This research was supported in part by National Institutes of Health, Institute of General Medical Sciences through Grant GM-70004-05 and by the U.S. Bureau of the Census through Joint Statistical Agreement JSA 75-2.

THE ASYMPTOTIC COVARIANCE STRUCTURE OF ESTIMATED PARAMETERS FROM CONTINGENCY TABLE LOG-LINEAR MODELS

by

Gary G. Koch, Daniel H. Freeman, Jr., & H. Dennis Tolley

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Memio Series No. 1046

December 1975

THE ASYMPTOTIC COVARIANCE STRUCTURE OF ESTIMATED PARAMETERS FROM CONTINGENCY TABLE LOG-LINEAR MODELS

by

Gary G. Koch, Daniel H. Freeman, Jr.,* and H. Dennis Tolley†
Department of Biostatistics
University of North Carolina at Chapel Hill

ABSTRACT

This paper presents a matrix formulation for log-linear model analysis of contingency tables. Both weighted least squares estimators and maximum likelihood estimators are considered in this framework with results being given for their corresponding covariance matrix structures. Moreover, a general analytical strategy is presented for the simultaneous use of these two estimation procedures in a manner which emphasizes their respective strengths. Four examples illustrating the application of this methodology are then provided.

* now at Department of Epidemiology and Public Health, Division of Biometry, Yale University School of Medicine, New Haven, Connecticut

† now at Department of Community Health Science, Duke University, Durham, North Carolina

THE ASYMPTOTIC COVARIANCE STRUCTURE OF ESTIMATED PARAMETERS FROM CONTINGENCY TABLE LOG-LINEAR MODELS

Gary G. Koch, Daniel H. Freeman, Jr., and H. Dennis Tolley
Department of Biostatistics
University of North Carolina at Chapel Hill
Chapel Hill, N. C., U. S. A.

1. INTRODUCTION

Log-linear models provide a useful framework for analyzing many types of multi-dimensional contingency tables. As such, they have received considerable attention in the recent statistical literature. In this regard, two general methodological strategies exist for estimating the parameters associated with such models. These are

I. Iterative Proportional Fitting (IPF) of marginal configurations which are statistically sufficient for the log-linear model

II. Weighted Least Squares (WLS) fitting of asymptotic regression models to log-linear functions of cell proportions.

Method I is discussed extensively in Bishop, Fienberg, and Holland [1975] (Hereafter, abbreviated BFH) as well as many other references including Goodman [1970] and Ku and Kullback [1974]; various aspects of Method II are outlined in Grizzle and Williams [1972] as an application of the general approach of Grizzle, Starmer, and Koch [1969] (Hereafter, abbreviated GSK).

For large samples, these two methods of analysis are asymptotically equivalent in the sense that their respective vectors of estimated parameters are both asymptotically unbiased and have the same asymptotic covariance matrix for models which do indeed adequately characterize the data. Thus, as indicated in Appendix 3 of Koch and Tolley [1975], the choice between them in such cases is largely a matter of personal

tastes and computational convenience. However, log-linear models are often of interest for situations involving moderate size samples in which many of the individual cell frequencies are very small (i.e., less than 5). In these cases, the GSK approach is not strictly applicable because the vector of observed cell proportions cannot be presumed as approximately having a multivariate normal distribution. On the other hand, if the log-linear model to be fitted is known a priori to be valid, then the IPF approach is appropriate within the context of the less restrictive assumptions that the vector of estimated parameters can be regarded as approximately having a multivariate normal distribution. However, these estimated parameters may be represented as certain functions of the vector of observed cell proportions and thus can be further analyzed by the GSK approach as extended in Tolley and Koch [1974] Koch and Tolley[1975], and Koch, Tolley, and Freeman [1976]. This type of parameter oriented methodology is motivated by growth curve model concepts as described in Potthoff and Roy [1964], Allen and Grizzle [1969], and Koch and Greenberg [1971]; and henceforth it will be referred to as Functional Asymptotic Regression Methodology (FARM). As such, FARM is useful for the following types of applications:

1. The characterization of the multivariate relationships among a specific set of attributes and their variation across a set of sub-populations in terms of their inherent main effects and interactions
2. The analysis of functions of log-linear model based predicted (smoothed) values for cell probabilities; e.g.,
    a. first order marginal probabilities in repeated measurement situations as described in Koch, Freeman, Freeman, and Lehnen [1974]
    b. measures of agreement based on diagonal probabilities as described in Landis and Koch [1976].

3. The analysis of data from complex sample survey designs and other situations for which the multinomial-based, maximum likelihood formulation of IPF estimators is not applicable

Although such FARM analyses may often appear difficult, they can be undertaken in a straightforward manner through the use of matrix operations similar to those described by GSK. In this regard, a fundamental component of this general methodology is the estimated covariance matrix for the estimated parameters based on IPF. Thus, the primary purpose of this paper is to present a matrix formulation for the covariance matrix of these IPF estimators and the corresponding set of cell probability predicted values. These results are obtained via the well-known δ-method as based on the first order linear Taylor series approximating counterparts to such estimators with the required matrix of first derivatives being determined by implicit function differentiation methods.

2. THEORY

In this section, attention is initially focused on the statistical properties of IPF estimators for the situation where a simple random sample has been selected from a population in which the multivariate attributes under study are a priori known to be characterized by a specific log-linear model. Such assumptions are then relaxed, and results are subsequently given under more general conditions.

2.1 Simple Random Sampling From a Single Population of Known Structure

For a particular population, suppose the multivariate relationships among a specific set of d attributes is of interest. Let $j_g=1,2,\ldots,L_g$ index the set of categories which correspond to the possible responses for the g-th attribute where g=1,2,...,d. If a simple random sample of size n is selected from this population, the resulting data can be summarized with a d-dimensional contingency table in which $n_{j_1 j_2 \cdots j_d}$ denotes the observed frequency of the multivariate response profile $(j_1,j_2,\ldots,j_d)$.

The vector $\underset{\sim}{n}$ where

$$\underset{\sim}{n}' = (n_{11\ldots1},\ldots,n_{j_1 j_2 \cdots j_d},\ldots,n_{L_1 L_2 \ldots L_d}) \qquad (2.1)$$

is assumed to follow the multinomial distribution with parameters n and $\underset{\sim}{\pi}$ where

$$\underset{\sim}{\pi}' = (\pi_{11\ldots1},\ldots,\pi_{j_1 j_2 \cdots j_d},\ldots,\pi_{L_1 L_2 \ldots L_d}) \qquad (2.2)$$

with $\pi_{j_1 j_2 \cdots j_d}$ representing the probability that a randomly selected element from the population is classified into the $(j_1,j_2,\ldots,j_d)$-th response profile. Thus, the relevant product multinomial probability model is

$$\phi = n! \prod_{j_1=1}^{L_1} \prod_{j_2=1}^{L_2} \cdots \prod_{j_d=1}^{L_d} \{\pi_{j_1 j_2 \cdots j_d}^{n_{j_1 j_2 \cdots j_d}} / n_{j_1 j_2 \cdots j_d}!\} \qquad (2.3)$$

with the constraint

$$\sum_{j_1=1}^{L_1}\sum_{j_2=1}^{L_2}\cdots\sum_{j_d=1}^{L_d}\pi_{j_1 j_2 \cdots j_d} = 1 \quad . \tag{2.4}$$

