phi_3\right]. \tag{3.14}$$

The three formulas (3.10), (3.13) and (3.14) are the unconditional

variance of P_{ij}, and the average conditional variances when that we condition on the rows and columns, respectively.

We propose the following estimation procedure.

(1)   Fit a grand mean to the whole data set and get an estimate of the left hand side of (3.4) using the Williams one sample approach where his **X** matrix is nothing but an (ab $\times$ 1) vector of 1's. The estimate of the extra binomial variation given by (3.7) is thus our estimate, say C_1, for the left hand side of (3.10).

(2)   Consider each row of data values as shown in Figure 2 as a data set on its own (remember the discussion above on conditioning on each row...). Fit a grand row mean to each of these "datasets of b $\times$ 1 observations" using the same procedure as described in step (1). Note that the algorithm provides us with estimates of the means, $\hat{\eta}_{1i}$'s, and estimates of the extra binomial parameter, $\hat{\phi}_{1i}$'s, for each of the rows. Therefore a natural estimate of the average conditional variance given by (3.13) is a method of moments estimate, say C_2, given by

$$\frac{\widehat{E\left[V\left(P_{ij}\mid\theta_2\right)\right]}}{\eta(1-\eta)}=\left[\frac{\hat{\phi}_{11}\hat{\eta}_{11}(1-\hat{\eta}_{11})+\hat{\phi}_{12}\hat{\eta}_{12}(1-\hat{\eta}_{12})+\ldots+\hat{\phi}_{1a}\hat{\eta}_{1a}(1-\hat{\eta}_{1a})}{a\,\hat{\eta}(1-\hat{\eta})}\right]. \quad (3.15)$$

(3)   Repeat the procedure in (2) for each column. The b independent fits will produce estimates for the b column means, $\hat{\eta}_{2j}$'s, and the b estimates of extra binomial parameters $\hat{\phi}_{2j}$'s. Again a natural estimate, say C_3 of the left hand side of (3.14) would be given by

$$\frac{\widehat{E\left[V\left(P_{ij}\mid\theta_1\right)\right]}}{\eta(1-\eta)}=\left[\frac{\hat{\phi}_{21}\hat{\eta}_{21}(1-\hat{\eta}_{21})+\hat{\phi}_{22}\hat{\eta}_{22}(1-\hat{\eta}_{22})+\ldots+\hat{\phi}_{2a}\hat{\eta}_{1a}(1-\hat{\eta}_{2b})}{b\,\hat{\eta}(1-\hat{\eta})}\right]. \quad (3.16)$$

(4)   Having the estimates C_1, C_2 and C_3, we solve the following system of three equations for the three unknown parameters ϕ_1, ϕ_2 and ϕ_3

$$C_1 = \left\{ \phi_3 - \frac{1}{4}\,\phi_3\phi_1 + \frac{1}{4}\,\phi_2 - \frac{1}{4}\,\phi_2\phi_3 \right\},$$

$$C_2 = \left\{ \phi_3 - \frac{1}{4}\,\phi_3\phi_1 + \frac{1}{4}\,\phi_1 - \frac{1}{4}\,\phi_2\phi_3 \right\},$$

$$C_3 = \left\{ \phi_3 + \frac{1}{4}\,\phi_1 + \frac{1}{4}\,\phi_2 - \frac{1}{4}\,\phi_3\phi_1 - \frac{1}{4}\,\phi_2\phi_3 \right\},$$

and obtain the following estimates of ϕ's:

$$\hat{\phi}_1 = 4\left[C_3 - C_1 \right],$$

$$\hat{\phi}_2 = 4\left[C_3 - C_2 \right],$$

$$\hat{\phi}_3 = \left[\frac{C_1 - C_3 + C_2}{1 - \left[C_3 - C_1 \right] - \left[C_3 - C_2 \right]} \right].$$

SAS code for implementing the above steps is given in Appendix B. The variance of the proposed estimators is addressed next together with discussion on the process of generating data from the above model for the simulations. For calculating the variance and standard errors of the estimates we propose the following resampling procedure where we generate a given number, say N, of two way random tables in such a way that each table is based on the same values of ϕ_1, ϕ_2, and ϕ_3. For each table we

use the algorithm outlined above and obtain estimates of the ϕ's and η's. Then we calculate the variances and standard errors from these N estimates.

Although it was mentioned above that we only need the mean and variance of the θ_{1i}'s and θ_{2j}'s in our simulation program, we use the Beta distribution for the obvious reason that beta is a very flexible distribution on the (0,1) interval. Also it provides us with obvious candidates for η's and ϕ's since if $\theta \sim$ Beta(a,b) then we have that the

$$E(\theta) = \frac{a}{a+b} = \eta$$

and

$$V(\theta) = \frac{ab}{(a+b)^2(1+a+b)} = \phi\eta(1-\eta).$$

Adopting the following notation, where $\mathscr{B}(\eta,\phi)$ indicates a Beta distribution with mean η and variance $\phi\eta(1-\eta)$, we proceed as follows. Choose η, ϕ_1, ϕ_2 and generate a random sample of θ_{1i}'s for the rows such that $\theta_1 \sim$ Beta(η,ϕ_1) and similarly generate a random sample of θ_{2j}'s for the columns such that $\theta_{2j} \sim \mathscr{B}(\eta,\phi_2)$. Using SAS we can generate Gamma random variables. A Beta random variable can be constructed from two independent Gamma random variables using the following fact. If X_1, X_2 are independent Gamma random variables with shape parameters a, b, respectively then $\dfrac{X_1}{X_1 + X_2} \sim$ Beta(a,b). Because the means of θ_1 and θ_2 must be equal, a and b must be chosen so that (a/b) is constant. Having generated the θ_1's and θ_2's, we proceed with generating the P_{ij} in such a way that

$$P_{ij} \sim \mathscr{B}\left(\frac{\theta_{1i} + \theta_{2j}}{2}, \phi_3\right). \quad \text{Then for each cell we generate } n_{ij} \text{ to be Uniform}$$

discrete random variables (between 1 and 76) and for a given p_{ij}, we proceed by generating directly $Y_{ij} \sim$ Binomial(n_{ij}, p_{ij}) random variables.

The algorithm for the generation of N tables is sketched in Figure 3. The choices of the a and b shape parameters used were the pairs (2,2) for the rows and (3,3) for the columns. Therefore, the "true" overdispersion parameters for the rows were 0.2 and for the columns were 0.1429. The overall overdispersion was chosen to be 0.25.

Estimates of the parameters and their standard errors based on 100 (4x5) tables are presented in Table I. The first two rows in Table I correspond to the overall mean (η^{*}) and the estimate of the extra variation parameter from the overall fit (PE). The following rows represent the estimates of the means (M) and extra variations (EV) from the individual row or column fittings. They are labeled RIM, and RIEV (or CIM and CIEV) and represend the estimated values values from fitting the ith row (column) respectively. The last three rows of the table are the resulting estimates of the row, column, and overall ϕ's.

These results are compared to the procedures and methodologies developed in Chapter IV and will be discussed further in Chapter V. If the data do not possess significant amounts of overdispersion or if the number of levels of the random factor is small such as the case in our example in Table I (i.e., five rows and four columns), then the algorithm can lead to negative estimates of the ϕ_i's, especially ϕ_1 and ϕ_2. In our simulation, approximately forty percent of the $\hat{\phi_1}$'s and $\hat{\phi_2}$'s were negative but all $\hat{\phi_3}$'s were positive. This is similar to the occurrence of negative variance component estimates when dealing with continuous responses. The estimated variances of the $\hat{\phi}$'s in Table I drop significantly as the number of levels for that random effect increases.

```
TITLE 'GENERATION OF &N DATASETS W/ P1=.2, P2=.1429, AND P3=.25';
%LET NADS=DS;
%LET N=100;
%LET NR=5;
%LET NC=4;
DATA GALL;
  A=2;
  B=2;
  C=3;
  D=3;
  HTA=A/(A+B);
  ARRAY THETAI (5) THETAI1-THETAI5;
  ARRAY THETAJ (4) THETAJ1-THETAJ4;
  DO DSI=1 TO &N;
    DO I=1 TO 5;
       X1=RANGAM(SEED1,A);
       X2=RANGAM(SEED1,B);
       THETAI(I)=X1/(X1+X2);
       PHI1=1/(A+B+1);
    END;
    DO J=1 TO 4;
       X3=RANGAM(SEED2,C);
       X4=RANGAM(SEED2,D);
       THETAJ(J)=X3/(X3+X4);
       PHI2=1/(C+D+1);
    END;
        DO II=1 TO &NR;
           DO JI=1 TO &NC;
           PHI3=.25;
           EIJ=(THETAI(II)+THETAJ(JI))/2;
           ALPHA=((1-PHI3)/PHI3)*EIJ;
           BETA=((1-PHI3)/PHI3)*(1-EIJ);
           X5=RANGAM(SEED3,ALPHA);
           X6=RANGAM(SEED3,BETA);
           PIJ=X5/(X5+X6);
           NIJ=INT(RANUNI(SEED)*75)+1;
           YIJ=RANBIN(SEED,NIJ,PIJ);
           R=YIJ;
           M=NIJ;
           Q=M-R;

           OUTPUT;
          END;
        END;
  END;
```

Figure 3. Algorithm for Generating N Two-Way Tables

TABLE I

SIMULATION RESULTS OF TYPE I MODELS

SIMULATION OF TYPE I RANDOM MODELS A=5 B=4			
SIMULATION BASED ON 100 DIFFERENT DATA SETS			
THE PARAMETERS WERE: PHI1=0.2 PHI2=0.1429 PHI3=0.25			
NEGATIVE ESTIMATES OF PHI1, PHI2 AND PHI3 WERE SET TO ZERO			
VARIABLE N	MEAN	VARIANCE	STD ERROR OF MEAN
PE 100	0.31657241	0.00685854	0.00828163
η^* 100	0.50536630	0.00649048	0.00805635
R1EV 100	0.31276634	0.05761472	0.02400307
R1M 100	0.47761046	0.02807627	0.01675597
R2EV 100	0.29291694	0.04953145	0.02225566
R2M 100	0.47723260	0.03469507	0.01862661
R3EV 100	0.29411562	0.04617265	0.02148782
R3M 100	0.52776939	0.03053893	0.01747539
R4EV 100	0.33425280	0.04567791	0.02137239
R4M 100	0.51081009	0.02555468	0.01598583
R5EV 100	0.30405356	0.05012191	0.02238792
R5M 100	0.53158340	0.03217772	0.01793815
C1EV 100	0.33289324	0.03837459	0.01958943
C1M 100	0.53384104	0.02267104	0.01505690
C2EV 100	0.28889868	0.02693330	0.01641137
C2M 100	0.49703233	0.02381207	0.01543116
C3EV 100	0.33615772	0.02710309	0.01646302
C3M 100	0.51172636	0.02005275	0.01416077
C4EV 100	0.31331473	0.04603604	0.02145601
C4M 100	0.47683573	0.02436305	0.01560867
ϕ_1 100	0.18374462	0.07175443	0.02678702
ϕ_2 100	0.12856119	0.03667621	0.01915103
ϕ_3 100	0.28305181	0.00634260	0.00796405

Comments on Williams' Methodology for the

Mixed Case

Consider data as shown in Figure 2 (a two-way mixed cross classification) which arose as a result of a random sample of a levels of the row factor and J fixed levels of the column factor. Following the approach used in the previous section, we assume that $\theta_{1i} \stackrel{i\ i\ d}{\sim} \mathscr{B}(\eta,\phi_1)$ for i=1,...,a, where $\phi_1 \geq 0$.

Now if we let $\theta^*_{ij} = \dfrac{\theta_{1i} + \eta_{2j}}{2}$, where η_{2j} is the mean of the jth column then

θ^*_{ij} has expectation and variance given by

$$E(\theta^*_{ij}) = \frac{\eta + \eta_{2j}}{2}, \text{ and } V(\theta^*_{ij}) = \frac{1}{4}\phi_1\eta(1-\eta).$$

Thus the P_{ij} are independently distributed random variables on (0,1) such that, given θ_{1i} and η_{2j}, the conditional expectation and conditional variance are given by

$$E(P_{ij}|\theta_{1i}) = \theta^*_{ij}, \text{ and } V(P_{ij}|\theta_{1i}) = \phi_3\theta^*_{ij}(1-\theta^*_{ij}).$$

Thus the unconditional mean and variance of P_{ij} are as follows

$$E(P_{ij}) = E_{\theta_{1j}}\left[E_{P_{ij}|\theta_{1i}}\left[P_{ij}|\theta_{1i}\right]\right] = E_{\theta_{1i}}\left[\theta^*_{ij}\right] = \frac{\eta + \eta_{2j}}{2} = \eta^*_{2j}.$$

The variance of P_{ij} is

$$V(P_{ij}) = E_{\theta_{1i}}\left[E_{P_{ij}|\theta_{1i}}(P^2_{ij}|\theta_{1i})\right] - \eta^{*2}_{2j}$$

$$= E_{\theta_{1i}}\left[V_{P_{ij}|\theta_{1i}}(P_{ij} \mid \theta_{1i}) + \theta^{*}_{ij}{}^{2}\right] \cdot \eta^{*}_{2j}{}^{2}$$

$$= E_{\theta_{i}}\left[\phi_3\theta^{*}_{ij}-\phi_3(\theta^{*}_{ij})^{2}+(\theta^{*}_{ij})^{2}\right] - \eta^{*}_{2j}{}^{2}$$

$$= \phi_3\eta^{*}_{2j} + \tfrac{1}{4}(1-\phi_3)\left[\eta^{2}+ \phi_1\eta(1-\eta) + 2\eta\eta_{2j}+ \eta^{2}_{2j}\right]\cdot \eta^{*}_{2j}{}^{2}$$

$$= \phi_3\eta^{*}_{2j} - \phi_3\eta^{*}_{2j}{}^{2} + \tfrac{1}{4}(1-\phi_3)\phi_1\eta(1-\eta)$$

$$=\left\{\phi_3 \left[\eta^{*}_{2j}(1-\eta^{*}_{2j})\right]\right\} + \tfrac{1}{4}(1-\phi_3)\phi_1\eta(1-\eta). \qquad (3.18)$$

So in the mixed case one is unable to express the unconditional variance of the P_{ij} as a constant involving the ϕ's (in the case ϕ_1 and ϕ_2) times (its mean)(1- its mean). It seems that the generation of θ's as discussed in the previous section does not work for the mixed model.

It is possible that in the mixed model situation one can generate different θ_i's for a given row by weighting them by c_j where c_j is some sort of weight depending on the ratio of the particular column mean η_{2j} to the overall mean $\overline{\eta}_{j.}$. We feel that although the procedure works well as an extension of Williams' algorithm for the two-way random case it fails to be of use for the mixed two way model.

CHAPTER IV

TYPE II MODELS

Introduction

In this chapter we will consider *Type II models*; that is, models that belong to the class of generalized linear models, and more specifically logistic models, which include random effects and overdispersion parameters. We apply these models to data of the form specified in the introduction of Chapter III and we are going to address both the mixed case and random case. We are going to expand along the lines of Pierce and Sands (1975). They suggested that random perturbations can be justified as effects due to unmeasured sources of experimental error which affect the group of the n_{ij} binary observations as a whole. Our approach follows more closely the methodology introduced by Hinde (1982). His model is similar to that of Pierce and Sands but is applied to Poisson overdispersed data instead of binomial. Also the method of fitting is slightly different.

Our extensions consider models that not only model the overdispersion on the same additive scale as the fixed effects (in this case the logistic scale) but also account for two random factors in the two-way random classification with no interaction. Our fitting methods are based on Hinde's suggestion, that when dealing with single parameter exponential models and using the canonical link to fit overdispersed data, one fits a generalized linear model on an expanded set of observations. In our case we will have to expand the observations more times due to the existence of more than one random factor.

In its simplest form, a generalized linear model (GLM) is specified by:
(i) independent observations $y_1,...,y_n$ distributed according to a member of the exponential family of distributions;
(ii) a set of explanatory variables $\underset{\sim}{x}'$ available for each observation, describing the systematic linear component $\eta = X\underset{\sim}{\beta}$, and
(iii) the link function $g(\mu_i)=\eta_i$, relating the mean of an observation to the systematic component.

Consider first that the row factor and the column factor are fixed and the overdispersion enters the model through the linear predictor as an unobserved random variable. We know that if all effects entering the linear predictor were fixed, then our model would belong to the above class of GLM's since R_{ij} given $\underset{\sim}{x}'$ and p_{ij} is distributed Binomial(n_{ij},p_{ij}) which belongs to the exponential family. For our setup, the X matrix would correspond to the design matrix for the two way structure without interaction and the link function would be the logistic function as given in (3.1). Estimating the unknown parameters via iterative reweighted least squares produces maximum likelihood estimates. For more details see McCullagh and Nelder (1983).

If the random effects are on the same scale chosen for transforming the p_{ij}, then the form of the generalized linear model is

$$\underset{\sim}{\eta} = X\,\underset{\sim}{\beta} + Z\,\underset{\sim}{u}, \tag{4.1}$$

where   X is the design matrix for the fixed effects,

$\underset{\sim}{\beta}$ is the vector of fixed unknown parameters,

Z is the design matrix for the random effects,

$\underset{\sim}{u}$ is a random unobserved vector and

η is the linear predictor (i.e. $\eta = g(\mu)$ where $E(\underset{\sim}{Y})=\underset{\sim}{\mu}$).

Throughout this chapter we will consider g to be the logit link function (i.e., $\log \frac{p_{ij}}{(1-p_{ij})}$ where p_{ij} is the binomial parameter).

Review of Hinde's Approach

Let us first consider the case where there is overdispersion and study the model above by considering the rows and the columns as fixed effects and the overdispersion u_{ij} as the only random effect. This is the model considered by Pierce and Sands and we are going to develop a methodology for getting the estimates directly, following the approaches of Hinde (1982), and Brillinger and Preisler (1983) as they were applied to log-linear models for overdispersed counts. Thus, given u_{ij} we have

$$f(R_{ij}| \mu,\alpha_i,\beta_j,u_{ij}) \sim Binomial(n_{ij},p_{ij})$$

where i=1,...,I, j=1,...,J and such that

$$\log \frac{p_{ij}}{(1-p_{ij})} = \eta_{ij} = \mu + \alpha_i + \beta_j + u_{ij} \tag{4.2}$$

where $u_{ij} \sim F$. Thus the log-likelihood given u_{ij} is given by

$$\prod_i \prod_j f\left(R_{ij}| \mu,\alpha_i,\beta_j,u_{ij}\right) = \prod_i \prod_j \left[\frac{p_{ij}}{1-p_{ij}}\right]^{r_{ij}}(1-p_{ij})^{n_{ij}} \tag{4.3}$$

where

$$p_{ij} = \frac{e^{(\mu + \alpha_i + \beta_j + u_{ij})}}{1 + e^{(\mu + \alpha_i + \beta_j + u_{ij})}}.$$

The simplest situation is the case where $u_{ij} \sim N(0,1)$. Then the marginal distribution of R_{ij} is given by

$$\int_{-\infty}^{+\infty} \left[\frac{p_{ij}}{1-p_{ij}}\right]^{r_{ij}} (1-p_{ij})^{n_{ij}} \phi(u_{ij}) du_{ij}$$

where $\phi(u_{ij})$ is the standard normal pdf, and the marginal log-likelihood is given by

$$\sum_i \sum_j \log \int f\left[R_{ij} \mid \mu, \alpha_i, \beta_j, u_{ij}\right] \phi(u_{ij}) du_{ij}. \tag{4.4}$$

Consider the case that F is known up to a scale parameter. Going a step further and following the usual assumptions of variance component methodology, we assume that $u_{ij} \sim N(0, \theta_1^2)$. Similar assumptions were made by Pearce and Sands, in the binomial setup, and by Hinde and later on by Brillinger and Preisler for the Poisson setup. The above assumption implies that we could replace u_{ij} by $\theta_1 z_{ij}$ where $z_{ij} \sim N(0,1)$, and θ_1 is the unknown scale parameter. Thus p_{ij} in (4.3) becomes

$$p_{ij} = \left[1 + e^{-(\mu + \alpha_i + \beta_j + \theta_1 z_{ij})}\right]^{-1}.$$

One technique for maximizing $\mathscr{L}(\mu, \alpha_i, \beta_j, \theta_1)$ in order to find the maximum likelihood estimates of the parameters, is to use the EM algorithm of Dempster, Laird, and Rubin (1977). For this we need the so called complete data likelihood

$$h(\mu,\alpha_i,\beta_j,\theta_1) = \prod_i \prod_j f\left[R_{ij} \mid \mu,\alpha_i,\beta_j,\theta_1,z_{ij}\right]\phi(z_{ij}).$$

For the complete data, (log h) would be maximized over μ, α_i, β_j and θ_1 (M step) but since z_{ij} is unobservable we replace log h by $E\left[\log h \mid z\right]$ and the current values of μ, α_i, β_j and θ_1 (E step). If we let $\psi = \{\mu,\alpha,\beta,\theta_1\}$ then the EM algorithm defines the following iterative process for the estimates of ψ.

 E step: Compute $Q(\psi \mid \psi^{(p)}) = E(\log h \mid z)$

 M step: Choose $\psi^{(p+1)}$ to maximize $Q(\psi \mid \psi^{(p)})$

where $\psi^{(p)}$ is the estimate of ψ from the pth iteration. Hinde showed that differentiating $Q(\psi \mid \psi^{(p)})$ results in the same equations as those obtained by differentiating the marginal log-likelihood given by (4.4). Thus after some simplification we could obtain the maximum likelihood estimates of the parameters by differentiating

$$\sum_i \sum_j \log \int_{-\infty}^{+\infty} \frac{e^{r_{ij}(\mu + \alpha_i + \beta_j + \theta_1 z_{ij})}}{\left[1 + e^{(\mu + \alpha_i + \beta_j + \theta_1 z_{ij})}\right]^{n_{ij}}} \phi(z_{ij})dz_{ij} \qquad (4.5)$$

Unfortunately, the integral in (4.5) has no closed form solution. However, since the integral is over a normal density this allows us to approximate it by a K-point Gaussian quadrature. This simply amounts to replacing the $\int_{-\infty}^{+\infty} f(z_{ij})dz_{ij}$ by $\sum_{k=1}^{K} f(z_{1k})$ and $\phi(z_{ij})$ by w_k, the associated quadrature weights. A table of quadrature points and their associated weights for different values of K are given in Table II. More extensive and detailed tables on Gaussian quadrature with K's up to 136 can be found in Stroud and Secrest (1966) (Table Five). Note that his x_i's correspond to

TABLE II

ABSCISSAS AND WEIGHT FACTORS
FOR GAUSSIAN QUADRATURE*

K	z_{1k}	w_{1k}
3	±1.73204	0.16667
	0.00000	0.66666
4	±0.74197	0.45412
	±2.33441	0.04588
5	±1.35562	0.22208
	±2.85696	0.01126
	0.00000	0.53333
6	±0.61670	0.40883
	±1.88918	0.08862
	±3.32426	0.00256
7	±1.15441	0.24012
	±2.36676	0.03076
	±3.75048	0.00055
	0.00000	0.45714
8	±0.53908	0.37301
	±1.63651	0.11724
	±2.80249	0.00964
	±4.14455	0.00011
9	±1.02325	0.24410
	±2.07684	0.04992
	±3.20543	0.00279
	±4.51274	0.00002
	0.00000	0.40635
10	±0.48493	0.34464
	±1.46599	0.13548
	±2.48432	0.02186
	±3.58182	0.00076
	4.85946	0.000004

*from Stroud and Secrest (1966)

$\left[z_{1k} \sqrt{2} \right]$ and his A_i is $\left[\dfrac{w}{\sqrt{2\pi}} \right]$. Similar tables are given on page 924 of

Abramowitz and Stegun (1972). Thus if we denote by z_{1k} the gaussian quadrature point associated with the weight w_k then (4.5) becomes

$$\sum_i \sum_j \log \sum_{k=1}^{K} \frac{e^{r_{ij}(\mu + \alpha_i + \beta_j + \theta_1 z_{1k})}}{\left[1 + e^{(\mu + \alpha_i + \beta_j + \theta_1 z_{1k})} \right]^{n_{ij}}} w_{1k}. \tag{4.6}$$

Estimates of ψ are obtained by taking the partial derivatives of (4.6) with respect to the parameters. If we denote the partial derivatives (i.e., the so called score vector) by $\zeta(r_{ij} | \psi)$ then the maximum likelihood equations for ψ are given by

$$\sum_i \sum_j \zeta(r_{ij} | \hat{\psi}) = 0$$

Brillinger and Preisler (1983) gave various conditions leading to the consistency and asymptotic normality of $\hat{\psi}$. For example, under conditions (B-1) to (B-4) given by Huber (1967), $\hat{\psi}$ can be shown to be consistent. Further, if ψ_o is the true parameter vector then $\sqrt{IJ}\,(\hat{\psi} - \psi_o)$ is asymptotically normal with mean 0 and covariance matrix $i(\psi_o)^{-1}$ (the inverse of the information matrix), under conditions (N-1) to (N-4) of Huber (1967) which are satisfied if $E\left[\zeta(r | \psi) \right]$ is differentiable at $\psi = \psi_o$ and

$$i(\psi) = E\left[\zeta(r | \psi)\, \zeta(r | \psi)' \right].$$

Thus maximizing the marginal likelihood (4.6) corresponds to solving the following equations :

$$\sum_{i} \sum_{j} \frac{1}{f^*_{ij}} \frac{\partial f^*_{ij}}{\partial \psi} = 0 \qquad (4.7)$$

where

$$f^*_{ij}(R_{ij}|\mu,\alpha_i,\beta_j,\theta_1,z_{1k}) = \sum_{k=1}^{K} \left\{ \frac{e^{r_{ij}(\mu + \alpha_i + \beta_j + \theta_1 z_{1k})}}{\left[1 + e^{(\mu + \alpha_i + \beta_j + \theta_1 z_{1k})}\right]^{n_{ij}}} \right\} w_{1k}$$

is the unconditional distribution of R_{ij}. If we use B_{ijk} to denote $f(R_{ij}|\mu,\alpha_i,\beta_j,\theta_1,z_{1k})$, then from (4.7) we have

$$\frac{\partial f^*_{ij}}{\partial \psi} = \sum_{k=1}^{K} w_{1k} \frac{\partial B_{ijk}}{\partial \psi} .$$

For notational convenience, let $\underset{\sim}{\beta}$ be the vector associated with the fixed effects (i.e., $\underset{\sim}{\beta}=(\mu,\alpha_1,...,\alpha_I,\beta_1,...,\beta_J)'$; then we can show that

$$\frac{\partial B_{ijk}}{\partial \beta_s} = B_{ijk} \, x_s \left[r_{ij} - n_{ij} \frac{e^{r_{ij}(\mu + \alpha_i + \beta_j + \theta_1 z_{1k})}}{\left[1 + e^{(\mu + \alpha_i + \beta_j + \theta_1 z_{1k})}\right]} \right]$$

where $x_s = 0$ or 1 and s is indexing the elements of $\underset{\sim}{\beta}$, and that

$$\frac{\partial B_{ijk}}{\partial \theta_1} = B_{ijk} \, z_{1r} \left[r_{ij} - n_{ij} \frac{e^{r_{ij}(\mu + \alpha_i + \beta_j + \theta_1 z_{1k})}}{\left[1 + e^{(\mu + \alpha_i + \beta_j + \theta_1 z_{1k})}\right]} \right]$$

where z_{1k} is the kth gaussian quadrature point. Thus we have shown that the

system (4.7) has taken the familiar weighted regression normal equations form (i.e., $X'V^{-1}(Y- X\beta)$) that corresponds to fitting a binomial regression model to an expanded set of observations. In general we could write these equations as follows:

$$\sum_i \sum_j \sum_k w^*_{ijk} \frac{\partial f}{\partial \psi} = 0 \tag{4.8}$$

where

$$w^*_{ijk} \equiv w^*_{ijk}(R_{ij}|\mu,\alpha_i,\beta_j,\theta_1,z_{1k}) = \frac{w_{1k} f(R_{ij}|\mu,\alpha_i,\beta_j,\theta_1,z_{1k})}{f^*_{ij}(R_{ij}|\mu,\alpha_i,\beta_j,\theta_1,z_{1k})} \ .$$

In other words, we have shown that the normal equations above could be used iteratively to produce approximate maximum likelihood estimates if they were being applied to the expanded set of observations

$$
\underset{\sim}{R} = \left[\overset{\text{ORIGINAL}}{\underset{\text{DATA}}{\overbrace{R_{11},\ldots,R_{IJ}}}}, R_{11},\ldots,R_{IJ},\ldots,\underbrace{R_{11},\ldots,R_{IJ}}_{} \right]' .
$$

$$K \quad \text{times the original data}$$

Thus the original data set is repeated K times and includes as prior weights the vector:

$$\underset{\sim}{W}^* = \left[w^*_{111},\ldots,w^*_{IJ1},w^*_{112},\ldots,w^*_{IJ2},\ldots,w^*_{11K},\ldots,w^*_{IJK} \right]'$$

where

$$w^{*}_{ijk}=f\left[R_{ij}|z_{1k},\mu^{(p)},\alpha_i^{(p)}\ \beta_j^{(p)},\theta_1^{(p)}\right]\frac{w_k}{f^{*(p)}_{ij}}$$

(for $i=1,...,I$, $j=1,...,J$ and $k=1,...,K$) are the weights for the $(p+1)^{th}$ iteration and $\mu^{(p)}$, $\alpha_i^{(p)}$, $\beta_j^{(p)}$ and $\theta_1^{(p)}$ denote the estimates of the parameters from the $(p)^{th}$ iteration.

This iterative procedure was programmed using the SAS IML and MA-CRO language. The sequence of Binomial fits is programmed using the SAS NLIN procedure for producing the parameter estimates and for calculating and updating the weights w_{ijk}.

For addressing questions about the fitted model, we suggest using the deviance which is defined as 2*log(saturated model) - 2log(likelihood) as given by

$$D(\hat{\mu},\hat{\alpha}_i,\hat{\beta}_j,\hat{\theta}_1)=-2\ \sum_i\ \sum_j\left[\log\hat{f}^{*}_{ij}-\ r_{ij}\log(r_{ij})-(n_{ij}-r_{ij})\log(n_{ij}-r_{ij})\right]. \qquad (4.9)$$

In this procedure the parameters μ, α_i, β_j, and θ_1 are estimated simultaneously. Estimating standard errors for the estimates of the parameters will involve obtaining the inverse of the observed information matrix

$$\left\{\sum_{i=1}^{I}\ \sum_{j=1}^{J}\left[\frac{\partial\ \log\ f^{*}_{ij}\ (R_{ij}|\ \underset{\sim}{\beta},\theta_1)}{\partial\ \psi}\right]\left[\frac{\partial\ \log\ f^{*}_{ij}\ (R_{ij}|\ \underset{\sim}{\beta},\theta_1)}{\partial\ \psi}\right]'\right\}^{-1}. \qquad (4.10)$$

In an attempt to verify the suggestion of Hinde that K as big as 3 or 5

could be used to produce reasonably good estimates of $\underset{\sim}{\beta}$ and θ_1, we used the above algorithm on a set of data reported by Crowder (1978) and also analyzed by Williams (1982) by the methodology discussed in Chapter 3. The fitted model for a 2x2 factorial experiment with overdispersion is given by

$$\log\left[\frac{P_{ij}}{1-P_{ij}}\right] = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \theta_1 z_{ij}$$

where i,j=1,2 and $z_{ij} \sim N(0,1)$. In Table III we report the number of Gaussian quadrature points (GQ) used, the number of iterations needed for obtaining convergence of the deviance (D) within .01 of the deviance from the previous iteration, and the estimates of the parameters. The last row in the table represents the equivalent estimates reported by Williams. The listing of the program that was used to compute the entries appearing in Table III is provided in Appendix C. The first important observation is that estimates of the parameters change considerably based on the number of quadrature points. The second important observations is that the estimates of the fixed effects are "closer" than the estimates of the overdispersion depending on the number of quadrature points used. Comparing the first two rows of Table III with the last row (i.e., the estimates obtained using Williams' algorithm) we conclude again that the fixed effects estimates are very close in all cases but, the overdispersion parameter estimate is much smaller in Williams's approach. In Chapter IV when comparing the models we will see that Williams's estimate for the overdispersion parameter can only be as large as $0.25(\theta_1)$ which certainly holds true.

TABLE III

RESULTS FROM ANALYSIS BASED ON CROWDER'S DATA SET USING 3 AND 5 QUADRATURE POINTS

# of GQ points	# of iter $\left[D_n < .01 \, D_o\right]$	$\hat{\mu}$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\alpha\beta}_{22}$	$\hat{\theta}_1$
3	7	-0.554	.104	1.35	-.819	.2146
5	6	-0.548	.094	1.34	-.815	.2424
Williams method	5	-0.535	.070	1.33	-.819	.0249

Extensions of Hinde's Methodology for the

Mixed Case

Consider the case of having one of the dimensions of the two way table (the columns for example) random in addition to the overdispersion. In the model given by (4.2), let the column effect be random. The setup is the same as the one which appears in Figure 1. Let us denote this random effect by b_j where $j=1,...,b$ and $i=1,...,a$. Thus we can rewrite the model as follows

$$f(R_{ij} \mid u_{ij}, b_j, \mu, \alpha_i) \sim \text{Binomial}(n_{ij}, p_{ij})$$

and

$$\log \frac{p_{ij}}{(1-p_{ij})} = \eta_{ij} = \mu + \alpha_i + b_j + u_{ij}$$

where μ and α_i are unknown fixed parameters and $b_j \sim F_1$ and $u_{ij} \sim F_2$.