The population is also assumed to have known structure in the sense that the variation among the elements of $\underset{\sim}{\pi}$ is assumed to be characterized by the log-linear model

$$\underset{(rx1)}{\underset{\sim}{\pi}} = \underset{\sim}{\pi}(\underset{\sim}{\beta}) = \{\underset{\sim\sim\sim}{\exp}(\underset{(rxt)}{\underset{\sim}{X}}\ \underset{(txl)}{\underset{\sim}{\beta}})\}/\{\underset{(lxr)}{\underset{\sim r}{1}'}\ [\underset{\sim\sim\sim}{\exp}(\underset{(rxt)}{\underset{\sim}{X}}\ \underset{(txl)}{\underset{\sim}{\beta}})]\} \tag{2.5}$$

where $r = \prod_{g=1}^{d} L_g$ denotes the total number of possible multivariate response profiles, $\underset{\sim}{X}$ denotes an appropriate (rxt) design (or independent variable) matrix of known coefficients whose t columns are linearly independent and represent a basis for the main effects and interactions which constitute the model, $\underset{\sim}{\beta}$ denotes the corresponding (txl) vector of unknown parameters, $\underset{\sim r}{1}$ denotes an (rxl) vector of 1's, and $\underset{\sim\sim\sim}{\exp}$ transforms a vector to the corresponding vector of exponential functions. Finally, the columns of the matrix $\underset{\sim}{X}$ are assumed without loss of generality to be orthogonal to $\underset{\sim r}{1}$; i.e.,

$$\underset{\sim}{X}'\underset{\sim r}{1} = \underset{\sim t}{0} \tag{2.6}$$

where $\underset{\sim t}{0}$ is a (txl) vector of 0's.

The method of maximum likelihood will be used to form estimators $\hat{\underset{\sim}{\beta}}$ for $\underset{\sim}{\beta}$. These estimators are characterized implicitly by the following set of equations

$$\left[\frac{\underset{\sim}{d}\ \log_e \phi}{\underset{\sim\sim}{d\beta}} \;\middle|\; \hat{\underset{\sim}{\beta}}=\underset{\sim}{\beta}\right] = \underset{\sim t}{0} \tag{2.7}$$

where the likelihood ϕ is defined by (2.3). By matrix differentiation methods similar to those used by GSK and Forthofer and Koch [1973], it follows that

$$\frac{d}{d\underset{\sim}{\beta}'}[\log_e \phi] = \frac{d}{d\underset{\sim}{\beta}'}[\underset{\sim}{n}'\underset{\sim}{X}\underset{\sim}{\beta} - n\log_e \{\underset{\sim}{1}_r' [\underset{\sim}{\exp}(\underset{\sim}{X}\underset{\sim}{\beta})]\}] \qquad (2.8)$$

$$= \underset{\sim}{n}'\underset{\sim}{X} - \frac{n[\underset{\sim}{\exp}(\underset{\sim}{X}\underset{\sim}{\beta})]'\underset{\sim}{X}}{\{\underset{\sim}{1}'[\underset{\sim}{\exp}(\underset{\sim}{X}\underset{\sim}{\beta})\}}$$

$$= \underset{\sim}{n}'\underset{\sim}{X} - n[\underset{\sim}{\pi}(\underset{\sim}{\beta})]'\underset{\sim}{X} \quad .$$

Thus, the equations (2.7) may be compactly written as

$$\underset{\sim}{X}'\hat{\underset{\sim}{\pi}} = \underset{\sim}{X}'\underset{\sim}{p} \qquad (2.9)$$

where $\hat{\underset{\sim}{\pi}} = \underset{\sim}{\pi}(\hat{\underset{\sim}{\beta}})$ represents the maximum likelihood estimator for $\underset{\sim}{\pi}$ based on $\hat{\underset{\sim}{\beta}}$ via (2.5), and $\underset{\sim}{p} = (\underset{\sim}{n}/n)$ represents the observed vector of sample proportions which correspond to the unrestricted maximum likelihood estimator for $\underset{\sim}{\pi}$ for the situation where there is no assumed model like (2.5).

Although the equations (2.9) appear straightforward, their non-linear structure often does not permit explicit solution. In those cases, iterative methods like IPF or successive search algorithms are required to determine $\hat{\underset{\sim}{\beta}}$ and $\hat{\underset{\sim}{\pi}}$. If these computational problems are presumed manageable, another important consideration is the estimation of the asymptotic covariance matrix for $\hat{\underset{\sim}{\beta}}$. Since the vector $\underset{\sim}{\pi}$ has been assumed to be characterized by the log-linear model (2.5), this question can be handled here by determining the Fisher Information Matrix

$$\frac{\partial}{\partial \underset{\sim}{\beta}\partial \underset{\sim}{\beta}'} [\log \phi] = -n\{ \frac{\partial}{\partial \underset{\sim}{\beta}'} [\underset{\sim}{X}'\{\underset{\sim}{\pi}(\underset{\sim}{\beta})\}]\} \tag{2.10}$$

$$= -n\underset{\sim}{X}' [\frac{\partial}{\partial \underset{\sim}{\beta}'} \{\underset{\sim}{\pi}(\underset{\sim}{\beta})\}]$$

By matrix differentiation methods similar to those used to obtain (2.8) it follows that

$$\frac{\partial}{\partial \underset{\sim}{\beta}'} \{\underset{\sim}{\pi}(\underset{\sim}{\beta})\} = \frac{\partial}{\partial \underset{\sim}{\beta}'} \{ \frac{\exp(\underset{\sim}{X}\underset{\sim}{\beta})}{\underset{\sim}{1}'[\exp(\underset{\sim}{X}\underset{\sim}{\beta})]} \} \tag{2.11}$$

$$= \frac{\{\underset{\sim}{1}_r'[\exp(\underset{\sim}{X}\underset{\sim}{\beta})]\}D_{\exp(\underset{\sim}{X}\underset{\sim}{\beta})}\underset{\sim}{X} - [\exp(\underset{\sim}{X}\underset{\sim}{\beta})][\exp(\underset{\sim}{X}\underset{\sim}{\beta})]'\underset{\sim}{X}}{\{\underset{\sim}{1}_r'[\exp(\underset{\sim}{X}\underset{\sim}{\beta})]\}^2}$$

where $D_{\underset{\sim}{y}}$ is a diagonal matrix with the elements of the vector $\underset{\sim}{y}$ on the main diagonal. However, by using the assumed model structure (2.5), expression (2.11) may be simplified to

$$\frac{\partial}{\partial \underset{\sim}{\beta}'} [\underset{\sim}{\pi}(\underset{\sim}{\beta})] = [D_{\underset{\sim}{\pi}(\underset{\sim}{\beta})}\underset{\sim}{X}] - [\underset{\sim}{\pi}(\underset{\sim}{\beta})][\underset{\sim}{\pi}(\underset{\sim}{\beta})]'\underset{\sim}{X} \tag{2.12}$$

$$= [D_{\underset{\sim}{\pi}} - \underset{\sim}{\pi}\underset{\sim}{\pi}']\underset{\sim}{X} \quad .$$

If the result in (2.12) is substituted into (2.10), then the Fisher Information Matrix is given by

$$\frac{\partial}{\partial \underset{\sim}{\beta}\partial \underset{\sim}{\beta}'} [\log \phi] = -n\, \underset{\sim}{X}'[D_{\underset{\sim}{\pi}} - \underset{\sim}{\pi}\underset{\sim}{\pi}']\underset{\sim}{X} \quad . \tag{2.13}$$

Thus, the asymptotic covariance matrix $V_{\hat{\underset{\sim}{\beta}}}(\underset{\sim}{\pi})$ for $\hat{\underset{\sim}{\beta}}$ is obtained by forming the negative inverse of the Fisher Information Matrix as shown in (2.14).