The situation we are going to consider is $b_j \sim N(0, \theta_1^2)$ and $u_{ij} \sim N(0, \theta_2^2)$ and that the two random variables are independent. Let us replace b_j and u_{ij} by $\theta_1 z_{1j}$ and $\theta_2 z_{2ij}$ and let $\underset{\sim}{\beta} = (\mu_1,...,\mu_I)'$, where $\mu_i = \mu + \alpha_i$, represent the vector of the row means from the fixed part of the model and $\underset{\sim}{\theta} = (\theta_1, \theta_2)'$ represent the vector of the two standard deviations for the two unobservable random variables z_{1j} and z_{2ij}, respectively. Let $H\left[z_{1j}, z_{2ij}\right]$ represent the conditional distribution of R_{ij} given z_{1j} and z_{2ij} which means that

$$H\left[z_{1j}, z_{2ij}\right] = \frac{e^{r_{ij}(\mu + \alpha_i + \theta_1 z_{1j} + \theta_2 z_{2ij})}}{\left[1 + e^{(\mu + \alpha_i + \theta_1 z_{1j} + \theta_2 z_{2ij})}\right]^{n_{ij}}}. \qquad (4.11)$$

The marginal likelihood is given by

$$L(\underset{\sim}{\beta},\underset{\sim}{\theta}) = \int\limits_{\{\underset{\sim}{z}\}} \prod_i \prod_j H\left[z_{1j}, z_{2ij}\right] \phi(\underset{\sim}{z}) \, d\underset{\sim}{z}$$

and thus the marginal log-likelihood $\mathscr{L}(\underset{\sim}{\beta},\underset{\sim}{\theta})$ simplifies to

$$\sum_j \log\left\{ \int\limits_{z_1} \left\{ \prod_i \int\limits_{z_2} H\left[z_{1j}, z_{2ij}\right] \phi(z_{2ij}) \, dz_{2ij} \right\} \phi(z_{1j}) \, dz_{1j} \right\}. \qquad (4.12)$$

Using a K point Gaussian quadrature for the inside integral and L point for the outside we replace the integrals by summations and the normal p.d.f's by the equivalent weights and we get from (4.12) that

$$\mathscr{L}(\underset{\sim}{\mu},\underset{\sim}{\theta}) = \sum_j \log\left\{ \sum_{l=1}^{L} \left[\prod_i \left(\sum_{k=1}^{L} B_{ijkl} \, w_k \right) \right] w_l \right\} \qquad (4.13)$$

where now

$$B_{ijkl} = \frac{e^{r_{ij}(\mu + \alpha_i + \theta_1 z_1 + \theta_2 z_k)}}{\left[1 + e^{(\mu + \alpha_i + \theta_1 z_1 + \theta_2 z_k)}\right]^{n_{ij}}}.$$

In the following steps we are going to outline the derivations for the appropriate weights that are needed for the iterations that provide solutions to the normal equations.

1). The partial derivatives of B_{ijkl} wrt μ_i and θ_s where $s=1,2$. take the form

$$\frac{\partial B_{ijkl}}{\partial \mu_i} = B_{ijkl} \; x_i \left\{ r_{ij} - n_{ij} \; \frac{e^{r_{ij}(\mu_i + \theta_1 z_{11} + \theta_2 z_{2k})}}{1 + e^{(\mu_i + \theta_1 z_{11} + \theta_2 z_{2k})}} \right\}$$

where $x_i = 0$ or 1 (depending on the row), and

$$\frac{\partial B_{ijkl}}{\partial \theta_s} = B_{ijkl} \; z_s \left\{ r_{ij} - n_{ij} \; \frac{e^{r_{ij}(\mu_i + \theta_1 z_{11} + \theta_2 z_{2k})}}{1 + e^{(\mu_i + \theta_1 z_{11} + \theta_2 z_{2k})}} \right\}$$

where $z_s = z_{11}$ or z_{2k} depending on s being equal to 1 or 2 respectively.

2). Using the rules of differentiation one can show that the partial derivative (say for example wrt μ_i) of the product of the sum of functions takes the following form

$$\frac{\partial}{\partial \mu_i} \prod_i \left[\sum_k B_{ijkl}(\mu_i) w_k \right] = \sum_{i'} \left[\prod_{i(\neq i)} \left\{ \left[\sum_k B_{ijkl}(\mu_i) w_k \right] \left[\sum_k B_{i'jkl}(\mu_i) w_k b_{i'jkl} \right] \right\} \right]$$

where $b_{i'jkl} = \dfrac{\partial}{\partial \mu_i} B_{ijkl}(\mu_i)$.

3). The partial of the marginal log-likelihood using the results of steps 1 and 2 becomes

$$\frac{\partial}{\partial \mu_i} \mathscr{L}(\mu, \theta) = \sum_j \frac{\left[\sum_1 w_1 \sum_{i'} \left\{ \prod_{i(\neq i)} \left[\sum_k B_{ijkl}(\mu_i) w_k \right] \left[\sum_k B_{i'jkl}(\mu_i) w_k b_{i'jkl} \right] \right\} \right]}{\sum_{1=1}^{L} \left[\prod_i \left(\sum_{k=1}^{L} B_{ijkl} \; w_k \right) \right] w_1}$$

and if we let $B_{ijl} = \left[\sum_k B_{ijkl}(\mu_i) w_k \right]$ then the above becomes

$$\frac{\partial}{\partial \mu_i} \mathscr{L}(\underset{\sim}{\mu},\underset{\sim}{\theta}) = \sum_j \frac{\left[\sum_l w_l \left[\prod_{i(\neq i')} B_{ijl} \right] \left\{ \sum_i \frac{\left[\sum_k B_{i'jkl}(\mu_i) w_k b_{i'jkl} \right]}{B_{ijl}} \right\} \right]}{\sum_{l=1}^{L} \left[\prod_i \left(\sum_{k=1}^{L} B_{ijkl} w_k \right) \right] w_l}$$

$$= \sum_i \sum_j \sum_k \sum_l \left\{ \frac{\left[w_l \prod_i B_{ijl} \right]}{\sum_l w_l \left[\prod_i B_{ijl} \right]} \frac{\left[w_k B_{ijkl} \right]}{B_{ijkl}} \right\} b_{ijkl} \qquad (4.14)$$

$$= \sum_k \sum_l \sum_i \sum_j \left\{ \begin{matrix} \text{approrpiate} \\ \text{weights} \\ \text{labeled } W_{ijkl} \end{matrix} \right\} x_i \left\{ r_{ij} - n_{ij} \frac{e^{r_{ij}(\mu_i + \theta_1 z_{11} + \theta_2 z_{2k})}}{1 + e^{(\mu_i + \theta_1 z_{11} + \theta_2 z_{2k})}} \right\} \qquad (4.15)$$

where $x_i = 0$ or 1 (depending on the appropriate row indicator variable). We get a similar result for the partial wrt θ_s with the only difference being that the x_i's in (4.15) are replaced by the z_s's where $s = 1,2$ depending on θ_s. Therefore we have shown that the normal equations could produce maximum likelihood estimates of the parameters by simply employing an iterative reweighted least squares algorithm on an expanded data set. In fact the original data set is expanded out to $N \times K \times L$ where N represents the total number of cells in the two-way table and K and L are the number of quadrature points chosen for each of the two random factors. The process represented below

ORIGINAL
DATA

$$\underset{\sim}{R} = \left[R_{11}, \ldots, R_{ab}, R_{11}, \ldots, R_{ab}, \ldots, R_{11}, \ldots, R_{ab} \right]'$$

L x K times the original data

amounts to augmenting not only the original response variable but also augmenting the X's (i.e. the indicator variables that make up the design matrix for the fixed part of the model) and the z's and using as prior weights the vector

$$\underset{\sim}{W}^* = \left[w^*_{1111}, \ldots, w^*_{ab11}, w^*_{1121}, \ldots, w^*_{ab21}, \ldots, w^*_{11KL}, \ldots, w^*_{abKL} \right]'$$

$$\underset{K \text{ x } L \text{ times the original data}}{\longleftarrow \text{———————————}|\text{———————————} \longrightarrow}$$

where

$$w^*_{ijrs} = \left(\frac{B_{jl}}{B_j}\right)\left(\frac{B_{ijkl}}{B_{ijl}}\right) \tag{4.16}$$

$B_{ijkl} = f(R_{ij}|z_k, z_l, \mu_i^{(p)}, \theta_1^{(p)}, \theta_2^{(p)})\, w_k;$ $B_{ijl} = \sum_k B_{ijlk};$ $B_{jl} = (\prod_i B_{ijl})w_l;$ $B_j = \sum_l B_{jl};$ (i=1,...,a, j=1,...,b, k=1,...,K, and l=1,...,L) and again the weights for the $(p+1)^{th}$ iteration are calculated based on $\mu_i^{(p)}$, and $\theta_1^{(p)}$. For addressing questions about the fitted model we suggest using the deviance given by

$$D(\hat{\mu}_i, \hat{\phi}_1, \hat{\phi}_2) = -2 \sum_i \sum_j \left[\log \hat{f}^*_{ij} - r_{ij}\log(r_{ij}) - (n_{ij}-r_{ij})\log(n_{ij}-r_{ij})\right] \tag{4.17}$$

The importance of the random effects can be assessed by constructing some form of interval estimate for ϕ_2 and ϕ_1. Estimation of both ϕ_1 and ϕ_2 is accomplished at the same time as μ and α_i. Estimating standard errors involves obtaining the inverse of the observed information matrix.

$$\left\{ \sum_{j=1}^{b} \left[\frac{\partial \log f^*_j(R_j| \beta, \phi_1, \phi_2)}{\partial \psi}\right]\left[\frac{\partial \log f^*_j(R_j| \beta, \phi_1, \phi_2)}{\partial \psi}\right]' \right\}^{-1} \tag{4.18}$$

where $f_j^*(R_j, \beta, \phi_1, \phi_2)$ is the likelihood for a fixed level j of the random effect which translates to

$$f_j^*(R_j | \beta, \phi_1, \phi_2) = \left\{ \sum_{l=1}^{L} \left[\prod_i \left(\sum_{k=1}^{L} B_{ijkl} \, w_k \right) \right] w_l \right\}$$

where again B_{ijkl} is the likelihood of any given observation on the expanded data set

$$B_{ijkl} = \frac{e^{r_{ij}(\mu + \alpha_i + \theta_1 z_1 + \theta_2 z_k)}}{\left[1 + e^{(\mu + \alpha_i + \theta_1 z_1 + \theta_2 z_k)} \right]^{n_{ij}}}.$$

It is important here to mentioned that the SAS code for implementing the above steps is given in Appendix D. More about the program and also discussions on the results from applying both types of models to the Virginia Pine tree data set is given in the following chapter.

Extensions of Hinde's Methodology for the
Random Model

Now let us extend our model to the case of both dimensions in our two-way table being random in addition to trying to allow for overdispersion. This situations is totally analogous to the situation depicted in Figure 2. The truth of the matter is that the algorithm becomes difficult to implement from the computing point of view. It is even more of a problem in our case since our programs are written in SAS using limited IML workspace. We do nothing more than outline the derivations. But in reality one has to go through an analogous procedure to get estimates of the parameters in the model. Thus in the model given by (4.1) we change the column effect from being fixed to being random. The setup is the same as the one appearing in Chapter III with the only difference that the random effects enter the

model at the same level as the systematic effects. Let us once again denote the random effects by a_i and b_j where $j=1,...,b$ and $i=1,...,a$. Thus we can rewrite the model as follows

$$f(R_{ij}|\ u_{ij},\mu,a_i,b_j) \sim Binomial(n_{ij},p_{ij})$$

and

$$\log \frac{p_{ij}}{(1-p_{ij})} = \eta_{ij} = \mu + a_i + b_j + u_{ij}$$

where μ is an unknown fixed parameter and $a_i\sim F_1$, $b_j\sim F_2$ and $u_{ij} \sim F_3$.

Again we will assume normality and independence of the random effects and replace the random effects in the model equation by the standard deviations times the z's. Now the only fixed effect is μ the overall mean and the vector $\underset{\sim}{\theta}=(\theta_1\ \theta_2\ \theta_3)'$ represents the standard deviations for the three random variables z_{1i}, z_{2j} and z_{3ij}, respectively. Let $H\left[z_{1i},z_{2j},z_{3ij}\right]$ represent the conditional distribution of R_{ij} given the z's, similar to (4.11); then the marginal likelihood is given by

$$L(\mu,\underset{\sim}{\theta})= \int\limits_{\{\underset{\sim}{z}\}} \prod_i \prod_j H\left[z_{1i},z_{2j},z_{3ij}\right]\ \phi(\underset{\sim}{z})\ d\underset{\sim}{z}.$$

The order of integration in terms of i or j is not important. Replacing the integrals by summations and the normal p.d.f's by the equivalent weights we could get the estimates of the θ's by fitting an intercept term and the appropriate combination of z's on the original data set after expanding it K*L*M times. Here again K, L and M are the numbers of quadrature points used to approximate the three different integrals in terms of the z's. We

do not need to have K=L=M. Our decision on the choices of K, L and M; although limited to numbers definitely smaller than 5 for a medium size data set, should reflect our suspicions on the size of the θ's. This means that if for any reason the size the standard deviation for the any of the random effects is large then we must use a higher number of quadrature points to approximate the particular integral associated with that effect. Bock and Aitkin (1981) suggest that numbers as big three, can be used to obtain good estimates of the parameters. What may also happen is that one might fail to show significance for some of these variance components. If, for example, one of the variance components fails to be significant, then for the resulting reduced model one could increase the orders of the individual quadratures. This would produce better accuracy for the remaining variance components.

CHAPTER V

COMPARISONS APPLICATIONS AND DISCUSSIONS

Comparisons of Type I and Type II Models

In this chapter we will attempt to compare the models considered in Chapters III and IV by means of comparing estimates and their properties. We will fit both models to a set of data from a genetic experiment considering the survival of different families of Virginia Pine. We will also elaborate on the difficulties that we encountered applying the models to a large data set. At the end of the chapter we will close with suggestions for further research.

In what follows we will attempt to further explain the meaning of the parameters for the two types of models. Consider both models with rows and columns fixed and both allowing for an overdispersion parameter. That means that $Y_{ijk}|p_{ij}$~Binomial$(1,p_{ij})$ was a binary observation on the kth unit, $k=1,...n_{ij}$ in row i and column j. For the Type II model

$$p_{ij} = \frac{e^{\mu_{ij} + \sigma z_{ij}}}{1 + e^{\mu_{ij} + \sigma z_{ij}}}, \tag{5.1}$$

where Z_{ij}~N$(0,1)$.

For the Type II model the covariance of two binary observations which belong to the same cell in the table is given by

$$\text{Cov}(Y_{ijk}, Y_{ijk'}) = E(Y_{ijk}Y_{ijk'}) - \left[E(Y_{ijk})E(Y_{ijk'}) \right]$$

$$= E_{P_{ij}} \left[E_{Y_{ijk}Y_{ijk'}|P_{ij}} \left[Y_{ijk}Y_{ijk'} \right] \right] - \left\{ E_{P_{ij}}(P_{ij}) \right\}^2$$

$$= E_{Z_{ij}} \left[E_{Y_{ijk}Y_{ijk'}|Z_{ij}} \left[Y_{ijk}Y_{ijk'} \right] \right] - \left\{ E_{Z_{ij}}(P_{ij}) \right\}^2.$$

Because Y_{ijk} and $Y_{ijk'}$ are independent given Z_{ij}, the covariance of two binary observations in the same cell is

$$\text{Cov}(Y_{ijk}, Y_{ijk'}) = \text{Var}_{Z_{ij}}(P_{ij}). \tag{5.2}$$

The variance of a binary observation is

$$\text{Var}(Y_{ijk}) = E_{Z_{ij}} \left[\text{Var}_{Y_{ijk}|Z_{ij}}(Y_{ijk}|P_{ij}) \right] + \text{Var}_{Z_{ij}} \left[E_{Y_{ijk}|Z_{ij}}(Y_{ijk}|P_{ij}) \right]$$

$$= E_{Z_{ij}} \left[P_{ij}(1-P_{ij}) \right] - \text{Var}_{Z_{ij}}(P_{ij}).$$

$$= E_{Z_{ij}} \left[P_{ij} \right] - \left[E_{Z_{ij}}(P_{ij}) \right]^2. \tag{5.3}$$

Thus, from (5.2) and (5.3) the correlation between two binary observations in the same cell, say ij, is

$$\text{Corr}(Y_{ijk}, Y_{ijk'}) = \frac{\text{Var}_{Z_{ij}}(P_{ij})}{E_{Z_{ij}} \left[P_{ij} \right] - \left[E_{Z_{ij}}(P_{ij}) \right]^2}. \tag{5.4}$$

Equation (5.4) in relation to the Williams' assumption that $E(P_{ij})=\pi_{ij}$,

where $\pi_{ij}=\dfrac{e^{\mu_{ij}}}{1+e^{\mu_{ij}}}$, and the $Var(P_{ij})=\phi\pi_{ij}(1-\pi_{ij})$ results in the following

$$Corr(Y_{ijk}, Y_{ijk'})=\frac{\phi\ \pi_{ij}(1-\pi_{ij})}{\pi_{ij}-\pi_{ij}^2}=\phi. \qquad (5.5)$$

Therefore (5.5) means that Williams' model corresponds to a constant intraclass correlation model. For the Type II overdispersion model we obtain the following by expanding (5.1) around $z_{ij}=0$. If we denote p_{ij} as $g(z_{ij})$ then

$$\frac{d}{dz}g(z)\Big|_{z=0}=\frac{\sigma\ e^{\mu_{ij}}}{\left[1\ +\ e^{\mu_{ij}}\right]^2}$$

and for $\pi_{ij}=\dfrac{e^{\mu_{ij}}}{1+e^{\mu_{ij}}}$ a first order Taylor's expansion of $g(z_{ij})$ gives that

$$p_{ij}\cong\pi_{ij}\left[1+\sigma z_{ij}(1-\pi_{ij})\right]. \qquad (5.6)$$

From the fact that $Z_{ij}\sim N(0,1)$ and (5.6) we obtain that

$$E_{Z_{ij}}(P_{ij})\cong\pi_{ij}\ \text{ and the }\ Var_{Z_{ij}}(P_{ij})\cong\sigma^2\pi_{ij}^2(1-\pi_{ij})^2.$$

Thus the above relations result in the following expression for the correlation among two binary observations of the Type II model (i.e. Pierce's overdispersion model) is given by

$$\text{Corr}(Y_{ijk}, Y_{ijk'}) = \frac{\sigma^2 \pi^2_{ij}(1-\pi_{ij})^2}{\pi_{ij}-\pi^2_{ij}} = \sigma^2 \pi_{ij}(1-\pi_{ij}). \tag{5.7}$$

That means that the intraclass correlation depends on the probability of success for that class. Thus the only way that there is direct correspondence among the parameters ϕ and the σ^2 (σ^2 in Chapter IV was denote by θ^2) is constant over cells. Generally speaking, one cannot use $\hat{\phi}$ fitting Williams' overdispersion model to obtain an estimate of σ^2 for Pierce's overdispersion model. One could certainly place an upper bound on ϕ such as $\phi \leq 0.25\sigma^2$ since $\pi_{ij}(1-\pi_{ij}) \leq 0.25$.

Using a second order approximation for g(z) we find

$$p_{ij} \cong \pi_{ij}\left[1+\sigma(1-\pi_{ij})z_{ij}+\sigma^2(1-\pi_{ij})(1-2\pi_{ij})z^2_{ij}\right]$$

with

$$E(P_{ij}) \cong \pi_{ij}\left[1+\sigma^2(1-\pi_{ij})(1-2\pi_{ij})\right]$$

and

$$\text{Var}_{Z_{ij}}(P_{ij}) \cong \sigma^2\pi^2_{ij}(1-\pi_{ij})^2\left[1+2\sigma^2(1-2\pi_{ij})^2\right].$$

Thus the the correlation between Y_{ijk} and $Y_{ijk'}$ is given by

$$\sigma^2\pi_{ij}(1-\pi_{ij})\left\{\frac{\left[(1-\pi_{ij})+2\sigma^2(1-\pi_{ij})(1-2\pi_{ij})^2\right]}{\left[1+\sigma^2(1-\pi_{ij})(1-2\pi_{ij})\right]\left[1+\pi_{ij}[1+\sigma^2(1-\pi_{ij})(1-2\pi_{ij})]\right]}\right\}.$$

Type II models can easily accommodate covariates at the binomial level (i.e., all the binary observations for the same cell ij share the same values of the given covariates). As we have seen in Chapters III and IV, Type II models can be extended more easily to cases where there are several levels

of overdispersion.

The methodology used for fitting both types of models is iterative weighted least squares and both fittings methodologies resemble the quasi-likelihood methodology. The estimates have good asymptotic properties such as: being asymptotically unbiased, consistent and asymptotically normal as shown for a variety of settings by several authors mentioned in Chapter II. Further comparisons of sorts will be discussed below based on the results from applying both models to a genetic study.

<div align="center">Application of Models to a Genetic Study</div>

A genetic experiment carried out by Oklahoma State Agricultural Experiment Station and in particular by Dr. Charles Tauer involved studies attempting to identify tree families and seed sources that best adapt to our state's conditions. Selling Christmas trees presents a very good opportunity as an alternative crop for the state's growers.
Before we discuss the experiment further we would like to mention here that discussions along the lines of Mandel (1983) may be appropriate with respect to the application of our models on these data. By this we mean that the data could be analyzed directly and more appropriately by other models. We believe, as Mandel and his discussants pointed out, that in an attempt to illustrate a methodology using real data in appropriate format albeit with an inappropriate context, it is still reasonable.

This particular experiment involved is a survival study of many families of Virginia Pine trees. The trees have been grown in five different locations from which the first three were private growers in different locations within the state. We are also given the different states (11 to be specific) and different stands within a state (ranging from 1-6). The seed was planted in a nursery for a one year period, after which the seedlings

were transplanted in the different locations. The response variable considered here is survival after a year's time in the different locations. It is worth mentioning that the data is being pooled over families (i.e. trees) and reps in the different locations. In our context we think of our two-way table as consisting of location and "source" where "source" is a specific state by stand combination. The total number of sources for each location is 39. A complete listing of the data is given in Appendix A. One could apply both the models of Chapter III and IV by considering both the locations and the sources as random, or locations as fixed and sources as random. The fact that the data is pooled over two other potential sources of variation (reps and families) justifies expecting overdispersion.

An overdispersion model such as the one discussed in Chapter III is fitted to the entire data set and the results of the analysis when treating both effects as random and is summarized in Table IV.

In Figure 4, the column EXVR and MEAN represent the estimates of extra-binomial variations (overdispersions) and means respectively from the individual fits and the overall fit along the lines of Chapter III.

Attempting to answer some of the investigators questions, we highlighted in Table IV what we think is probably the "best state sources" (as far as high average survival rate and low variability). These happen to be sources from Tennessee and North Carolina (i.e., Sources: 19-23, and 24-28 respectively). Sources 18 and 23 (which are in Kentucky and Tennessee, respectively), seem to be the "best sources" using the same criteria as above. It is clear that the first three locations, (which were carried out by private owners whereas the last two locations were carried out by the investigator), seem to have lower average survival and higher variability.

	EXVR	MEAN			
OVER	.177	.742			

	EXVR	MEAN
LOC1	.101	.437
LOC2	.081	.814
LOC3	.064	.780
LOC5	.039	.846
LOC6	.028	.792

	EXVR	MEAN		EXVR	MEAN		EXVR	MEAN
SRC1	.276	.649	SRC14	.879	.294	SRC27	.601	.805
SRC2	.532	.705	SRC15	.027	.732	SRC28	.274	.656
SRC3	.300	.703	SRC16	.090	.765	SRC29	.128	.735
SRC4	.400	.700	SRC17	.223	.677	SRC30	.384	.740
SRC5	.524	.589	SRC18	.048	.861	SRC31	.189	.768
SRC6	.209	.461	SRC19	0	.806	SRC32	.092	.846
SRC7	.092	.814	SRC20	.049	.665	SRC33	.294	.852
SRC8	.411	.722	SRC21	.045	.812	SRC34	.031	.753
SRC9	.174	.706	SRC22	.096	.652	SRC35	.211	.790
SRC10	.111	.806	SRC23	.056	.862	SRC36	.373	.807
SRC11	.089	.762	SRC24	.084	.804	SRC37	.244	.694
SRC12	.082	.681	SRC25	.195	.757	SRC38	0	.787
SRC13	.571	.824	SRC26	.021	.787	SRC39	.075	.685

$\hat{\phi}_1 = 0.467925$

$\hat{\phi}_2 = 0$ (ACTUALLY IT WAS NEGATIVE -0.07)

$\hat{\phi}_3 = 0.0876216$

Figure 4. Results from Analysis Based on the Tree
Data Using Type I Models (Both Effects
Are Considered Random).

The overall estimate of ϕ_2 turns out to be zero and the estimate of ϕ_3, that is, the overdispersion, turns out to be 0.0876. In the context of this study both these estimates are interpreted as variance components. Thus, ϕ_2 measures variability between families within sources (i.e., tree to tree variability) as opposed to ϕ_3 which measures variability within families and among reps within locations. The SAS code that produced the above results is given in Appendix B. Estimates of the variances and standard errors of the overall ϕ's were not obtained since simulating 100 data sets and refitting the model can be very costly and CPU intensive for such a large data set (remember that one of the effects has 39 levels).

We will discuss the results provided in Table I concerning the variability of the ϕ's in a simulated small two-way table situation. We simulated 100 (5x4) data sets for a given set of ϕ's and observed the mean estimate of the ϕ's and their variances (bottom part of the table). We see that, on the average, the estimates of the ϕ's are "good". Keep in mind that we are dealing with a variance component problem and that negative estimates of ϕ's were truncated to zero. The variances could be considered relatively large (that is, in comparison with the estimated variances of the $\hat{\theta}$'s that we calculated when we fitted Type II models to the data). We probably have bias in the estimates and their variances since these estimates are based on a small number of levels for the random effects. Notice that the estimates of the variances goes down as the number of levels increases.

Before discussing the results of applying models Type II to the tree data, we will mention some of the difficulties encountered in the computations due to the nature of the data. First, it was impossible for us to consider analyzing the data by considering both effects as random and

accommodate an overdispersion parameter. The reason is that when we expanded the data set three times (once for each of the variance components) we exceeded the maximum memory that a user can allocate at our installation. Even when we considered the mixed case (that is considering the location fixed and the sources as random) we were only able to analyze three locations.

Another complication arose from the fact that the binomial denominators become too large when we included the last two locations in the analysis and in addition the observed proportions ranged from [0,1]. The n_{i+} (i.e., the total number of trials for a given location) in several instances turned out to be a "very large" number. As a consequence, in calculating the weights for the iterations and the corresponding log likelihood in several instances, we had to evaluate expressions such as

$$\left[\frac{p_{ij}}{(1-p_{ij})} \right]^{r_{ij}} (1-p_{ij})^{n_{ij}}.$$

These calculations for very small or large values of p_{ij} resulted in weights being far too small (zero as far as the machine was concerned). As a result of that, in subsequent iterations, observations that had zero weights were excluded.

We were able to analyze portions of the data. In particular, the first three locations were analyzed by considering location fixed, sources random and allowing overdispersion. Because of our inability to use Type II models when considering the whole data set, even for the mixed model with two variance components, the results that we found cannot really be compared to the results of fitting Type I models.

For the two random effects (the source effect and the overdispersion effect) we expanded the data set 3x5 times using three point (K=3)

quadrature for the overdispersion random effect and (L=5) for the source random effect. The estimates of the variances and the three μ_i's (for the three "fixed" locations) are given below:

$$\hat{\mu}_1 = -0.0356 \qquad\qquad \hat{\theta}_1 = 0.5591$$
$$\hat{\mu}_2 = -0.0356 + 1.9032 \qquad \text{and}$$
$$\hat{\mu}_3 = -0.0356 + 1.7369 \qquad\qquad \hat{\theta}_2 = 0.4551$$

The first thing that strikes us from the above estimates is the estimate of $\hat{\theta}_1$. Therefore the estimates of the variance components are: $\hat{\theta}_1^2 \cong 0.3125$ and $\hat{\theta}_2^2 \cong 0.2071$. Both estimates (i.e., the source and overdispersion variance components) seem small, when compared using the upper bound rule to the equivalent estimates in Figure 4. Let us not forget though that the results on Figure 4 are based on analysis using Type I models in the random case for the entire data as opposed to a mixed model on the first three locations. One also has to keep in mind that the upper bound rule was verified only for the overdispersion models when all other effects were considered fixed. The variances and covariances of the parameter estimates given above obtained by inverting the observed information matrix and are given below:

I IM	LOC1	LOC2	LOC3	OVER	SRVC
LOC1	0.028794	-0.006639	-0.018459	0.007976	-0.014127
LOC2	-0.006639	0.056793	0.008982	0.000254	-0.010140
LOC3	-0.018459	0.008982	0.066445	0.002690	0.006167
OVER	0.007976	0.000254	0.002690	0.015497	-0.000797
SRVC	-0.014127	-0.010140	0.006167	-0.000797	0.022315

Suggestions for Further Research

The effect of using different number of quadrature points on the estimates still needs to be addressed. Also an extensive simulation study with smaller binomial denominators on a smaller size data set is needed. An alternative weighting technique needs to be proposed for data sets with very large denominators. These methods should be compared with more direct methods for maximizing the likelihood.

Also for the models in Chapters III, one needs to investigate the possibility of using the deviance instead of the Pearson's Chi square statistic for calculating the estimates of the the ϕ's from the individual fits.

BIBLIOGRAPHY

ABRAMOWITZ, M., and STEGUN, I., (1972), *Handbook of Mathematical Functions,* New York, Dover.

AITKIN, M., (1987), "Modeling Variance Heterogeneity in Normal Regression Using GLIM," *Applied Statistics,* 36, No. 3, 332-339.

ALTHAM, P. M. E., (1978), "The Generalizations of the Binomial Distribution," *Applied Statistics,* 27, 162-167.

ANDERSON, D. A., and AITKIN, M., (1985), "Variance Component Models with Binary Response: Interviewer Variability," *Journal of Royal Statistical Society, B,* 47, 203-210.

ANSCOMBE, F. J., (1954), "Discussion of Fisher (1954)," *Biometrics,* 10, 141-144.

BARTLETT, M. S., (1954), "Discussion of Fisher (1954)," *Biometrics,* 10, 140-141.

BEITLER, P. J., and LANDIS J. R., (1985), "A Mixed-Effects Model for Categorical Data," *Biometrics,* 41, 991-1000.

BERKSON, J., (1951), "Why I Prefer Logits to Probits," *Biometrics,* 7, 327-339.

BISHOP, Y. M., FIENBERG, S. E., and HOLLAND, P. W., (1975), *Discrete Multivariate Analysis: Theory and Practice,* Cambridge, Mass., The MIT Press.

BOCK, D., and AITKIN, M. (1981) "Marginal Maximum Likelihood Estimation for Item Parameters:Application of the EM Algorithm," Psychometrika, 46, 443-459.459.

BRESLOW, N., (1984), "Extra-Poisson Variation in Log-linear Models," *Applied Statistics,* 33, 38-44.

BRILLINGER, D. R., and PREISLER, M. K., (1983) "Maximum Likelihood Estimation in a Latent Variable," in *Studies in Econometrics, Time Series, and Multivariate Statistics,* eds S. KARLIN, T. AMEMIYA and L.A.GOODMAN, NEW YORK, Academic Press, 31-65.

BRIER, S. S., (1980), "Analysis of Contingency Tables Under Cluster Sampling," *Biometrika,* 67, 591-596.

CHATFIELD, C., and GOODHARDT, G. J., (1970), "The Beta-Binomial Model for Consumer Purchasing Behavior," *Applied Statistics,* 19, 240-250.

COCHRAN, W. G., (1943), "Analysis of Variances for Percentages Based on Unequal Numbers," *Journal of the American Statistical Association*, 38, 287-301.

COCHRAN, W. G., (1954), "Discussion of Fisher (1954)," *Biometrics*, 10, 144-147.

COX, D. R., (1970), *Analysis of Binary Data*, London, Methuen.

COX, D. R., (1983), "Some Remarks on Overdispersion," *Biometrika* 70, 269-274.

CROWDER, M. J., (1978), "Beta-Binomial Anova for Proportions," *Applied Statistics*, 27, 234-37.

CROWDER, M. J., (1979), "Inference About the Intraclass Correlation Coefficient in the Beta-binomial ANOVA for Proportions," *Journal of Royal Statistical Society*, B, 41, No. 2, 230-234.

CROWDER, M. J., (1985), "Gaussian Estimation for Correlated Binomial Data," *Journal of Royal Statistical Society*, B, 47, No. 2, 229-237.

CROWDER, M. J., (1987), "On Linear and Quadratic Estimating Equations," *Biometrika* 74, 3, 591-597.

DEMPSTER A. P., LAIRD, N. M., and RUBIN, D. B., (1977). "Maximum Likelihood from Incomplete Data Via the EM Algorithm (with discussion)," *Journal of Royal Statistical Society*, B, 39, 1-38.

EDWARDS, A. W. F., (1958). "An Analysis of Geissler's Data on the Human Ratio," *Annals of Human Genetics*, 23, 6-15.

FAHRMEIR, L., and KAUFMANN, H., (1985), "Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models," *The Annals of Statistics*, 13, No 1, 342-368.

FINNEY, D. J., (1971), *Probit Analysis*, (Third Edition) London, Cambridge University Press.

FIRTH, D., (1987), "On the efficiency of Quasi-Likelihood Estimation," *Biometrika*, 74, 233-245.

FISHER, R. A., (1954), "The Analysis of Variance with Various Binomial Transformations," *Biometrics*, 10, 130-139.

GILMOUR, A. R., ANDERSON, R. D., and RAE, A. L. (1985), "The Analysis of Binomial Data by a Generalized Linear Mixed Model," *Biometrika*, 72, 593-599.