$$V_{\hat{\underset{\sim}{\beta}}}(\underset{\sim}{\pi}) = \frac{1}{n} \{\underset{\sim}{X}'[D_{\underset{\sim}{\pi}} - \underset{\sim}{\pi}\underset{\sim}{\pi}']\underset{\sim}{X}\}^{-1} \tag{2.14}$$

Since $\hat{\underset{\sim}{\pi}} = \underset{\sim}{\pi}(\hat{\underset{\sim}{\beta}})$ is a consistent estimator for $\hat{\underset{\sim}{\beta}}$, a consistent estimator for the covariance matrix $\underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}}(\underset{\sim}{\pi})$ in (2.14) is

$$\underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}} = \underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}}(\hat{\underset{\sim}{\pi}}) = \frac{1}{n}\{\underset{\sim}{X}'[\underset{\sim}{D}_{\hat{\underset{\sim}{\pi}}} - \hat{\underset{\sim}{\pi}}\hat{\underset{\sim}{\pi}}']\underset{\sim}{X}\}^{-1} \quad . \tag{2.15}$$

Although the results given for the equations (2.9) and the covariance matrix (2.14) are applicable to any design matrix $\underset{\sim}{X}$ satisfying (2.6), they are of particular interest for the special case in which the columns of the matrix $\underset{\sim}{X}$ can be expressed in terms of the following types of indicator functions of the response profile indexes $\underset{\sim}{j}$ where $\underset{\sim}{j}' = (j_1, j_2, \ldots, j_d)$

$$x_{k_1 0 \ldots 0}(\underset{\sim}{j}) = x_{1k_1}(\underset{\sim}{j}) = \begin{cases} 1 \text{ if } j_1 = k_1 \\ -1 \text{ if } j_1 = L_1 \\ 0 \text{ otherwise} \end{cases} \quad \text{for } k_1 = 1,2,\ldots,(L_1 - 1) \tag{2.16}$$

$$\ldots \qquad \ldots \qquad \ldots$$

$$x_{00 \ldots k_d}(\underset{\sim}{j}) = x_{dk_d}(\underset{\sim}{j}) = \begin{cases} 1 \text{ if } j_d = k_d \\ -1 \text{ if } j_d = L_d \\ 0 \text{ otherwise} \end{cases} \quad \text{for } k_d = 1,2,\ldots,(L_d - 1)$$

and their respective higher order cross-products as generally represented by

$$x_{k_1 k_2 \ldots k_d}(\underset{\sim}{j}) = \prod_{g=1}^{d} \{x_{gk_g}(\underset{\sim}{j})\} \tag{2.17}$$

where $x_{g0}(\underset{\sim}{j}) \equiv 1$ by definition. In addition, $\underset{\sim}{X}$ is required to have a hierarchical structure that includes with any given interaction variable $x_{k_1 k_2 \ldots k_d}(\underset{\sim}{j})$, all interaction variables of the same order and all corresponding lower order interaction terms; i.e., it includes all terms in the family

$$F_{\underset{\sim}{\delta}(\underset{\sim}{k})} = \{\text{all } \pi_{h_1 h_2 \ldots h_d}(\underset{\sim}{j}) \text{ with } h_g = 0,1,\ldots,[\delta(h_g)][L_g - 1] \text{ for } g = 1,2,\ldots,d\} \tag{2.18}$$

where

$$\delta(h_g) = \begin{cases} 0 \text{ if } h_g = 0 \\ 1 \text{ if } h_g > 0 \end{cases} \tag{2.19}$$

and where $\underset{\sim}{k}$ and $\underset{\sim}{\delta}(\underset{\sim}{k})$ are defined by $\underset{\sim}{k}' = (k_1, k_2, \ldots, k_d)$ and $\underset{\sim}{\delta}'(\underset{\sim}{k}) = \{\delta(k_1), \delta(k_2), \ldots, \delta(k_d)\}$. Thus, the family $F_{\underset{\sim}{\delta}(\underset{\sim}{k})}$ contains

$$\nu\{\underset{\sim}{\delta}(\underset{\sim}{k})\} = [\prod_{g=1}^{d} L_g^{\{\delta(k_g)\}}] - 1 \tag{2.20}$$

such variables in all; and their simultaneous sets of values for all response profiles $\underset{\sim}{j}$ may be represented by the $\{rx[\nu\{\underset{\sim}{\delta}(\underset{\sim}{k})\}]\}$ submatrix $\underset{\sim}{X}_{\underset{\sim}{\delta}(\underset{\sim}{k})}$ of $\underset{\sim}{X}$ which has the respective members of $F_{\underset{\sim}{\delta}(\underset{\sim}{k})}$ as its columns. As a consequence of these considerations, it follows that the maximum likelihood equations (2.9) for such hierarchically structured $\underset{\sim}{X}$ matrices can be partitioned into a series of subset equations

$$\underset{\sim}{X}'_{\underset{\sim}{\delta}(\underset{\sim}{k})} \hat{\underset{\sim}{\pi}} = \underset{\sim}{X}'_{\underset{\sim}{\delta}(\underset{\sim}{k})} \underset{\sim}{p} \tag{2.21}$$

corresponding to the distinct families $F_{\underset{\sim}{\delta}(\underset{\sim}{k})}$, where any such family which is entirely contained within some other family is excluded from further consideration without loss of generality since those equations based on it are a subset of those based on the larger family which contain it. However, by adjoining the constraint

$$\underset{\sim}{1}' \hat{\underset{\sim}{\pi}} = \underset{\sim}{1}' \underset{\sim}{p} = 1 \tag{2.22}$$

to each of the equation sets (2.21), it follows that there exist non-singular $[\nu\{\underset{\sim}{\delta}(\underset{\sim}{k})\} + 1]$-dimensional square matrices $\underset{\sim}{C}_{\underset{\sim}{\delta}(\underset{\sim}{k})}$ which

can be used to transform the augmented equation sets (2.21) - (2.22) into the sets

$$\underset{\sim}{X}'_{\underset{\sim}{\delta}(\underset{\sim}{k})}\hat{\underset{\sim}{\pi}} = \underset{\sim}{C}_{\underset{\sim}{\delta}(\underset{\sim}{k})}\begin{bmatrix}\underset{\sim}{1}' \\ \underset{\sim}{X}'_{\underset{\sim}{\delta}(\underset{\sim}{k})}\end{bmatrix} \quad \hat{\underset{\sim}{\pi}} = \underset{\sim}{C}_{\underset{\sim}{\delta}(\underset{\sim}{k})}\begin{bmatrix}\underset{\sim}{1}' \\ \underset{\sim}{X}'_{\underset{\sim}{\delta}(\underset{\sim}{k})}\end{bmatrix} \quad \underset{\sim}{p} = \underset{\sim}{X}'_{\underset{\sim}{\delta}(\underset{\sim}{k})}\underset{\sim}{p} \qquad (2.23)$$

where the columns of the matrix $\underset{\sim}{X}_{\underset{\sim}{\delta}(\underset{\sim}{k})}$ correspond to the following types of indicator functions

$$\underset{\sim}{x}_{h_1 h_2 \ldots h_d}(\underset{\sim}{j};\underset{\sim}{\delta}(\underset{\sim}{k})) = \begin{cases} 1 \text{ if } j_g = h_g \text{ for all } g=1,2,\ldots,d \text{ such that } \delta(k_g)=1 \\ 0 \text{ otherwise} \end{cases} \qquad (2.24)$$

$$\text{with } h_g = \begin{cases} 0 \text{ if } \delta(k_g)=0 \\ 1,2,\ldots,L_g \text{ if } \delta(k_g)=1 \end{cases} \text{ for } g=1,2,\ldots,d \ .$$