GLADEN, B., (1979), "The Use of the Jacknife to Estimate Proportions From Toxicological Data in the Presence of Litter Effects," *Journal of the American Statistical Association*, 74, 278-283.

GREEN, P. J., (1984), "Iteratively Reweighted Least Squares for Maximum Likelihood Estimation and some Robust and Resistant Alternatives (with discussion)," *Journal of Royal Statistical Society, B, 46,* No. 2, 149-192.

GRIFFITHS, D. A., (1973), "Maximum Likelihood Estimation for the Beta-Binomial Distribution and an Application to the Household Distribution of the Total Number of Cases of a Disease," *Biometrics,* 29, 637-648.

GRIZZLE J. E., STAMNER C. F., and KOCH, G. G., (1969), "Analysis of Categorical Data by Linear Models," *Biometrics,* 25, 489-504.

HARVILLE, D. A., and MEE, R. W., (1984), "A Mixed-Model Procedure for Analyzing Ordered Categorical Data," *Biometrics,* 40, 393-408.

HASEMAN, J. K., and KUPPER, L. L., (1979), "Analysis of Dichotomous Response Data form Certain Toxicological Experiments," *Biometrics,* 35, 281-293.

HILL, J. R., and TSAI C., (1988) "Calculating the Efficiancy of Maximum Quasilikelihood Estimation," *Applied Statistics,* 37, No. 2, 219-230.

HINDE, J., (1982), "Compound Poisson Regression Models," in *GLIM 82: Proceedings of the International Conference in Generalized Linear Models,* 109-121. Springer-Verlag, New York 1982.

HUBER, P. J., (1967), "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions," in *Proceedings Fifth Berkley Symposium Mathematical Statistics and Probability,* vol. 1, 221-233.

IM, S., and GIANOLA (1988), "Mixed Models For Binomial Data with an Application to Lamb Mortality," *Applied Statistics,* 37, No 2, 196-204.

JORGENSEN, B., (1983), "Maximum Likelihood Estimation and Large Sample Inference for Generalized Linear Models and Nonlinear Regression Models," *Biometrika* 70, 1, 19-28.

KEMPTHORNE, J. W., (1982), "The Analysis of Clustered Attribute Data from Nested and/or Classification Designs using Random Effects Models," *University of North Carolina at Chapel Hill, Institute of Statistics, Mimeo Series No. 1429.*

KEMPTHORNE, J. W., and KOCH, G. G., (1983), "Statistical Method for Binary Data from a Set of Partially Structure Two-Way Random Cross-Classification," *Communication in Statistics-Theory and Methods,* 12, 2761-2777.

KLEINMAN, J. C., (1973), "Proportions with Extraneous Variance: Single and Independent Samples," *Journal of the American Statistical Association,* 68, 46-54.

KLEINMAN, J. C., (1975), "Proportions with Extraneous Variance: Two Dependent Samples," *Biometrics,* 31 737-743.

KUPPER, L. L., and HASEMAN, J. K., (1978) "The Use of the Correlated Binomial Model for the Analysis of Certain Toxicological Experiments," *Biometrics,* 34, 69-76.

LAIRD, N. M., WARE, J. H., (1982), "Random-Effects Models for Longitudinal Data," *Biometrics,* 38, 963-974.

LAIRD, N. M., (1978), "Empirical Bayes Methods for Two-Way Contingency Tables," *Biometrika,* 65, 581-590.

LEONARD, T., (1975), "Bayesian Estimation Methods for Two-Way Contigency Tables," *Journal of the Royal Statistical Society, Series B,* 37, 23-37.

MANTEL, N., (1983), "Creation of a Statistical Problem," *The American Statistician,* Vol 39, No 4 , 334-336.

McCULLAGH, P., (1983), "Quasi-Likelihood Functions," *The Annals of Statistics,* 11, 59-67.

McCULLAGH, P., and NELDER, J., A., (1983), *Generalized Linear Models* London, Chapman and Hall.

MOORE, D. F., (1986), "Asymptotic Properties of Moment Estimators for Overdispersed Counts and Proportions," *Biometrika,* 73, 583-588.

MOORE, D. F., (1987), "Modeling the Extraneous Variance in the Presence of Extra-Binomial Variation," *Applied Statistics,* No. 1, 8-14.

MOSER, B. K., (1985), *Missing Data in Linear Models with Constant Interclass Correlation and Random Block Designs with Binary Data,* Unpublished Ph.D. Dissertation, Purdue University.

MOSIMANN, J. E., (1962) "On the Compound Multiracial, the Multivariate β Distribution, and Correlation Between Proportions," *Biometrika,* 49, 65-69.

NELDER, J. A., and WEDDERBURN, R. W. M., (1972), "Generalized Linear Models," *Journal of the Royal Statistical Society, Series A,* 135, 370-384.

NELDER, J. A., (1985), "Quasi-Likelihood and GLIM," *GLIM 85 Proceedings* 120-127. Springer-Verlag, New York.

NELDER, J. A., and PREGIBON, D., (1987), "An Extended Quasi-Likelihood Function," *Biometrika,* 74, 221-232.

OTAKE M., and PRENTICE, R., (1984), "The Analysis of Chromosomally Aberrant Cells Based on Beta-Binomial Distribution," *Radiation Research,* 98, 456-470.

PAUL, S. R., (1982), "Analysis of Proportions of Affected Fetuses in Teratological Experiments," *Biometrics*, 38, 361-370.

PIERCE, D. A., and SANDS, B. R., (1975), "Extra-Bernoulli Variation in Binary Data." *Technical Report 46, Oregon State University, Dept. of Statistics.*

PIERCE, D. A., (1976), "A Random Effects Model for Matched Pairs of Binomial Data," *Technical Report 55, Oregon State University, Dept. of Statistics.*

SHIRLEY, E. A. C., and HICKLING R., (1981), "An Evaluation of some Statistical Methods for Analyzing Numbers of Abnormalities Found Amongst Liters in Teratology Studies," *Biometrics*, 37, 819-829.

SKELLAM, J. G., (1948), "A Probability Distribution Derived for the Binomial Distribution by Regarding the Probability of Success as Variable Between the Sets of Trials," *Journal of the Royal Statistical Society, Ser. B,* 10, 257-261.

SMITH, D. M., (1983), "Maximum Likelihood Estimation of the Parameters of the Beta Binomial Distribution," *Applied Statistics*, 32, No. 2, 196-204.

SMYTH, G. K., (1987), "Modeling the Dispersion in the Generalized Linear Models," *Technical Report No. 10, University of Santa Barbara, Dept. of Statistics..*

STIRATELLI, R., LAIRD, N. M., and WARE, J. H., (1984), "Random-Effects Models for Serial Observations with Binary Response," *Biometrics*, 40, 961-971.

STIRLING, W. D., (1984), "Iteratively Reweighted Least Squares for Models with a Linear Part," *Applied Statistics*, 33, No. 1, 7-17.

STIRLING, W. D., (1985), "Heteroscedastic Models and an Application to Block Designs," *Applied Statistics*, 34, No. 1, 33-41.

STOKES, L., (1988), "Estimation of Interviewer Effects for Categorical Items in a Random Digit Telephone Survey," *Journal of the American Statistical Asssociation*, Vol 83, No 403, 623-630.

STROUD, A. H., and SECREST, D., (1966), *Gaussian Quadrature Formulas,* Englewood Cliffs (N.J.): Prentice Hall.

TORNQVIST, L., VARTIA, P., and VARTIA, Y. O., (1985), "How Should Relative Changes be Measured?" *American Statistician*, 39, 43-46.

WEDDERBURN, R. M. W., (1974), "Quasi-likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method," *Biometrika*, 61, 439-447.

WILLIAMS, D. A., (1975), "The Analysis of Binary Responses from Toxicological Experiments Involving Reproduction and Teratogenicity," *Biometrics*, 31, 949-952.

WILLIAMS, D. A., (1982), "Extra-Binomial Variation in Logistic Linear Models," *Applied Statistics*, 31, No. 2, 144-148.

WHITTLE, P., (1961) "Gaussian estimation in Stationary Time Series," *Bulletin of International Statistical Institute*, 39, 1-26.

APPENDIX A

VIRGINIA PINE TREE DATA SET LISTING

VIRGINIA PINE TREE DATA

LOCATION	STATED	STAND	TOTAL COUNT	# SURVIVE
1	2	1	24	6
1	2	2	40	5
1	2	3	24	7
1	2	4	24	5
1	4	2	16	0
1	4	3	6	2
1	4	4	6	3
1	4	5	20	4
1	8	1	22	8
1	8	2	8	4
1	8	3	72	38
1	9	1	24	10
1	9	2	8	3
1	10	1	32	15
1	10	2	28	16
1	10	3	32	17
1	10	4	40	11
1	10	5	16	12
1	11	1	8	6
1	11	2	24	10
1	11	3	32	27
1	11	5	24	11
1	11	7	24	17
1	12	1	16	9
1	12	2	14	6
1	12	4	40	27
1	12	5	16	4
1	12	6	14	3
1	13	1	40	18
1	13	2	24	6
1	13	3	16	7
1	13	4	44	31
1	13	5	24	12
1	13	6	8	4
1	14	1	32	14
1	14	2	8	3
1	14	3	16	4

1	15	1	16	11
1	15	2	40	17
2	2	1	24	21
2	2	2	32	26
2	2	3	24	15
2	2	4	20	19
2	4	2	16	11
2	4	3	6	0
2	4	4	6	5
2	4	5	20	17
2	8	1	22	16
2	8	2	8	6
2	8	3	72	55
2	9	1	24	20
2	9	2	8	8
2	10	1	32	28
2	10	2	40	28
2	10	3	32	26
2	10	4	40	33
2	10	5	16	15
2	11	1	8	5
2	11	2	16	12
2	11	3	24	23
2	11	5	22	18
2	11	7	24	24
2	12	1	16	15
2	12	2	14	12
2	12	4	40	32
2	12	5	16	16
2	12	6	14	12
2	13	1	40	29
2	13	2	24	19
2	13	3	16	11
2	13	4	40	37
2	13	5	20	20
2	13	6	8	7
2	14	1	32	27
2	14	2	7	5
2	14	3	16	13
2	15	1	16	13
2	15	2	40	30
3	2	1	24	12
3	2	2	36	26
3	2	3	24	18
3	2	4	24	15
3	4	2	16	8
3	4	3	6	3
3	4	4	6	6
3	4	5	24	18
3	8	1	22	15
3	8	2	8	7
3	8	3	72	58
3	9	1	24	18
3	9	2	8	8
3	10	1	32	22

3	10	2	40	29
3	10	3	32	27
3	10	4	40	33
3	10	5	16	15
3	11	1	8	8
3	11	2	16	11
3	11	3	32	27
3	11	5	22	18
3	11	7	23	21
3	12	1	16	12
3	12	2	14	14
3	12	4	40	35
3	12	5	16	15
3	12	6	14	8
3	13	1	40	34
3	13	2	24	22
3	13	3	16	15
3	13	4	44	41
3	13	5	24	23
3	13	6	8	5
3	14	1	32	28
3	14	2	8	8
3	14	3	16	13
3	15	1	16	13
3	15	2	40	29
5	2	1	111	90
5	2	2	124	118
5	2	3	92	88
5	2	4	114	102
5	4	2	48	45
5	4	3	24	16
5	4	4	24	19
5	4	5	48	43
5	8	1	71	64
5	8	2	24	23
5	8	3	216	187
5	9	1	67	50
5	9	2	24	22
5	10	1	110	85
5	10	2	120	101
5	10	3	96	86
5	10	4	119	90
5	10	5	48	45
5	11	1	28	23
5	11	2	48	34
5	11	3	96	77
5	11	5	72	50
5	11	7	72	61
5	12	1	48	41
5	12	2	43	35
5	12	4	120	88
5	12	5	48	46
5	12	6	48	36
5	13	1	120	105
5	13	2	72	67

5	13	3	48	43
5	13	4	120	113
5	13	5	72	67
5	13	6	24	20
5	14	1	96	87
5	14	2	24	23
5	14	3	48	36
5	15	1	48	36
5	15	2	120	88
6	2	1	97	76
6	2	2	100	91
6	2	3	60	52
6	2	4	95	76
6	4	2	40	32
6	4	3	20	13
6	4	4	20	18
6	4	5	40	36
6	8	1	60	49
6	8	2	20	17
6	8	3	180	150
6	9	1	60	39
6	10	1	95	73
6	10	2	100	74
6	10	3	80	58
6	10	4	100	70
6	10	5	40	30
6	11	1	20	16
6	11	2	40	30
6	11	3	80	54
6	11	5	60	30
6	11	7	60	51
6	12	1	40	35
6	12	2	35	24
6	12	4	100	85
6	12	5	40	35
6	12	6	40	34
6	13	1	100	76
6	13	2	60	48
6	13	3	40	34
6	13	4	100	73
6	13	5	60	52
6	13	6	20	16
6	14	1	80	70
6	14	2	20	19
6	14	3	40	33
6	15	1	40	34
6	15	2	100	77

APPENDIX B

PROGRAM LISTING FOR TYPE I MODELS

```
//U12558AA   JOB  (12558,AND-RO-NIKO),'BISIM',CLASS=4,TIME=(30,00),
//   MSGCLASS=X,NOTIFY=*,MSGLEVEL=(1,1)
/*PASSWORD ?
/*ROUTE  PRINT  LOCAL
/*JOBPARM  ROOM=V,FORMS=9001
//     EXEC  SASN,REGION=8500K,OPTIONS='MACRO'
//FT20F001 DD  DUMMY
//LIB  DD  DSN=&&TEMP,SPACE=(CYL,(60,1)),UNIT=VIO
//AM   DD  DSN=U12558A.T15LOC.SAS.DATA,DISP=OLD
//CT   DD  DSN=U12558A.BIN1.CNTL(CTREE5L),DISP=OLD
//SYSIN DD  *
   OPTIONS DQUOTE NOMPRINT NOMACROGEN   NOSYMBOLGEN;

%LET  NADS=DS;
%LET  N= 1 ;                          /*     SOS     */
%LET  NR=5;
%LET  NC=39;


TITLE 'ANALYSIS  OF  CHRISTMAS  TREE  DATA  (ALL  5  LOCATIONS)';

DATA L3;INFILE CT;
   INPUT LOC STATE STAND FAM R M ;
   KEEP DSI LOC FAM R M Q;
   Q=M-R;
   DSI=1;
PROC PRINT;


DATA AM.GALL3DS;SET L3;
   KEEP DSI II JI R M Q;
   II=LOC;
   JI=FAM;


DATA LIB.GALL;SET AM.GALL3DS;
PROC PRINT;

%MACRO GEN;
   %DO KI=1 %TO &N;
      DATA LIB.&NADS&KI;SET LIB.GALL;
         XX=SYMGET('KI'); IF XX=DSI;
         DROP XX;
         INT=1;RHO=0;ESTPAR=.5;
         PROC PRINT;
      %DO LI=1 %TO &NR;
         DATA LIB.&NADS&KI.R&LI;SET LIB.&NADS&KI;
         VV=II;
         YY=SYMGET('LI'); IF YY=VV;
         DROP YY VV;
         INT=1;RHO=0;ESTPAR=.5;
         PROC PRINT;
      %END;
      %DO MI=1 %TO &NC;
         DATA LIB.&NADS&KI.C&MI;SET LIB.&NADS&KI;
         WW=JI;
         ZZ=SYMGET('MI'); IF ZZ=WW;
         DROP ZZ WW;
```

```
            INT=1;RHO=0;ESTPAR=.5;
            PROC PRINT;
        %END;
    %END;
%MEND GEN;


%GEN


%LET N=1   ;                            /*    SOS    */
PROC PRINTTO UNIT=20;

%MACRO BINOMIAL(DATA=,RESPONSE=,NUMBER=,VARS=);

/*- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -*/
/* VARIABLE      FUNCTION                                          */
/* - - - - - - - -    - - - - - - - -                             */
/*                                                                 */
/* DATA          INPUT DATA SET                                   */
/* RESPONSE      VARIABLE CONTAINING THE NUMBER OF RESPONDANTS */
/* NUMBER        VARIABLE CONTAINING THE NUMBER IN GROUP        */
/* VARS          LIST OF INDEPENDENT VARIABLES                   */
/* P             RESPONSE PROB AS FUNCTION OF Z (=XB)            */
/* PHI           DERIVATIVE OF P AS A FUNCTION OF Z AND/OR P    */
/*                                                                 */
/*- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -*/
/* RESPONSE~BIN(NUMBER,P)                                          */
/* E(RESPONSE)=NUMBER*P                                            */
/* Z=XB                                                            */
/* LOGIT  LINK FUNCTION  P=1/(1+EXP(-Z))                          */
/* PROBIT LINK FUNCTION  P=PROBNORM(Z)                            */
/* PHI=DERIVATIVE   P/WRT(Z)                                      */
/* FOR LOGIT  LINK  PHI=NUMBER*P*(1-P)                            */
/* FOR PROBIT LINK  PHI=NUMBER*EXP(-Z*Z/2)/SQRT(8*ATAN(1))       */
/* VAR(RESPONSE)=NUMBER*P*(1-P)                                   */
/* MODEL RESPONSE=NUMBER*P                                        */
/* WI=(1+(RHO*(NUMBER-1)))                                        */
/* _WEIGHT_=(VAR(RESPONSE))*(1/WI)=((NUMBER*P*(1-P))*(1/WI)      */
/* _LOSS_= (-RESPONSE*LOG(P)-(NUMBER-RESPONSE)*LOG(1-P))         */
/*                 /_WEIGHT_                                       */
/* DER.B=PHI*DER(Z)/WRT(B)                                        */
/*                                                                 */
/*- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -*/

        %LET N=0;        /* SPLIT OUT INDIVIDUAL NAMES */
        %LET OLD=;
        %DO %WHILE(%SCAN(&VARS,&N+1)^=);
            %LET N=%EVAL(&N+1);
            %LET VAR&N=%SCAN(&VARS,&N);
            %LET OLD=&OLD _OLD&N;
        %END;
        /* DO MLE WITH NONLINEAR LEAST SQUARES */

    PROC PRINTTO UNIT=20 NEW;
```

```
PROC NLIN NOHALVE SIGSQ=1 DATA=&DATA(RENAME=(
    %DO I=1 %TO &N;
        &&VAR&I=_OLD&I
    %END; )) OUTEST=W2OUT;
    RETAIN LOGLIKE 0;
    /* START INITIAL VALUES AT ZERO */
    PARMS
        INTERCPT=0
        %DO I=1 %TO &N; &&VAR&I=0 %END; ;
    /* COMPUTE INNER PRODUCT */
    Z=INTERCPT %DO I=1 %TO &N; + &&VAR&I*_OLD&I %END; ;
    /* MODEL RESPONSE PROBABILITY: CHANGE THIS FOR DIFFERENT MODEL */
    *P=PROBNORM(Z);        /*PROBIT REGRESSION*/
    P=1/(1+EXP(-Z));    /*LOGIT   REGRESSION*/
    IF _MODEL_=1 THEN DO;
        IF _OBS_=1 THEN DO; PUT LOGLIKE=; LOGLIKE=0; END;
        LOGLIKE= LOGLIKE +
                &RESPONSE*LOG(P)+(&NUMBER-&RESPONSE)*LOG(1-P);
    END;
    MODEL &RESPONSE=&NUMBER*P;
    WI=(1+(RHO*(&NUMBER-1)));
    V=(&NUMBER*P*(1-P));
    _WEIGHT_=(1/WI)*(1/V);
    _LOSS_=(-&RESPONSE*LOG(P)-(&NUMBER-&RESPONSE)*LOG(1-P))/_WEIGHT_;
    /* CHANGE THIS FOR DIFFERENT PROBABILITY MODEL */
    *PHI=&NUMBER*EXP(-Z*Z/2)/SQRT(8*ATAN(1)); /* PROBIT REGRESSION */
    PHI=&NUMBER * P * (1-P);                          /* LOGIT   REGRESSION */
    DER.INTERCPT=PHI;
    %DO I=1 %TO &N;
        DER.&&VAR&I=PHI*_OLD&I;
    %END;
%MEND BINOMIAL;

%MACRO DOIT(DSN);
  %DO K =1 %TO  5 ;
    %PUT ITERATION &K;
    %BINOMIAL(DATA=&DSN,RESPONSE=R,NUMBER=M,VARS=%STR( ))
    DATA X;SET &DSN;
      KEEP R M INT RHO ESTPAR;
    DATA OUT;SET W2OUT;
      DROP _TYPE_ _NAME_ _ITER_;
      IF _ITER_=.;
    PROC MATRIX ;
      FETCH XALL DATA=X;
      R=XALL(,1);
      N=NROW(R);
      NC=NCOL(XALL);
      RHOV=XALL(,NC-1);
      RHO=RHOV(1,);
      OP=J(N,1,1);
      M=XALL(,2);
      X=XALL(,3:NC-2);
      FETCH OUTMAT DATA=OUT;
      NPAR=NCOL(OUTMAT);
      DFM=NPAR-1;
```

```
           EST=OUTMAT(1,);
           DFE=N-DFM;
           X2=EST(,1);
           MSE=EST(,1)#/DFE;
           PAR=EST(1,2:NPAR);
           BET=PAR  %STR(%');
           Z=X*BET;
           PV=OP#/(OP+EXP(-Z));
           P=PV(1,);
           COVMAT=OUTMAT(2:NPAR, 2:NPAR);
           VARP=COVMAT#/MSE;
           VL=X*VARP*X  %STR(%');
           VL=DIAG(VL);
           VL=VL(,+);


           PW=1#/(OP+(RHO#/(M-OP)));


           WT=(M#P#(OP-P));
           WVQ=PW#WT#VL;
           NEX2=PW#(OP-WVQ);
           SNEX2=NEX2(+,);
           DEX2=((M-OP)#PW#(OP-WVQ));
           SDEX2=DEX2(+,);
           RH=(X2-SNEX2)#/SDEX2;
           RHV=OP#RH;
           WDAT=R||M||OP||RHOV||RHV||PV;
           OUTPUT WDAT OUT=&DSN(RENAME=(COL1=R COL2=M COL3=INT
                                 COL4=RHO COL5=RH COL6=ESTPAR));
           PRINT RHO RH P ;


           DATA &DSN;SET &DSN;
           DROP RHO DIF DIFF ;
           RENAME RH=RHO;
           DIF=ABS(RH-RHO);
           DIFF=INT(DIF*10000);
           IF  RH<0 THEN    RH=0;
           CALL SYMPUT('RHM',RH);
           CALL SYMPUT('DIFFM',DIFF);

           PROC PRINT;

           DATA &DSN.F;SET &DSN;
             KEEP RHO ESTPAR;
             IF _N_=1;
           %IF &RHM<0 OR &DIFFM<1 %THEN %LET K=10;
           %ELSE K=&K;
           %PUT FLAG1 K IS &K RH IS &RHM DIFFM IS &DIFFM;
             ;
        %END;

%OUT: %PUT FLAG2 THE VALUES OF K IS &K ;
%MEND DOIT;

%MACRO ALLDS;
```

```
    %DO L=1 %TO &N ;                                        /* SOS */
    %DOIT(LIB.DS&L)
    %DOIT(LIB.DS&L.R1)
    %DOIT(LIB.DS&L.R2)
    %DOIT(LIB.DS&L.R3)
    %DOIT(LIB.DS&L.R4)
    %DOIT(LIB.DS&L.R5)
    %DOIT(LIB.DS&L.C1)
    %DOIT(LIB.DS&L.C2)
    %DOIT(LIB.DS&L.C3)
    %DOIT(LIB.DS&L.C4)
    %DOIT(LIB.DS&L.C5)
    %DOIT(LIB.DS&L.C6)
    %DOIT(LIB.DS&L.C7)
    %DOIT(LIB.DS&L.C8)
    %DOIT(LIB.DS&L.C9)
    %DOIT(LIB.DS&L.C10)
    %DOIT(LIB.DS&L.C11)
    %DOIT(LIB.DS&L.C12)
    %DOIT(LIB.DS&L.C13)
    %DOIT(LIB.DS&L.C14)
    %DOIT(LIB.DS&L.C15)
    %DOIT(LIB.DS&L.C16)
    %DOIT(LIB.DS&L.C17)
    %DOIT(LIB.DS&L.C18)
    %DOIT(LIB.DS&L.C19)
    %DOIT(LIB.DS&L.C20)
    %DOIT(LIB.DS&L.C21)
    %DOIT(LIB.DS&L.C22)
    %DOIT(LIB.DS&L.C23)
    %DOIT(LIB.DS&L.C24)
    %DOIT(LIB.DS&L.C25)
    %DOIT(LIB.DS&L.C26)
    %DOIT(LIB.DS&L.C27)
    %DOIT(LIB.DS&L.C28)
    %DOIT(LIB.DS&L.C29)
    %DOIT(LIB.DS&L.C30)
    %DOIT(LIB.DS&L.C31)
    %DOIT(LIB.DS&L.C32)
    %DOIT(LIB.DS&L.C33)
    %DOIT(LIB.DS&L.C34)
    %DOIT(LIB.DS&L.C35)
    %DOIT(LIB.DS&L.C36)
    %DOIT(LIB.DS&L.C37)
    %DOIT(LIB.DS&L.C38)
    %DOIT(LIB.DS&L.C39)

DATA AM.FI&L.N1;SET LIB.&NADS&L.F;
    PE=RHO;
    HE=ESTPAR;
    KEEP PE HE;

%DO LI=1 %TO &NR;
DATA AM.FI&L.N2&LI;SET LIB.&NADS&L.R&LI.F;
    QE&LI=RHO;
```

```
AE&LI=ESTPAR;
KEEP QE&LI AE&LI;
%END;


%DO MI=1 %TO &NC;
DATA AM.FI&L.N1&MI;SET LIB.&NADS&L.C&MI.F;
  ME&MI=RHO;
  BE&MI=ESTPAR;
  KEEP ME&MI BE&MI;
  %END;


PROC PRINTTO;

DATA AM.FIN&L;MERGE AM.FI&L.N1
                         AM.FI&L.N21  AM.FI&L.N22  AM.FI&L.N23
                         AM.FI&L.N24  AM.FI&L.N25
                         AM.FI&L.N11  AM.FI&L.N12  AM.FI&L.N13
                         AM.FI&L.N14  AM.FI&L.N15  AM.FI&L.N16
                         AM.FI&L.N17  AM.FI&L.N18  AM.FI&L.N19
                         AM.FI&L.N110  AM.FI&L.N111  AM.FI&L.N112
                         AM.FI&L.N113  AM.FI&L.N114  AM.FI&L.N115
                         AM.FI&L.N116  AM.FI&L.N117  AM.FI&L.N118
                         AM.FI&L.N119  AM.FI&L.N120  AM.FI&L.N121
                         AM.FI&L.N122  AM.FI&L.N123  AM.FI&L.N124
                         AM.FI&L.N125  AM.FI&L.N126  AM.FI&L.N127
                         AM.FI&L.N128  AM.FI&L.N129  AM.FI&L.N130
                         AM.FI&L.N131  AM.FI&L.N132  AM.FI&L.N133
                         AM.FI&L.N134  AM.FI&L.N135  AM.FI&L.N136
                         AM.FI&L.N137  AM.FI&L.N138  AM.FI&L.N139  ;
  ZE=((QE1*(AE1*(1-AE1)))
     +(QE2*(AE2*(1-AE2)))
     +(QE3*(AE3*(1-AE3)))
     +(QE4*(AE4*(1-AE4)))
     +(QE5*(AE5*(1-AE5))))  /  (5*(HE*(1-HE)));
  VE= ((ME1*(BE1  *(1-BE1)))
     +  (ME2*(BE2  *(1-BE2)))
     +  (ME3*(BE3  *(1-BE3)))
     +  (ME4*(BE4  *(1-BE4)))
     +  (ME5*(BE5  *(1-BE5)))
     +  (ME6*(BE6  *(1-BE6)))
     +  (ME7*(BE7  *(1-BE7)))
     +  (ME8*(BE8  *(1-BE8)))
     +  (ME9*(BE9  *(1-BE9)))
     +(ME10*(BE10*(1-BE10)))
     +(ME11*(BE11*(1-BE11)))
     +(ME12*(BE12*(1-BE12)))
     +(ME13*(BE13*(1-BE13)))
     +(ME14*(BE14*(1-BE14)))
     +(ME15*(BE15*(1-BE15)))
     +(ME16*(BE16*(1-BE16)))
     +(ME17*(BE17*(1-BE17)))
     +(ME18*(BE18*(1-BE18)))
     +(ME19*(BE19*(1-BE19)))
     +(ME20*(BE20*(1-BE20)))
```

```
                    +(ME21*(BE21*(1-BE21)))
                    +(ME22*(BE22*(1-BE22)))
                    +(ME23*(BE23*(1-BE23)))
                    +(ME24*(BE24*(1-BE24)))
                    +(ME25*(BE25*(1-BE25)))
                    +(ME26*(BE26*(1-BE26)))
                    +(ME27*(BE27*(1-BE27)))
                    +(ME28*(BE28*(1-BE28)))
                    +(ME29*(BE29*(1-BE29)))
                    +(ME30*(BE30*(1-BE30)))
                    +(ME31*(BE31*(1-BE31)))
                    +(ME32*(BE32*(1-BE32)))
                    +(ME33*(BE33*(1-BE33)))
                    +(ME34*(BE34*(1-BE34)))
                    +(ME35*(BE35*(1-BE35)))
                    +(ME36*(BE36*(1-BE36)))
                    +(ME37*(BE37*(1-BE37)))
                    +(ME38*(BE38*(1-BE38)))
                    +(ME39*(BE39*(1-BE39))))
                                              /  (39*(HE*(1-HE)));
   GE=(PE-ZE)*4;
   IE=(PE-VE)*4;
   JE=(ZE-PE+VE)  /  (1-(PE-ZE)-(PE-VE))  ;
PROC  PRINT;
%END;
%MEND  ALLDS;


%ALLDS

%MACRO  FINALS;
   %DO  OO=1  %TO  &N ;                                    /* SOS */
     AM.FIN&OO
   %END;
%MEND  FINALS;


DATA  AM.FINAL  ;SET  %FINALS ;
   RENAME  GE=PHI1;
   RENAME  IE=PHI2;
   RENAME  JE=PHI3;
PROC  PRINT  DATA=AM.FINAL ;
//
```