Since the definition (2.24) implies that the matrix $\underset{\sim}{X}_{\underset{\sim}{\delta}(\underset{\sim}{k})}$ generates the

$$d_{\underset{\sim}{\delta}(\underset{\sim}{k})} = \sum_{g=1}^{d} \delta(k_g) \qquad (2.25)$$

dimensional sub-tables (which correspond to the attributes with $\delta(k_g)=1$) for the observed and predicted proportion vectors $\underset{\sim}{p}$ and $\hat{\underset{\sim}{\pi}}$, then the equations (2.23) mean that the maximum likelihood parameter estimator $\hat{\underset{\sim}{\beta}}$ must produce predicted values $\hat{\underset{\sim}{\pi}}$ which have such marginal distributions identical to those based on $\underset{\sim}{p}$. This structural aspect of the maximum likelihood equations (2.9) was initially presented by Birch [1963] in the context of three-dimensional contingency tables; and subsequently, it has been discussed by many other authors concerned with the analysis of log-linear models since it provides much of the theoretical justification for the use of the IPF algorithm to determine $\hat{\underset{\sim}{\pi}}$. Briefly stated, the IPF algorithm starts with an initial pseudo-estimator like

$$\hat{\underset{\sim}{\pi}}_0 = \left(\frac{1}{r}\right) \underset{\sim}{1}_r \tag{2.26}$$

which clearly satisfies the model (2.5), and then iteratively adjusts it to conform successively with the observed marginal configurations based on $\underset{\sim}{p}$ which are associated with the respective sets of equations (2.23) until convergence to the stable estimator $\hat{\underset{\sim}{\pi}}$ which satisfies all of the equations (2.9) has occurred. A more detailed description of the IPF algorithm together with its computational and statistical properties is given in BFH. In particular, the IPF procedure is known to converge to the solution $\hat{\underset{\sim}{\pi}}$ which satisfies the relationships implied by (2.5) and (2.9) provided such a solution exists; and as the sample size n becomes increasingly large, the probability of observing vectors $\underset{\sim}{p}$ for which solutions $\hat{\underset{\sim}{\pi}}$ exist approaches certainty when (2.5) is known to be valid; i.e., as $n \to \infty$

$$\Pr\{\text{IPF convergence given (2.5) is valid}\} \to 1 \ . \tag{2.27}$$

Thus, the IPF algorithm may be regarded as a methodologically effective computing procedure for solving the equations (2.9) for situations involving hierarchical models $\underset{\sim}{X}$ as defined by (2.16) - (2.18). Moreover, expression (2.15) can be used to obtain the consistent estimator $\underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}}$ for the estimated parameter vector $\hat{\underset{\sim}{\beta}}$. If the sample size n is sufficiently large that the vector $\hat{\underset{\sim}{\beta}}$ can be presumed approximately to have a multivariate normal distribution, then further analysis can be undertaken by means of the extended GSK approach described in Koch and Tolley [1975] which involves the testing of various linear hypotheses of interest involving $\hat{\underset{\sim}{\beta}}$ by generalized Wald [1943] statistics and the fitting of corresponding linear regression models by weighted least

squares. In particular, an appropriate test statistic for the hypothesis

$$H_0: \ \underset{\sim}{C}\underset{\sim}{\beta} = \underset{\sim}{0}_q \tag{2.28}$$

where $\underset{\sim}{C}$ is a known (qxt) matrix of full rank $q \leq t$ is

$$Q_{\underset{\sim}{C}} = Q_{\underset{\sim}{C}}(\hat{\underset{\sim}{\beta}}) = \hat{\underset{\sim}{\beta}}'\underset{\sim}{C}'[\underset{\sim}{C}\underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}}\underset{\sim}{C}']^{-1}\ \underset{\sim}{C}\hat{\underset{\sim}{\beta}} \tag{2.29}$$

which has approximately a chi-square distribution with D.F.=q in large samples. If $Q_{\underset{\sim}{C}}$ is sufficiently small to suggest that the data are supportive for (i.e., do not clearly contradict) the hypothesis (2.28), it then follows that the vector $\hat{\underset{\sim}{\beta}}$ may be characterized by the linear regression model

$$\underset{\sim}{E}_{\underset{\sim}{A}}\{\hat{\underset{\sim}{\beta}}\} = \underset{\sim}{\beta} = \underset{\sim}{X}_{\beta R}\underset{\sim}{\beta}_R \tag{2.30}$$

where $\underset{\sim}{X}_{\beta R}$ is an [tx(t-q)] full rank matrix which is the orthocomplement of $\underset{\sim}{C}$ in (2.28) (i.e., $\underset{\sim}{C}\underset{\sim}{X}_{\beta R} = \underset{\sim}{0}_{q,(t-q)}$ with $\underset{\sim}{0}_{q,(t-q)}$ being a [qx(t-q)] matrix of 0's) and "$\underset{\sim}{E}_{\underset{\sim}{A}}$" means "asymptotic expectation."

Within the context of this formulation, the GSK weighted least squares computational algorithms can be used to determine a BAN estimator

$$\underset{\sim}{b}_R = (\underset{\sim}{X}'_{\beta R}\underset{\hat{\underset{\sim}{\beta}}}{V}^{-1}\underset{\sim}{X}_{\beta R})^{-1}\underset{\sim}{X}'_{\beta R}\underset{\hat{\underset{\sim}{\beta}}}{V}^{-1}\hat{\underset{\sim}{\beta}} \tag{2.31}$$

for $\underset{\sim}{\beta}_R$ and its corresponding estimated covariance matrix

$$\underset{\sim}{V}_{\underset{\sim}{b}_R} = (\underset{\sim}{X}'_{\beta R}\underset{\hat{\underset{\sim}{\beta}}}{V}^{-1}\underset{\sim}{X}_{\beta R})^{-1} \tag{2.32}$$

In addition, the weighted least squares residual statistic

$$Q = Q(X_{\beta R}, \hat{\beta}) = (\hat{\beta} - b_R)' V_{\hat{\beta}}^{-1} (\hat{\beta} - b_R) \tag{2.33}$$

is identically equal to $Q_C(\hat{\beta})$ in (2.29) , which may be interpreted as an operational reflection of the equivalence between the goodness of fit of the model (2.30) and the validity of the hypothesis (2.28) with respect to β. On the basis of this equivalence, the statistic (2.29) can be used to test hypotheses pertaining to the goodness of fit of corresponding models (2.30) in a framework which does not require the actual fitting of such models; and, in certain cases, its effective application permits substantial savings in computational costs for those analyses that require interactive model building as opposed to the direct fitting of a specific model which is known a priori to be valid. However, it has the obvious disadvantage of only indicating whether the data are supportive for such models, but gives no information regarding the estimation of corresponding model parameters and their respective covariance matrices. Thus, given that an appropriate reduced model $X_{\beta R}$ has been identified by either a priori considerations or the interactive application of (2.29), the weighted least squares fitting procedures (2.31) - (2.33) can be used for such estimation purposes. After having fitted the model $X_{\beta R}$ to $\hat{\beta}$, further analysis of the data can be undertaken by applying to b_R the same basic methodological strategy which was outlined for $\hat{\beta}$ in (2.28) - (2.33). This successive reduction process can then be continued until an appropriate overall final model has been formulated.

Since the ultimate results b_R and V_{b_R} of this analysis are based on a two-stage estimation process involving IPF initially and then WLS, its computational and statistical properties merit further clarification

with respect to the original log-linear model framework (2.5) from which all of this discussion originated. From a conceptual point of view, this can be accomplished simply by introducing the structure of the model (2.30) for $\underset{\sim}{\beta}$ into the model (2.5) for $\underset{\sim}{\pi}$ to obtain

$$\begin{aligned}
\underset{\sim}{\pi} = \underset{\sim}{\pi}(\underset{\sim}{\beta}) &= \{\underset{\sim\sim}{\exp}(\underset{\sim\sim}{X}\underset{\sim}{\beta})\}/\{\underset{\sim}{1}_r'[\underset{\sim\sim}{\exp}(\underset{\sim\sim}{X}\underset{\sim}{\beta})]\} \qquad (2.34)\\
&= \{\underset{\sim\sim}{\exp}(\underset{\sim\sim}{X}\underset{\sim\sim}{X}_{\beta R}\underset{\sim}{\beta}_R)\}/\{\underset{\sim}{1}'[\underset{\sim\sim}{\exp}(\underset{\sim\sim}{X}\underset{\sim\sim}{X}_{\beta R}\underset{\sim}{\beta}_R)]\}\\
&= \{\underset{\sim\sim}{\exp}(\underset{\sim}{X}_R\underset{\sim}{\beta}_R)\}/\{\underset{\sim}{1}'[\underset{\sim\sim}{\exp}(\underset{\sim}{X}_R\underset{\sim}{\beta}_R)]\}\\
&= \underset{\sim}{\pi}(\underset{\sim}{\beta}_R)
\end{aligned}$$

where $\underset{\sim}{X}_R = \underset{\sim\sim}{X}\underset{\sim\sim}{X}_{\beta R}$ is a reduced [rx(t-q)] design matrix satisfying the same conditions as $\underset{\sim}{X}$ in (2.5), although not necessarily the hierarchical family structure summarized by (2.18). In other words, this two-stage estimation procedure yields estimators $\underset{\sim}{b}_R$ for the parameters $\underset{\sim}{\beta}_R$ which correspond to a log-linear model of the type (2.5); and as such, it represents an effective method for fitting such models for situations where the IPF algorithm is not applicable because the corresponding design matrix $\underset{\sim}{X}_R$ does not have the hierarchical structure (2.18).

Additional insights can be gained by substituting expression (2.15) for $\underset{\sim\sim}{V}_{\hat{\underset{\sim}{\beta}}}$ into expression (2.31) for $\underset{\sim}{b}_R$ to obtain

$$\begin{aligned}
\underset{\sim}{b}_R &= (\underset{\sim\sim}{X}_{\beta R}'\underset{\sim\sim}{V}_{\hat{\underset{\sim}{\beta}}}^{-1}\underset{\sim\sim}{X}_{\beta R})^{-1}\underset{\sim\sim}{X}_{\beta R}'\underset{\sim\sim}{V}_{\hat{\underset{\sim}{\beta}}}^{-1}\hat{\underset{\sim}{\beta}} \qquad (2.35)\\
&= (\underset{\sim\sim}{X}_{\beta R}'\underset{\sim}{X}'[\underset{\sim\sim}{D}_{\hat{\underset{\sim}{\pi}}}-\hat{\underset{\sim}{\pi}}\hat{\underset{\sim}{\pi}}']\underset{\sim\sim}{X}\underset{\sim\sim}{X}_{\beta R})^{-1}\underset{\sim}{X}_{\beta R}'\underset{\sim}{X}'[\underset{\sim\sim}{D}_{\hat{\underset{\sim}{\pi}}}-\hat{\underset{\sim}{\pi}}\hat{\underset{\sim}{\pi}}']\underset{\sim\sim}{X}\hat{\underset{\sim}{\beta}}\\
&= (\underset{\sim}{X}_R'[\underset{\sim\sim}{D}_{\hat{\underset{\sim}{\pi}}}-\hat{\underset{\sim}{\pi}}\hat{\underset{\sim}{\pi}}']\underset{\sim}{X}_R)^{-1}\underset{\sim}{X}_R'[\underset{\sim\sim}{D}_{\hat{\underset{\sim}{\pi}}}-\hat{\underset{\sim}{\pi}}\hat{\underset{\sim}{\pi}}'][\underset{\sim\sim}{\log}_e\{\underset{\sim}{\pi}(\hat{\underset{\sim}{\beta}})\}]
\end{aligned}$$

where $\underset{\sim}{\log}_e$ transforms a vector to the corresponding vector of natural logarithms, and relating this result to those given by Grizzle and Williams [1972] for the standard one-stage application of the GSK procedure for fitting the log-linear model (2.34) directly to $\underset{\sim}{p}$ (although as mentioned previously, much larger sample sizes are required to ensure the validity of this approach than for the two-stage method currently under discussion, given that (2.34) is a priori known to be appropriate). In this regard, let $\underset{\sim}{X}_{RC}$ be an [rx(r-u-1)] full rank matrix which is orthogonal to both $\underset{\sim}{X}_R$ and $\underset{\sim}{1}_R$ where u = (t-q).

From (2.34), it then follows that the log-linear functions $\underset{\sim}{F}$ of the observed proportion vector $\underset{\sim}{p}$ which are defined by

$$\underset{\sim}{F} = \begin{bmatrix} \underset{\sim}{F}_1 \\ \underset{\sim}{F}_2 \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}_R' \\ \underset{\sim}{X}_{RC}' \end{bmatrix} [\underset{\sim}{\log}_e(\underset{\sim}{p})] \tag{2.36}$$

may be characterized by the model

$$\underset{\sim}{E}_A\{\underset{\sim}{F}\} = \begin{bmatrix} \underset{\sim}{X}_R' \\ \underset{\sim}{X}_{RC}' \end{bmatrix} [\underset{\sim}{X}_R \underset{\sim}{\beta}_R] = \begin{bmatrix} \underset{\sim}{X}_R'\underset{\sim}{X}_R \\ \underset{\sim}{0}_{(r-u-1),u} \end{bmatrix} \underset{\sim}{\beta}_R \quad . \tag{2.37}$$

Given this formulation, a BAN estimator $\underset{\sim}{\overline{b}}_R$ for $\underset{\sim}{\beta}_R$ can be obtained by using the GSK weighted least squares computational algorithms as summarized in (2.31) - (2.33), but with $\underset{\sim}{F}$ replacing $\underset{\sim}{\beta}$ and the consistent estimator

$$\underset{\sim}{V}_F = \frac{1}{n} \begin{bmatrix} \underset{\sim}{X}_R' \\ \underset{\sim}{X}_{RC}' \end{bmatrix} \underset{\sim}{D}_p^{-1} [\underset{\sim}{D}_p - \underset{\sim}{p}\underset{\sim}{p}'] \underset{\sim}{D}_p^{-1} [\underset{\sim}{X}_R \underset{\sim}{X}_{RC}] \tag{2.38}$$

$$= \frac{1}{n} \begin{bmatrix} \underset{\sim}{X}_R' \\ \underset{\sim}{X}_{RC}' \end{bmatrix} \underset{\sim}{D}_p^{-1} [\underset{\sim}{X}_R, \underset{\sim}{X}_{RC}]$$

for its covariance matrix replacing V_{β}. Thus,

$$\bar{b}_R = \left\{ [X'_R X_R, 0_{u,(r-u-1)}] V_F^{-1} \begin{bmatrix} X'_R X_R \\ 0_{(r-u-1),u} \end{bmatrix} \right\}^{-1} [X'_R X_R, 0_{u,(r-u-1)}] V_F^{-1} F \quad . \tag{2.39}$$