APPENDIX C

PROGRAM LISTING FOR TYPE II MODELS (1)

```
//U12558AA   JOB  (12558,AND-RO-NIKO),'BIOSIM',CLASS=4,TIME=(2,59),
//   MSGCLASS=X,NOTIFY=*,MSGLEVEL=(1,1)
/*PASSWORD  ?
/*ROUTE  PRINT  LOCAL
/*JOBPARM  ROOM=V,FORMS=9001
//     EXEC  SAS,REGION=2500K,OPTIONS='MACRO'
//SYSIN  DD  *
OPTIONS  DQUOTE  MPRINT  SYMBOLGEN;

%MACRO  BINOMIAL(DATA=,RESPONSE=,NUMBER=,WEIGHT=,VARS=);

   /*----------------------------------------------------------*/
   /* VARIABLE       FUNCTION                                  */
   /* --------       --------                                  */
   /*                                                          */
   /* DATA           INPUT DATA SET                            */
   /* RESPONSE       VARIABLE CONTAINING THE NUMBER OF RESPONDANTS */
   /* NUMBER         VARIABLE CONTAINING THE NUMBER IN GROUP   */
   /* VARS           LIST OF INDEPENDENT VARIABLES             */
   /* P              RESPONSE PROB AS FUNCTION OF Z (=XB)       */
   /* PHI            DERIVATIVE OF P AS A FUNCTION OF Z AND/OR P */
   /*                                                          */
   /*----------------------------------------------------------*/

       %LET  N=0;          /* SPLIT OUT INDIVIDUAL NAMES */
       %LET  OLD=;
       %DO  %WHILE(%SCAN(&VARS,&N+1)^=);
           %LET  N=%EVAL(&N+1);
           %LET  VAR&N=%SCAN(&VARS,&N);
           %LET  OLD=&OLD _OLD&N;
       %END;
       /* DO MLE WITH NONLINEAR LEAST SQUARES */

*PROC  PRINTTO  UNIT=20  NEW;

PROC  NLIN  NOHALVE  SIGSQ=1  DATA=&DATA(RENAME=(
    %DO  I=1  %TO  &N;
        &&VAR&I=_OLD&I
    %END; ))  OUTEST=BOIOUT;
    RETAIN  LOGLIKE   TERM2  DEVIANCE  0;
    /* START INITIAL VALUES AT ZERO */
    PARMS
        INTERCPT=0
        %DO  I=1  %TO  &N;  &&VAR&I=0  %END;  ;
    /* COMPUTE INNER PRODUCT */
    Z=INTERCPT  %DO  I=1  %TO  &N;  +  &&VAR&I*_OLD&I  %END;  ;
    /* MODEL RESPONSE PROBABILITY */
    P=1/(1+EXP(-Z));

    IF  _MODEL_=1  THEN  DO;
        IF  _OBS_=1  THEN  DO;
            PUT  LOGLIKE=;
            LOGLIKE=0;
            TERM2=0;
            PUT  DEVIANCE;
```

```
                DEVIANCE=0;
            END;
            IF  P < =0  THEN  P=.001;
            IF  P > =1  THEN  P=.999;
            LOGLIKE=  LOGLIKE  +  &RESPONSE*LOG(P)+
                      (&NUMBER-&RESPONSE)*LOG(1-P);
            IF
            &RESPONSE  NE  0  THEN  DO;
            TERM2=TERM2+&RESPONSE*LOG(&RESPONSE/&NUMBER)+
                      (&NUMBER-&RESPONSE)*LOG(1-&RESPONSE/&NUMBER);
            END;
            DEVIANCE=2*TERM2-2*LOGLIKE;
        END;
            IF  P < =0  THEN  P=.001;
            IF  P > =1  THEN  P=.999;
        MODEL  &RESPONSE=&NUMBER*P;
        V=(&NUMBER*P*(1-P));
        _WEIGHT_=(1/V)*&WEIGHT;
        _LOSS_=(-&RESPONSE*LOG(P)-(&NUMBER-&RESPONSE)*
                 LOG(1-P))/_WEIGHT_;
        PHI=&NUMBER * P *(1-P);
        DER.INTERCPT=PHI;
        %DO I=1 %TO &N;
            DER.&&VAR&I=PHI*_OLD&I;
        %END;
        OUTPUT OUT=NLINBOI P=PR PARMS=BO B1 B2 B3 SSE=X2;
                ID  _WEIGHT_ LOGLIKE TERM2 DEVIANCE;
        PROC PRINT;
    %MEND ;

DATA WEX;
    KEEP S2 R2 SR R M W;
    INPUT SEED ROOT R M Q;
    S2=0;  R2=0;  SR=0;
    IF  SEED=2  THEN  DO;  S2=1;  END;
    IF  ROOT=2  THEN  DO;  R2=1;  END;
    IF  SEED=2  AND  ROOT=2  THEN  DO;  SR=1;  END;
    W=1;
CARDS;
1 1 10 39 29
1 1 23 62 39
1 1 23 81 58
1 1 26 51 25
1 1 17 39 22
1 2  5  6  1
1 2 53 74 21
1 2 55 72 17
1 2 32 51 19
1 2 46 79 33
1 2 10 13  3
2 1  8 16  8
2 1 10 30 20
2 1  8 28 20
2 1 23 45 22
2 1  0  4  4
```

```
2 2  3 12  9
2 2 22 41 19
2 2 15 30 15
2 2 32 51 19
2 2  3  7  4
;
PROC PRINT;

        %BINOMIAL (DATA=WEX, RESPONSE=R, NUMBER=M, WEIGHT=W,
           VARS=%STR(S2 R2 SR ));

     PROC IML;
        RESET          FW=6;
        USE NLINBOI;
        READ ALL INTO XALL;
        DR=XALL(|,1|);
        N=NROW(DR);
        JV=J(N,1,1);
        DM=XALL(|,2|);
        DS2=XALL(|,3|);
        DR2=XALL(|,4|);
        DSR=XALL(|,5|);
        W=XALL(|,6|);
        LOGL=XALL(|N,8|);
        DE=XALL(|N,10|);
        FV=XALL(|,11|);
        LP=LOG((FV/DM)/(JV-(FV/DM)));
        BOS=XALL(|N,12|);
        B1S=XALL(|N,13|);
        B2S=XALL(|N,14|);
        B3S=XALL(|N,15|);
        PI=DR/DM;
        PI=PI+((PI=0)*.001);
        PI=PI-((PI=1)*.001);
        FN=(DR#LOG(PI))+((DM-DR)#LOG(JV-PI));
        F= 2*FN(|+,|);
        PRINT BOS B1S B2S B3S;
        PRINT LOGL DE F ;
        K=5;
        NK=N*K;
        OP=J(NK,1,1);
        F=OP#F;
        DE=OP#DE;
        BOS=OP#BOS;
        B1S=OP#B1S;
        B2S=OP#B2S;
        B3S=OP#B3S;
        J1=J(N,1,1);
        J2=J(N,1,2);
        J3=J(N,1,3);
        J4=J(N,1,4);
        J5=J(N,1,5);
        J=J1//J2//J3//J4//J5;
        I=DO(1,N,1);
        B={1,1,1,1,1};
```

```
        I=B@I';
*    DZ={1.732,0,-1.732};
*    DP={.1667,.6667,.1667};
     DZ={2.856966,1.35562,0,-1.35562,-2.856966};
     DP={0.011257,0.222076,0.5333,0.222076,0.011257};
     S2=B@DS2;
     R2=B@DR2;
     SR=B@DSR;
     R=B@DR;
     M=B@DM;
     FV=B@FV;
     JV=B@JV;
     Z=DZ@J1;
     P=DP@J1;
     TLP=B@LP;
     II=1;
     LP=TLP+Z#II;
     PH=JV/(JV+EXP(-LP));
     PH=PH+((PH=0)*.001);
     PH=PH-((PH=1)*.001);
     RH=PH#M;
     W=LOG(P)+((R#LOG(PH))+((M-R)#LOG(JV-PH)));
     W=EXP(W);
     PRINT R  M  FV  PI  Z  LP  PH  RH  W;
     HW=SHAPE(W,K);
     HW=HW(|+,|);
     H=B@HW';
     W=W/H;
     PRINT W;
     WDAT=R||M||S2||R2||SR||Z||W||P||F||DE||BOS||B1S||B2S||B3S;
     CREATE  BOIOUT  FROM  WDAT;
     APPEND  FROM  WDAT;
     QUIT;

DATA  BOIOUT;
  SET  BOIOUT;
CALL  SYMPUT('BOSM',COL11);
CALL  SYMPUT('B1SM',COL12);
CALL  SYMPUT('B2SM',COL13);
CALL  SYMPUT('B3SM',COL14);

%LET  B4SM=0;



DATA  BOIOUT;
  SET  BOIOUT;
%PUT  BOS  &BOSM;
%PUT  B1S  &B1SM;
%PUT  B2S  &B2SM;
%PUT  B3S  &B3SM;
%PUT  B4S  &B4SM;
DROP  COL11  COL12  COL13  COL14;
```

```
%MACRO BINOMIA2(DATA=_LAST_,RESPONSE=,NUMBER=,WEIGHT=,VARS=);

/*- - - - - - - -—- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -*/
/* VARIABLE    FUNCTION                                        */
/* - - - - - - - -   - - - - - - - -                          */
/*                                                            */
/* DATA        INPUT DATA SET                                 */
/* RESPONSE    VARIABLE CONTAINING THE NUMBER OF RESPONDANTS */
/* NUMBER      VARIABLE CONTAINING THE NUMBER IN GROUP        */
/* VARS        LIST OF INDEPENDENT VARIABLES                  */
/* P           RESPONSE PROB AS FUNCTION OF Z (=XB)           */
/* PHI         DERIVATIVE OF P AS A FUNCTION OF Z AND/OR P    */
/*                                                            */
/*- - - - - - - -—- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -*/


        %LET N=0;          /* SPLIT OUT INDIVIDUAL NAMES */
        %LET OLD=;
        %DO %WHILE(%SCAN(&VARS,&N+1)^=);
            %LET N=%EVAL(&N+1);
            %LET VAR&N=%SCAN(&VARS,&N);
            %LET OLD=&OLD _OLD&N;
        %END;
        /* DO MLE WITH NONLINEAR LEAST SQUARES */

   *PROC PRINTTO UNIT=20 NEW;

   PROC NLIN NOHALVE SIGSQ=1 DATA=&DATA(RENAME=(
       %DO I=1 %TO &N;
           &&VAR&I=_OLD&I
       %END; )) OUTEST=BOIOUT;
       RETAIN LOGLIKE   TERM2 DEVIANCE 0;
       /* START INITIAL VALUES AT ZERO */
       PARMS
           INTERCPT=0
           %DO I=1 %TO &N; &&VAR&I=0 %END; ;
       /* COMPUTE INNER PRODUCT */
       Z=INTERCPT %DO I=1 %TO &N; + &&VAR&I*_OLD&I %END; ;
       /* MODEL RESPONSE PROBABILITY */
       P=1/(1+EXP(-Z));

            IF _ITER_=0 THEN IF _N_=1 THEN DO;
                INTERCPT=&BOSM;
                %DO I=1 %TO &N;
                    &&VAR&I=&&B&I.SM;
                %END;
                END;
        IF _MODEL_=1 THEN DO;
            IF _OBS_=1 THEN DO;
                PUT LOGLIKE=;
                LOGLIKE=0;
                TERM2=0;
                PUT DEVIANCE;
                DEVIANCE=0;
            END;
```

```
        IF  P < =0  THEN  P=.001;
        IF  P > =1  THEN  P=.999;
        LOGLIKE=  LOGLIKE  +  &RESPONSE*LOG(P)+
                (&NUMBER-&RESPONSE)*LOG(1-P);
        IF
        &RESPONSE  NE  0  THEN  DO;
        TERM2=TERM2+&RESPONSE*LOG(&RESPONSE/&NUMBER)+
                (&NUMBER-&RESPONSE)*LOG(1-&RESPONSE/&NUMBER);
        END;
        DEVIANCE=2*TERM2-2*LOGLIKE;
    END;
        IF  P < =0  THEN  P=.001;
        IF  P > =1  THEN  P=.999;
        MODEL  &RESPONSE=&NUMBER*P;
        V=(&NUMBER*P*(1-P));
        _WEIGHT_=(1/V)*&WEIGHT;
*   _LOSS_=(-&RESPONSE*LOG(P)-(&NUMBER-&RESPONSE)*
            LOG(1-P))/_WEIGHT_;
        PHI=&NUMBER  *  P  *(1-P);
        DER.INTERCPT=PHI;
        %DO  I=1  %TO  &N;
            DER.&&VAR&I=PHI*_OLD&I;
        %END;
        OUTPUT  OUT=NLINBOI  P=PR  PARMS=B0  B1  B2  B3  B4;
        *        ID   _WEIGHT_  LOGLIKE  TERM2  DEVIANCE;
        PROC  PRINT;
%MEND  ;

    %MACRO  DOIT(DSN);
        %DO  R  =1  %TO  10;
            %PUT  ITERATION  &R;

    %BINOMIA2  (DATA=&DSN,  RESPONSE=COL1,  NUMBER=COL2,  WEIGHT=COL7,
        VARS=%STR(COL3  COL4  COL5  COL6));

PROC  IML;
    RESET           FW=6;
    C=.01;
    N=21;
    K=5;
    NK=N*K;
    USE  NLINBOI;
    READ  ALL  INTO  XALL;
    R=XALL(|,1|);
    M=XALL(|,2|);
    S2=XALL(|,3|);
    R2=XALL(|,4|);
    SR=XALL(|,5|);
    Z=XALL(|,6|);
    W=XALL(|,7|);
    P=XALL(|,8|);
    F=XALL(|1,9|);
    DE=XALL(|1,10|);
    RH=XALL(|,11|);
    BOS=XALL(|1,12|);
```

```
        B1S=XALL(|1,13|);
        B2S=XALL(|1,14|);
        B3S=XALL(|1,15|);
        B4S=XALL(|1,16|);
        OP=J(NK,1,1);
        LP=LOG((RH/M)/(OP-(RH/M)));
        W=LOG(P)+((R#LOG((RH/M)))+((M-R)#LOG(OP-(RH/M))));
        W=EXP(W);
        PRINT W;
        HW=SHAPE(W,K);
        PRINT HW;
        HW=HW(|+,|);
        LHW=LOG(HW);
        PRINT HW LHW;
        CHW=LHW(|,+|);
        E=-2*CHW;
        D=E+F(|1,|);
        CONV=INT(ABS(D-DE)*100);
        PRINT BOS B1S B2S B3S B4S;
        PRINT CHW DE E F D CONV;
        D=OP#D;
        BOS=OP#BOS;
        B1S=OP#B1S;
        B2S=OP#B2S;
        B3S=OP#B3S;
        B4S=OP#B4S;
        CONV=OP#CONV;
        F=OP#F;
        B={1,1,1,1,1};
        H=B@HW';
        W=W/H;
        W=W#(W>.0000001);
WDAT=R||M||S2||R2||SR||Z||W||P||F||D||CONV||
        BOS||B1S||B2S||B3S||B4S;
        CREATE &DSN FROM WDAT;
        APPEND FROM WDAT;
        QUIT;

DATA &DSN;SET &DSN;
CALL SYMPUT('CONVM',COL11);
CALL SYMPUT('BOSM',COL12);
CALL SYMPUT('B1SM',COL13);
CALL SYMPUT('B2SM',COL14);
CALL SYMPUT('B3SM',COL15);
CALL SYMPUT('B4SM',COL16);

DATA &DSN;SET &DSN;
%PUT CONV &CONVM;
%PUT BOS &BOSM;
%PUT B1S &B1SM;
%PUT B2S &B2SM;
%PUT B3S &B3SM;
%PUT B4S &B4SM;
DROP COL11 COL12 COL13 COL14 COL15 COL16;
%IF &CONVM=0 %THEN %LET R=10;
```

```
    %PUT R  &R;

        %END;
%MEND  DOIT;

    %DOIT(BOIOUT)

//
```