Since it can be verified that

$$V_F^{-1} = n \begin{bmatrix} (X'_R X_R)^{-1} X'_R \\ (X'_{RC} X_{RC})^{-1} X'_{RC} \end{bmatrix} [D_p - pp'][X_R (X'_R X_R)^{-1} X_{RC} (X'_{RC} X_{RC})^{-1}] \tag{2.40}$$

by using direct multiplication together with the identity

$$[X_R (X'_R X_R)^{-1} X'_R + X_{RC} (X'_{RC} X_{RC})^{-1} X'_{RC}] = [I_r - \frac{1}{r} 1_{rr}] \quad , \tag{2.41}$$

it follows that the expression (2.39) for $\bar{b}_R$ can be simplified to

$$\begin{aligned} \bar{b}_R &= \{X'_R [D_p - pp'] X_R\}^{-1} \{X'_R [D_p - pp'] [X_R (X'_R X_R)^{-1} X'_R + X_{RC} (X'_{RC} X_{RC})^{-1} X'_{RC}] [\log_e (p)]\} \\ &= \{X'_R [D_p - pp'] X_R\}^{-1} \{X'_R [D_p - pp'] [I_r - \frac{1}{r} 1_{rr}] [\log_e (p)]\} \\ &= \{X'_R [D_p - pp'] X_R\}^{-1} \{X'_R [D_p - pp'] [\log_e (p)]\}. \end{aligned} \tag{2.42}$$

However, this result demonstrates that expression (2.42) for the standard GSK estimator $\bar{b}_R$ and expression (2.35) for the two-stage estimator b_R are identical except for the fact that the former involves the observed proportion vector p while the latter involves the IPF predicted proportion vector $\pi(\hat{\beta})$ based on the first stage model X with the hierarchical family structure (2.18). Thus, b_R may be obtained by applying the standard one-stage GSK procedure for fitting log-linear models outlined in (2.36) - (2.38) to the pseudo-contingency table involving the set of IPF predicted frequencies $n\pi(\hat{\beta})$. Similar remarks of this type also apply to the estimated

covariance matrix $V_{\underset{\sim}{b}_R}$ in (2.32) for $\underset{\sim}{b}_R$ which may be written as

$$V_{\underset{\sim}{b}_R} = \frac{1}{n}\{X_R'[D_{\hat{\pi}} - \hat{\pi}\,\hat{\pi}']X_R\}^{-1} \tag{2.43}$$

and the correspondingly analogous estimated covariance matrix

$$V_{\underset{\sim}{\bar{b}}_R} = \frac{1}{n}\{X_R[D_p - p\,p']X_R\}^{-1} \tag{2.44}$$

for $\underset{\sim}{\bar{b}}_R$. With these considerations of computational equivalence in mind, it then follows that the various aspects of interactive model fitting described for $\hat{\beta}$ in the context of (2.28) - (2.33) can be undertaken by the suitable application of the GSK procedures to the pseudo-contingency table involving the IPF predicted frequencies $n\pi(\hat{\beta})$. Moreover, one of the useful by-products of such analysis are the estimators $\hat{\beta}$ and $V_{\hat{\beta}}$ themselves as they pertain to the first stage hierarchical model X since the IPF algorithm is specifically directed at determining $n\pi(\hat{\beta})$ rather than these quantities. Thus, this two-stage extended GSK approach provides a useful framework for

i. Computing the estimated parameter vector $\hat{\beta}$ and its corresponding estimated covariance matrix $V_{\hat{\beta}}$ for hierarchical log-linear models X for which predicted values $n\pi(\hat{\beta})$ have been previously determined by IPF;

ii. Fitting non-hierarchical models X_R by first identifying an appropriate hierarchical model X_H from which X_R can be obtained by a non-singular transformation $X_R = X_H X_{HR}$, and then fitting X_H by IPF to determine $n\pi(\hat{\beta}_H)$ which are then analyzed by WLS to determine the estimators b_R and V_{b_R} in (2.31) - (2.32).

However, (i) may be regarded as a special case of (ii) by assuming $X_R = X_H$ where X_H is hierarchical. In addition, if X_H is complete (or saturated) in the sense of containing all the main effects and interactions (2.16) - (2.17), then

(ii) becomes identical with the standard one stage GSK procedure since $n\underset{\sim}{\pi}(\hat{\underset{\sim}{\beta}}_H) = \underset{\sim}{n}$ for this situation. Finally, although the estimators $\underset{\sim}{b}_R$ have reasonable statistical properties (i.e., they can be shown to be BAN estimators by arguments like those given in Koch and Tolley [1975], Appendix III), they are not the strict maximum likelihood estimators $\hat{\beta}_R$ which are defined by (2.9). If the sample size n is sufficiently large that $\underset{\sim}{b}_R$ can be presumed approximately to have a multivariate normal distribution, this consideration poses no real difficulty because the asymptotic equivalence of $\underset{\sim}{b}_R$ and $\hat{\underset{\sim}{\beta}}_R$ under such conditions causes their respective results to be suitably similar for most practical purposes. On the other hand, if such assumptions are not realistic, then $\underset{\sim}{b}_R$ is no longer valid; and hence $\hat{\underset{\sim}{\beta}}_R$ must be determined. For this purpose, one useful method is direct maximization of the likelihood function (2.3) itself by successive approximation numerical methods like those given in Kaplan and Elston [1972] with $\underset{\sim}{b}_R$ representing an initial starting value which should be in most applications adequately close to $\hat{\underset{\sim}{\beta}}_R$ for convergence to occur rapidly. As soon as $\hat{\underset{\sim}{\beta}}_R$ has been obtained by this procedure, its corresponding predicted frequency vector $n\underset{\sim}{\pi}(\hat{\underset{\sim}{\beta}}_R)$ can then be analyzed by the extended GSK approach in the sense of (ii) in order to compute its estimated covariance matrix $\underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}}$. Thus, regardless of whether the estimation of model parameters is considered from the point of view of the standard GSK procedure, IPF, the extended GSK procedures, or strict maximum likelihood, a consistent estimator for the corresponding covariance structure can be determined by using the two-stage procedure based on $\underset{\sim}{b}_R$ and $\underset{\sim}{V}_{\underset{\sim}{b}_R}$ as formulated in (2.35) and (2.43) respectively, which may be conceptually regarded as predicted pseudo-contingency table oriented Functional Asymptotic Regression Methodology (FARM).

To complete this part of the discussion, an estimator for the covariance matrix of the vector of predicted proportions $\hat{\underset{\sim}{\pi}} = \underset{\sim}{\pi}(\hat{\underset{\sim}{\beta}})$ will be given to supplement the estimator $\underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}}$ given in (2.15) for the vector of estimated parameters $\hat{\underset{\sim}{\beta}}$. This result is obtained via the well-known δ-method as based on the first order linear Taylor series approximating counterpart function vector

$$\underset{\sim}{\pi}_L(\hat{\underset{\sim}{\beta}}) = \underset{\sim}{\pi}(\underset{\sim}{\beta}) + \left[\frac{d\underset{\sim}{\pi}(\underset{\sim}{y})}{d\underset{\sim}{y}}\bigg|_{\underset{\sim}{y}=\underset{\sim}{\beta}}\right]'[(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})]. \tag{2.45}$$

Thus, it follows that the asymptotic covariance matrix of $\hat{\underset{\sim}{\pi}}$ is given by

$$\underset{\sim}{V}_{\hat{\underset{\sim}{\pi}}}(\underset{\sim}{\beta}) = \left[\frac{d\underset{\sim}{\pi}(\underset{\sim}{y})}{d\underset{\sim}{y}}\bigg|_{\underset{\sim}{y}=\underset{\sim}{\beta}}\right]'[\underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}}(\underset{\sim}{\pi})]\left[\frac{d\underset{\sim}{\pi}(\underset{\sim}{y})}{d\underset{\sim}{y}}\bigg|_{\underset{\sim}{y}=\underset{\sim}{\beta}}\right], \tag{2.46}$$

which may be rewritten as

$$\underset{\sim}{V}_{\hat{\underset{\sim}{\pi}}}(\underset{\sim}{\pi}) = \frac{1}{n}\{[\underset{\sim}{D}_{\underset{\sim}{\pi}} - \underset{\sim}{\pi}\,\underset{\sim}{\pi}']\underset{\sim}{X}\}\{\underset{\sim}{X}'[\underset{\sim}{D}_{\underset{\sim}{\pi}} - \underset{\sim}{\pi}\,\underset{\sim}{\pi}']\underset{\sim}{X}\}^{-1}\{\underset{\sim}{X}'[\underset{\sim}{D}_{\underset{\sim}{\pi}} - \underset{\sim}{\pi}\,\underset{\sim}{\pi}']\} \tag{2.47}$$

by applying (2.11). Since $\hat{\underset{\sim}{\pi}}$ is a consistent estimator for $\underset{\sim}{\pi}$, a consistent estimator for the covariance matrix of $\hat{\underset{\sim}{\pi}}$ is

$$\underset{\sim}{V}_{\hat{\underset{\sim}{\pi}}} = \underset{\sim}{V}_{\hat{\underset{\sim}{\pi}}}(\hat{\underset{\sim}{\pi}}) = \frac{1}{n}\{[\underset{\sim}{D}_{\hat{\underset{\sim}{\pi}}} - \hat{\underset{\sim}{\pi}}\,\hat{\underset{\sim}{\pi}}']\underset{\sim}{X}\}\{\underset{\sim}{X}'[\underset{\sim}{D}_{\hat{\underset{\sim}{\pi}}} - \hat{\underset{\sim}{\pi}}\,\hat{\underset{\sim}{\pi}}']\underset{\sim}{X}\}^{-1}\{\underset{\sim}{X}'[\underset{\sim}{D}_{\hat{\underset{\sim}{\pi}}} - \hat{\underset{\sim}{\pi}}\,\hat{\underset{\sim}{\pi}}']\}. \tag{2.48}$$

Thus, $\hat{\underset{\sim}{\pi}}$ and $\underset{\sim}{V}_{\hat{\underset{\sim}{\pi}}}$ can be subsequently used as the basis for other FARM analyses. In this regard, one class of examples would be repeated measurement research designs as discussed in Koch et al. [1974] where linear hypotheses

$$H_0: \ \underset{\sim}{A}\,\underset{\sim}{\pi} = \underset{\sim}{0} \tag{2.49}$$

are of interest with $\underset{\sim}{A}$ being a known coefficient matrix which produces linear functions of the first order marginal probabilities for the d-dimensional

contingency table under consideration. For these situations, attention would then be directed at the estimators

$$\underset{\sim}{F} = \underset{\sim}{A}\,\hat{\underset{\sim}{\pi}}, \quad \underset{\sim}{V}_{\underset{\sim}{F}} = \underset{\sim}{A}\,\hat{\underset{\sim}{V}}_{\underset{\sim}{\pi}}\,\underset{\sim}{A}' \tag{2.50}$$

with $\underset{\sim}{V}_{\underset{\sim}{F}}$ being assumed to be asymptotically non-singular by suitable construction of $\underset{\sim}{A}$.

2.2 Stratified Simple Random Sampling from Multiple Populations of Known Structure

This section is concerned with the covariance structure of estimators based on log-linear models pertaining to the multivariate relationships among a specific set of d attributes and their variation across a set of s sub-populations. In this regard, let $i=1,2,\ldots,s$ index the sub-populations and let $\underset{\sim}{j}=(j_1,j_2,\ldots,j_d)$ index the response profiles as in Section 2.1. If a stratified simple random sample involving $n_{1.},n_{2.},\ldots,n_{s.}$ elements from the i=1,2,...,s sub-populations respectively is selected, the resulting data can be summarized with a (d+1)-dimensional contingency table in which $n_{i\underset{\sim}{j}}$ denotes the observed frequency of the multivariate response profile $\underset{\sim}{j}$ for the elements from the i-th sub-population. The observed frequency vectors $\{\underset{\sim}{n}_i\}$ where

$$\underset{\sim}{n}_i' = (n_{i,11\ldots1},\ldots,n_{i,j_1j_2\ldots j_d},\ldots,n_{i,L_1L_2\ldots L_d}) \tag{2.51}$$

are assumed to be statistically independent and to follow corresponding multinomial distributions with parameters $\{\underset{\sim}{n}_{i.}\}$ and $\{\underset{\sim}{\pi}_i\}$ where

$$\underset{\sim}{\pi}_i' = (\pi_{i,11,\ldots1},\ldots,\pi_{i,j_1j_2\ldots j_d},\ldots,\pi_{i,L_1L_2\ldots L_d}) \tag{2.52}$$

with $\pi_{\underset{\sim}{i},j_1j_2\ldots j_d}$ representing the probability that a randomly selected element from the i-th sub-population is classified into the $\underset{\sim}{j}$-th response

profile. Thus, the relevant product multinomial model for the overall frequency vector $\underset{\sim}{n}$ where

$$\underset{\sim}{n}' = (\underset{\sim 1}{n}', \underset{\sim 2}{n}', \ldots, \underset{\sim s}{n}') \tag{2.53}$$

is given by

$$\phi = \prod_{i=1}^{s} \left[n_{i\cdot}! \prod_{j_1=1}^{L_1} \prod_{j_2=1}^{L_2} \cdots \prod_{j_d=1}^{L_d} \{ \pi_{i,j_1 j_2 \cdots j_d}^{n_{i,j_1 j_2 \cdots j_d}} / n_{i,j_1 j_2 \cdots j_d}! \} \right] \tag{2.54}$$

with the s constraints

$$\sum_{j_1=1}^{L_1} \sum_{j_2=1}^{L_2} \cdots \sum_{j_d=1}^{L_d} \pi_{i,j_1 j_2 \cdots j_d} = 1 \text{ for } i=1,2,\ldots,s \quad . \tag{2.55}$$

Each of the sub-populations is assumed to be characterized by log-linear models as defined in (2.5); i.e.,

$$\underset{\sim i}{\pi} = \underset{\sim i}{\pi}(\underset{\sim}{\beta}) = \{\underset{\sim\sim}{\exp}(\underset{\sim i}{X}\underset{\sim}{\beta})\}/\{\underset{\sim r}{1}' [\underset{\sim\sim}{\exp}(\underset{\sim i}{X}\underset{\sim}{\beta})]\} \tag{2.56}$$

for i=1,2,...,s with $\underset{\sim i}{X}$ being the corresponding sub-matrix for the i-th sub-population of an appropriate (rs x t) design matrix $\underset{\sim}{X}$ where

$$\underset{\sim}{X}' = (\underset{\sim 1}{X}', \underset{\sim 2}{X}', \ldots, \underset{\sim s}{X}') \quad . \tag{2.57}$$

In addition, each of the $\underset{\sim i}{X}$ are assumed without loss of generality to be orthogonal to $\underset{\sim r}{1}$ which means that $\underset{\sim}{X}$ satisfies

$$\underset{\sim}{X}' [\underset{\sim rr}{1} \otimes \underset{\sim s}{I}] = \underset{\sim t}{0} \tag{2.58}$$

where $\otimes$ denotes Kronecker product matrix multiplication. Finally, the separate models (2.56) may be simultaneously expressed for the overall

response profile vector $\underset{\sim}{\pi}$ where

$$\underset{\sim}{\pi}' = (\underset{\sim}{\pi}_1', \underset{\sim}{\pi}_2', \ldots, \underset{\sim}{\pi}_s') \tag{2.59}$$

by the composite log-linear model

$$\underset{\sim}{\pi} = \underset{\sim}{\pi}(\underset{\sim}{\beta}) = \underset{\sim}{D}_{\underset{\sim}{\eta}}^{-1} \{\underset{\sim\sim\sim}{\exp}(\underset{\sim\sim}{X}\underset{\sim}{\beta})\} \tag{2.60}$$

where $\underset{\sim}{\eta} = [\underset{\sim rr}{1} \times \underset{\sim s}{I}]\{\underset{\sim\sim\sim}{\exp}(\underset{\sim\sim}{X}\underset{\sim}{\beta})\}$.