APPENDIX D

PROGRAM LISTING FOR TYPE II MODELS (2)

```
//U12558AA   JOB  (12558,AND-RO-NIKO),'BIN35L3',CLASS=4,TIME=(10,50),
//    MSGCLASS=X,NOTIFY=*,MSGLEVEL=(1,1)
/*PASSWORD  ?
/*ROUTE  PRINT  LOCAL
/*JOBPARM  ROOM=V,FORMS=9001
//      EXEC  SAS,REGION=8500K,OPTIONS='MACRO'
//IN  DD  DSN=U10063B.BSIM.SAS.DATA,DISP=OLD
//CT  DD  DSN=U12558A.BIN1.CNTL(CTREE3L),DISP=OLD
//SYSIN  DD  *

   OPTIONS  DQUOTE  NOMPRINT  NOSYMBOLGEN  NOMACROGEN;


%MACRO  BINOMIAL(DATA=,RESPONSE=,NUMBER=,WEIGHT=,VARS=);


      %LET  N=0;          /*  SPLIT  OUT  INDIVIDUAL  NAMES  */
      %LET  OLD=;
      %DO  %WHILE(%SCAN(&VARS,&N+1)^=);
          %LET  N=%EVAL(&N+1);
          %LET  VAR&N=%SCAN(&VARS,&N);
          %LET  OLD=&OLD  _OLD&N;
      %END;
      /*  DO  MLE  WITH  NONLINEAR  LEAST  SQUARES  */


   *PROC  PRINTTO  UNIT=20  NEW;


   PROC  NLIN  NOHALVE  SIGSQ=1  DATA=&DATA(RENAME=(
      %DO  I=1  %TO  &N;
          &&VAR&I=_OLD&I
      %END;  ))  OUTEST=BOIOUT;
      RETAIN  LOGLIKE   TERM2  DEVIANCE  0;
      /*  START  INITIAL  VALUES  AT  ZERO  */
      PARMS
          INTERCPT=0
          %DO  I=1  %TO  &N;  &&VAR&I=0  %END;  ;
      /*  COMPUTE  INNER  PRODUCT  */
      Z=INTERCPT  %DO  I=1  %TO  &N;  +  &&VAR&I*_OLD&I  %END;  ;
      /*  MODEL  RESPONSE  PROBABILITY  */
      P=1/(1+EXP(-Z));

      IF  _MODEL_=1  THEN  DO;
          IF  _OBS_=1  THEN  DO;
              PUT  LOGLIKE=;
              LOGLIKE=0;
              TERM2=0;
              PUT  DEVIANCE;
              DEVIANCE=0;
          END;
          IF  P<=0  THEN  P=.000001;
          IF  P>=1  THEN  P=.999999;
          LOGLIKE=  LOGLIKE  +  &RESPONSE*LOG(P)+
                    (&NUMBER-&RESPONSE)*LOG(1-P);
          IF  0.5<&RESPONSE  AND  &RESPONSE<&NUMBER-0.5  THEN  DO;
          TERM2=TERM2+&RESPONSE*LOG(&RESPONSE/&NUMBER)+
                    (&NUMBER-&RESPONSE)*LOG(1-&RESPONSE/&NUMBER);
          END;
```

```
                  DEVIANCE=2*TERM2-2*LOGLIKE;
              END;
                  IF  P<=0  THEN  P=.000001;
                  IF  P>=1  THEN  P=.999999;
              MODEL  &RESPONSE=&NUMBER*P;
              V=(&NUMBER*P*(1-P));
              _WEIGHT_=(1/V)*&WEIGHT;
    *     _LOSS_=(-&RESPONSE*LOG(P)-(&NUMBER-&RESPONSE)*
                     LOG(1-P))/_WEIGHT_;
              PHI=&NUMBER  *  P  *(1-P);
              DER.INTERCPT=PHI;
              %DO  I=1  %TO  &N;
                  DER.&&VAR&I=PHI*_OLD&I;
              %END;
              OUTPUT  OUT=NLINBOI  P=PR  PARMS=BO  B1  B2  SSE=X2;
                  ID  _WEIGHT_  LOGLIKE  TERM2  DEVIANCE;

              PROC  PRINT  DATA=NLINBOI(OBS=15);

          %MEND  ;


DATA L3;INFILE  CT;
    INPUT  LOC  STATE  STAND  FAM  R  M  ;
    KEEP  LOC  FAM  R  M  T2  T3  W;
    T2=0;  T3=0;
    IF  LOC=2  THEN  DO;  T2=1;  END;
    IF  LOC=3  THEN  DO;  T3=1;  END;
    W=1;
PROC  PRINT;


%BINOMIAL  (DATA=L3,  RESPONSE=R,  NUMBER=M,  WEIGHT=W,
           VARS=%STR(T2  T3));


DATA LOCEX;MERGE  L3  NLINBOI;BY  LOC  FAM;
    DROP  LOC  FAM;
    I=LOC;J=FAM;
    DO  K=1  TO  3;
        DO  L=1  TO  5;
    IF  K=1  AND  L=1  THEN  DO;ZK=-1.732;ZL=-2.857;PK=.1667;PL=.01126;END;
    IF  K=1  AND  L=2  THEN  DO;ZK=-1.732;ZL=-1.356;PK=.1667;PL=.22208;END;
    IF  K=1  AND  L=3  THEN  DO;ZK=-1.732;ZL= 0      ;PK=.1667;PL=.53330;END;
    IF  K=1  AND  L=4  THEN  DO;ZK=-1.732;ZL= 1.356;PK=.1667;PL=.22208;END;
    IF  K=1  AND  L=5  THEN  DO;ZK=-1.732;ZL= 2.857;PK=.1667;PL=.01126;END;
    IF  K=2  AND  L=1  THEN  DO;ZK= 0      ;ZL=-2.857;PK=.6667;PL=.01126;END;
    IF  K=2  AND  L=2  THEN  DO;ZK= 0      ;ZL=-1.356;PK=.6667;PL=.22208;END;
    IF  K=2  AND  L=3  THEN  DO;ZK= 0      ;ZL= 0      ;PK=.6667;PL=.53330;END;
    IF  K=2  AND  L=4  THEN  DO;ZK= 0      ;ZL= 1.356;PK=.6667;PL=.22208;END;
    IF  K=2  AND  L=5  THEN  DO;ZK= 0      ;ZL= 2.857;PK=.6667;PL=.01126;END;
    IF  K=3  AND  L=1  THEN  DO;ZK=-1.732;ZL=-2.857;PK=.1667;PL=.01126;END;
    IF  K=3  AND  L=2  THEN  DO;ZK=-1.732;ZL=-1.356;PK=.1667;PL=.22208;END;
    IF  K=3  AND  L=3  THEN  DO;ZK=-1.732;ZL= 0      ;PK=.1667;PL=.53330;END;
    IF  K=3  AND  L=4  THEN  DO;ZK=-1.732;ZL= 1.356;PK=.1667;PL=.22208;END;
    IF  K=3  AND  L=5  THEN  DO;ZK=-1.732;ZL= 2.857;PK=.1667;PL=.01126;END;
        OUTPUT;
    END;
```

```
    END;

PROC SORT DATA=LOCEX;BY  K  I  L  J;
PROC PRINT DATA=LOCEX(OBS=40);

        PROC IML;
            RESET            FW=10;
            USE LOCEX;
            READ ALL INTO XALL;
            R=XALL(|,1|);
            M=XALL(|,2|);
            T2=XALL(|,3|);
            T3=XALL(|,4|);
            IJKL=NROW(R);
            N=IJKL;
            LOGL=XALL(|N,9|);
            F=XALL(|N,10|);
            DE=XALL(|N,11|);
            FV=XALL(|,12|);
            BOS=XALL(|N,13|);
            B1S=XALL(|N,14|);
            B2S=XALL(|N,15|);
            ZK=XALL(|,21|);
            ZL=XALL(|,22|);
            PK=XALL(|,23|);
            PL=XALL(|,24|);
        PRINT BOS B1S B2S;
        PRINT LOGL DE F ;

            FREE XALL;
            SHOW ALL;
            K=3;
            L=5;
            BK=J(K,1,1);
            BL=J(L,1,1);
            B=J(K*L,1,1);
            I=3;J=39;N=I*J;
            NK=N*K;  NL=N*L;  KL=K*L;
            NKL=N*K*L;
            OP=J(NKL,1,1);

            LP=LOG((FV/M)/(OP-(FV/M)));
            LP=LP+ZL+ZK;
            PH=OP/(OP+EXP(-LP));
            FREE LP;

    LBIJKL=LOG(PK)+((R#LOG((PH)))+((M-R)#LOG(OP-(PH))));
    BIJKL=EXP(LBIJKL);
    BIJL1=SHAPE(BIJKL,K);
    BIJL2=BIJL1(|+,|);
    BIJL=BK@BIJL2';
    WP1=BIJKL/BIJL;

            SHOW ALL;
```

```
FREE LBIJKL BIJKL BIJL1 BIJL;
BJL1=SHAPE(BIJL2,I);
PBJL1=BJL1(|#,|);
PBJL=J(K*I,1,1)@PBJL1';
BJL=PBJL#PL;
PBJL2=BJL(|1:J*L|);

FREE BJL1 PBJL1 PBJL BIJL2;
BJ1=SHAPE(PBJL2,L);
BJ3=BJ1(|+,|);
BJ4=LOG(BJ3);

FREE BJ1 PBJL2;
LL=BJ4(|,+|);
BJ=J(K*I*L,1,1)@BJ3';
WP2=BJL/BJ;

FREE BJ BJ3 BJ4 BJL;
W=WP1#WP2;
FREE WP1 WP2;
    WDAT=              R||M||T2||T3||ZK||ZL||
            W||PK||PL||(OP#F)||(OP#LL)||(OP#B0S)||(OP#B1S)||(OP#B2S);
    CREATE BOIOUT FROM WDAT;
    APPEND FROM WDAT;

  SHOW ALL;
    QUIT;

%MACRO BINOMIA2(DATA=_LAST_,RESPONSE=,NUMBER=,WEIGHT=,VARS=);

    %LET N=0;          /* SPLIT OUT INDIVIDUAL NAMES */
    %LET OLD=;
    %DO %WHILE(%SCAN(&VARS,&N+1)^=);
        %LET N=%EVAL(&N+1);
        %LET VAR&N=%SCAN(&VARS,&N);
        %LET OLD=&OLD _OLD&N;
    %END;
    /* DO MLE WITH NONLINEAR LEAST SQUARES */

  *PROC PRINTTO UNIT=20 NEW;

  PROC NLIN NOHALVE SIGSQ=1 DATA=&DATA(RENAME=(
    %DO I=1 %TO &N;
        &&VAR&I=_OLD&I
    %END; )) OUTEST=BOIOUT;
    RETAIN LOGLIKE  TERM2 DEVIANCE 0;
    /* START INITIAL VALUES AT ZERO */
    PARMS
        INTERCPT=0
        %DO I=1 %TO &N; &&VAR&I=0 %END; ;
    /* COMPUTE INNER PRODUCT */
    Z=INTERCPT %DO I=1 %TO &N; + &&VAR&I*_OLD&I %END; ;
    /* MODEL RESPONSE PROBABILITY */
    P=1/(1+EXP(-Z));
```

```
        IF _ITER_=0 THEN IF _N_=1 THEN DO;
            INTERCPT=&BOSM;
            %DO I=1 %TO &N;
                &&VAR&I=&&B&I.SM;
            %END;
            END;

    IF _MODEL_=1 THEN DO;
        IF _OBS_=1 THEN DO;
            PUT LOGLIKE=;
            LOGLIKE=0;
            TERM2=0;
            PUT DEVIANCE;
            DEVIANCE=0;
        END;
        IF P<=0 THEN P=.000001;
        IF P>=1 THEN P=.999999;
        LOGLIKE= LOGLIKE + &RESPONSE*LOG(P)+
                (&NUMBER-&RESPONSE)*LOG(1-P);
        IF 0.5<&RESPONSE AND &RESPONSE<&NUMBER-0.5 THEN DO;
        TERM2=TERM2+&RESPONSE*LOG(&RESPONSE/&NUMBER)+
                (&NUMBER-&RESPONSE)*LOG(1-&RESPONSE/&NUMBER);
        END;
        DEVIANCE=2*TERM2-2*LOGLIKE;
    END;
        IF P<=0 THEN P=.000001;
        IF P>=1 THEN P=.999999;
    MODEL &RESPONSE=&NUMBER*P;
    V=(&NUMBER*P*(1-P));
    _WEIGHT_=(1/V)*&WEIGHT;
*  _LOSS_=(-&RESPONSE*LOG(P)-(&NUMBER-&RESPONSE)*
            LOG(1-P))/_WEIGHT_;
    PHI=&NUMBER * P *(1-P);
    DER.INTERCPT=PHI;
    %DO I=1 %TO &N;
        DER.&&VAR&I=PHI*_OLD&I;
    %END;
    OUTPUT OUT=NLINBOI P=PR PARMS=BO B1 B2 B3 B4;
    *        ID  _WEIGHT_ LOGLIKE TERM2 DEVIANCE;
    PROC PRINT DATA=NLINBOI(OBS=20);
%MEND ;

    %MACRO DOIT(DSN);
        %DO R =1 %TO 20 ;
            %PUT ITERATION &R;

    %BINOMIA2 (DATA=&DSN, RESPONSE=COL1, NUMBER=COL2, WEIGHT=COL7,
        VARS=%STR(COL3 COL4 COL5 COL6));

PROC IML;
    RESET           FW=5 ;
    C=.01;
    I=3;J=39;N=I*J;
    K=3; L=5;
    IKL=I*K*L;IJKL=I*J*K*L;
```

```
    OP=J(IJKL,1,1);
    NTR=3;NBL=39;
    NK=N*K;
    NL=N*L;
    KL=K*L;
    NKL=N*K*L;
    BK=J(K,1,1);
    BL=J(L,1,1);
    B=J(K*L,1,1);
      USE NLINBOI;
      READ ALL INTO XALL;
      R=XALL(|,1|);
      M=XALL(|,2|);
      T2=XALL(|,3|);
      T3=XALL(|,4|);
      ZK=XALL(|,5|);
      ZL=XALL(|,6|);
      PK=XALL(|,8|);
      PL=XALL(|,9|);
      F=XALL(|1,10|);
      LL=XALL(|1,11|);
      RH=XALL(|,12|);
      BOS=XALL(|1,13|);
      B1S=XALL(|1,14|);
      B2S=XALL(|1,15|);
      B3S=XALL(|1,16|);
      B4S=XALL(|1,17|);
  FREE XALL;
  SHOW ALL;
  LBIJKL=LOG(PK)+((R#LOG((RH/M)))+((M-R)#LOG(OP-(RH/M))));
BIJKL=EXP(LBIJKL);
BIJL1=SHAPE(BIJKL,K);
BIJL2=BIJL1(|+,|);
BIJL=BK@BIJL2';
WP1=BIJKL/BIJL;

FREE LBIJKL BIJKL BIJL1 BIJL;
BJL1=SHAPE(BIJL2,I);
PBJL1=BJL1(|#,|);
PBJL=J(K*I,1,1)@PBJL1';
BJL=PBJL#PL;
PBJL2=BJL(|1:J*L|);

FREE BJL1 PBJL1 PBJL BIJL2;
BJ1=SHAPE(PBJL2,L);
BJ3=BJ1(|+,|);
BJ4=LOG(BJ3);

FREE BJ1 PBJL2;
L=BJ4(|,+|);
D=2*(L-F);
BJ=J(IKL,1,1)@BJ3';
WP2=BJL/BJ;

FREE BJ BJ3 BJ4 BJL;
```

```
W=WP1#WP2;
FREE WP1 WP2;
    CONV=INT(ABS(L-LL)*100);
    PRINT BOS B1S B2S B3S B4S;
    PRINT   LL (|FORMAT=12.4|) L (|FORMAT=12.4|)
              F (|FORMAT=12.4|)
              D (|FORMAT=12.4|)   CONV;
WDAT=              R||M||T2||T3||ZK||ZL||W||PK||PL||
     (OP#F)||(OP#L)||(OP#CONV)||
     (OP#BOS)||(OP#B1S)||(OP#B2S)||(OP#B3S)||(OP#B4S);
     CREATE &DSN FROM WDAT;
     APPEND FROM WDAT;
     QUIT;

     PROC PRINT DATA=BOIOUT(OBS=50);

DATA &DSN;
SET &DSN;
CALL SYMPUT('SMM',COL10);
CALL SYMPUT('LLM',COL11);
CALL SYMPUT('CONVM',COL12);
CALL SYMPUT('BOSM',COL13);
CALL SYMPUT('B1SM',COL14);
CALL SYMPUT('B2SM',COL15);
CALL SYMPUT('B3SM',COL16);
CALL SYMPUT('B4SM',COL17);

DATA &DSN;
   SET &DSN;
DROP   COL12 COL13 COL14 COL15 COL16 COL17;
  %PUT CONV &CONVM;
  %PUT SM &SMM;
  %PUT LL &LLM;
  %PUT BOS &BOSM;
  %PUT B1S &B1SM;
  %PUT B2S &B2SM;
  %PUT B3S &B3SM;
  %PUT B4S &B4SM;
  %IF &R=1 %THEN %LET CONVM=1;
  %IF &CONVM=0 %THEN %LET R=20;
 %PUT R &R;

PROC PRINT DATA=BOIOUT(OBS=5 );

    %END;
%MEND DOIT;

   %DOIT(BOIOUT)

 DATA IN.II3L35BI;SET NLINBOI;

%MACRO GENV;
%DO N=1 %TO &NP;
V&N.IJL1=SHAPE(V&N.IJKL,K);
V&N.IJL=V&N.IJL1(|+,|);
```

```
V&N.JL1=SHAPE(V&N.IJL',I);
V&N.JL=V&N.JL1(|+,|);
V&N.J1=SHAPE(V&N.JL',L);
V&N.J=V&N.J1(|+,|);
    SHOW ALL;
%END;
%MEND GENV;


%MACRO GENCH;
%DO N=1 %TO &NC;
  CH&N=V(|,&N|)*(V(|,&N|))';
%END;
    SHOW ALL;


%MEND GENCH;

PROC PRINT DATA=IN.II3LBINL(OBS=10);
      PROC IML;
          RESET           FW=10;
          USE IN.II3LBINL;
          READ ALL INTO XALL;
          I=3;J=39;
          K=3;L=3;
          R=XALL(|,1|);
          M=XALL(|,2|);
          L2=XALL(|,3|);
          L3=XALL(|,4|);
          ZK=XALL(|,5|);
          ZL=XALL(|,6|);
          W=XALL(|,7|);
          PK=XALL(|,8|);
          PL=XALL(|,9|);
          IJKL=NROW(R);
          N=IJKL;
          F=XALL(|N,10|);
          LOGL=XALL(|N,11|);
          RH=XALL(|,12|);
          BO=XALL(|N,13|);
          B1=XALL(|N,14|);
          B2=XALL(|N,15|);
          S1=XALL(|N,16|);
          S2=XALL(|N,17|);
        PRINT BO B1 B2 S1 S2;
        PRINT LOGL F ;
          OP=J(N,1,1);
V1IJKL=W#(R-RH);
V2IJKL=W#L2#(R-RH);
V3IJKL=W#L3#(R-RH);
V4IJKL=W#ZK#(R-RH);
V5IJKL=W#ZL#(R-RH);
          SHOW ALL;

%LET NV=V;
%LET NR=3;
%LET NC=39;
```

```
%LET  NP=5;

%GENV

RESET  PRINT;
V=V1J//V2J//V3J//V4J//V5J;

%GENCH
IM=CH1  +  CH2   +  CH3   +  CH4   +  CH5   +  CH6   +  CH7   +  CH8   +  CH9  +   CH10
   +CH11+  CH12  +  CH13  +  CH14  +  CH15  +  CH16  +  CH17  +  CH18  +  CH19  +  CH20
   +CH21+  CH22  +  CH23  +  CH24  +  CH25  +  CH26  +  CH27  +  CH28  +  CH29  +  CH30
   +CH31+  CH32  +  CH33  +  CH34  +  CH35  +  CH36  +  CH37  +  CH38  +  CH39  ;
      SHOW  ALL;
  IIM=INV(IM);
  AIIM=ABS(IIM);
  SIIM=SQRT(IIM);
  CHECK=IM*IIM;
      SHOW  ALL;
  QUIT;
//
```

VITA

Andronikos Mauromoustakos

Candidate fot the Degree of

Doctor of Philosophy

Thesis: MODELS FOR PROPORTIONS IN A TWO-WAY CROSS CLASSIFICATION WITHOUT INTERACTION

Major Field: Statistics

Biographical:

> Personal Data: Born in Thessaloniki, Greece, March 17, 1959, the son of Antonios and Eleni Mauromoustakos.
>
> Education: Graduated from the 5th Male High School, Thessaloniki, Greece in 1976; received Bachelor of Science from Oral Roberts University, Tulsa, Oklahoma in May 1982; received a Masters July 1984; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in May of 1989.
>
> Professional Experience: Worked as a tennis teaching professional in Greece from 1976 to 1979. Served as a teaching and or as a research assistant fot the Department of Statistics at Oklahoma State University, during 1982-1987 . Worked as a research assistant for the Agricultural Experiment Station of Oklahoma State University from 1983 to 1987. Worked as systems analyst for the University Computer Center at Oklahoma State University from 1987 to 1988.