As in Section 2.1, the method of maximum likelihood will be used to form estimators $\hat{\underset{\sim}{\beta}}$ for $\underset{\sim}{\beta}$. From (2.54) - (2.60), it follows that

$$\frac{d}{d\underset{\sim}{\beta}'}[\log \phi] = \frac{d}{d\underset{\sim}{\beta}'}[\underset{\sim}{n}'\underset{\sim\sim}{X}\underset{\sim}{\beta} - \underset{\sim}{n}'(\underset{\sim\sim\sim}{\log_e}\{[\underset{\sim rr}{1} \otimes \underset{\sim s}{I}][\underset{\sim\sim\sim}{\exp}(\underset{\sim\sim}{X}\underset{\sim}{\beta})]\})] \tag{2.61}$$

$$= \underset{\sim}{n}'\underset{\sim\sim}{X} - \underset{\sim}{n}'\underset{\sim\eta}{D}^{-1}[\underset{\sim rr}{1} \otimes \underset{\sim s}{I}]\underset{\sim\sim\sim}{D}_{\exp(\underset{\sim\sim}{X}\underset{\sim}{\beta})}\underset{\sim\sim}{X}$$

$$= \underset{\sim}{n}'\underset{\sim\sim}{X} - \underset{\sim}{n}'[\underset{\sim rr}{1} \otimes \underset{\sim s}{I}]\underset{\sim\sim}{D}_{\underset{\sim}{\pi}(\underset{\sim}{\beta})}\underset{\sim\sim}{X}$$

$$= \underset{\sim}{n}'\underset{\sim\sim}{X} - [\underset{\sim r}{1} \otimes \underset{\sim}{n}_{\cdot}]'\underset{\sim\sim}{D}_{\underset{\sim}{\pi}(\underset{\sim}{\beta})}\underset{\sim\sim}{X}$$

where $\underset{\sim}{n}_{\cdot}' = (n_{1\cdot}, n_{2\cdot}, \ldots, n_{s\cdot})$ represents the vector of sample sizes from the respective sub-populations. Thus, if

$$\underset{\sim}{m}(\underset{\sim}{\beta}) = [\underset{\sim\sim}{D}_{\underset{\sim}{\pi}(\underset{\sim}{\beta})}][\underset{\sim r}{1} \otimes \underset{\sim}{n}_{\cdot}] \tag{2.62}$$

denotes the expected frequency vector for $\underset{\sim}{n}$ under the model (2.54) - (2.60), then the maximum likelihood equations for $\underset{\sim}{\beta}$ may be compactly written as

$$\underset{\sim}{X}'\hat{\underset{\sim}{m}} = \underset{\sim}{X}'\underset{\sim}{n} \tag{2.63}$$

where $\hat{\underset{\sim}{m}} = \underset{\sim}{m}(\hat{\underset{\sim}{\beta}})$; and thus have the same general form as the equations (2.9) which pertained to a single population. As a consequence of this

result, it follows that their solution may be undertaken in terms of computational algorithms which are similar to those given in Section 2.1. More specifically, IPF may be used to solve the equations (2.63) whenever $\underset{\sim}{X}$ has an hierarchical structure analogous to that outlined in (2.16) - (2.18) with respect to indicator functions based on the composite (sub-population x response profile) indexes (i,j). In this context, the scope of such considerations also applies to situations which allow the further partition of i as a vector subscript $\underset{\sim}{i}=(i_1,i_2,\ldots,i_c)$ to reflect a c-dimensional factor classification of the sub-populations.

The covariance matrix $\underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}}(\underset{\sim}{\pi})$ for $\hat{\underset{\sim}{\beta}}$ may be obtained by forming the negative inverse of the Fisher Information Matrix via methods like those shown in (2.11). Thus,

$$\underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}}(\underset{\sim}{\pi}) = -\{\frac{d^2}{d\underset{\sim}{\beta}d\underset{\sim}{\beta}'}[\log \phi]\}^{-1} \qquad (2.64)$$

$$= -\{\frac{d}{d\underset{\sim}{\beta}'}[-\underset{\sim}{X}'\underset{\sim}{D}_{\underset{\sim}{\pi}(\underset{\sim}{\beta})}(\underset{\sim}{1}_r \otimes \underset{\sim}{n}_{\cdot})]\}^{-1}$$

$$= \underset{\sim}{X}'\underset{\sim}{D}_{(\underset{\sim}{1}_r \otimes \underset{\sim}{n}_{\cdot})}[\frac{d}{d\underset{\sim}{\beta}'}\{\underset{\sim}{\pi}(\underset{\sim}{\beta})\}]\}^{-1}$$

$$= \left\{\underset{\sim}{X}'[\underset{\sim}{I}_r \otimes \underset{\sim}{D}_{\underset{\sim}{n}_{\cdot}}]\begin{bmatrix} (\underset{\sim}{D}_{\underset{\sim}{\pi}_1} - \underset{\sim}{\pi}_1\underset{\sim}{\pi}_1')\underset{\sim}{X}_1 \\ (\underset{\sim}{D}_{\underset{\sim}{\pi}_2} - \underset{\sim}{\pi}_2\underset{\sim}{\pi}_2')\underset{\sim}{X}_2 \\ \cdots \\ (\underset{\sim}{D}_{\underset{\sim}{\pi}_s} - \underset{\sim}{\pi}_s\underset{\sim}{\pi}_s')\underset{\sim}{X}_s \end{bmatrix}\right\}^{-1}$$

$$= \{\sum_{i=1}^{s} n_{i\cdot}\underset{\sim}{X}_i'[\underset{\sim}{D}_{\underset{\sim}{\pi}_i} - \underset{\sim}{\pi}_i\underset{\sim}{\pi}_i']\underset{\sim}{X}_i\}^{-1} \quad ,$$

which is directly analogous to the result given for a single population in (2.14). Since $\hat{\underset{\sim}{\pi}} = \underset{\sim}{\pi}(\hat{\underset{\sim}{\beta}})$ is a consistent estimator for $\hat{\underset{\sim}{\beta}}$, a consistent

estimator for the covariance matrix $\underset{\sim}{V}_{\hat{\beta}}(\underset{\sim}{\pi})$ in (2.64) is

$$\underset{\sim}{V}_{\hat{\beta}} = \underset{\sim}{V}_{\hat{\beta}}(\hat{\underset{\sim}{\pi}}) = \{ \sum_{i=1}^{s} n_i \underset{\sim i}{X}' [D_{\hat{\underset{\sim}{\pi}}_i} - \hat{\underset{\sim i}{\pi}} \hat{\underset{\sim i}{\pi}}'] \underset{\sim i}{X} \}^{-1} \qquad (2.65)$$

If the overall sample size $n = \sum_{i=1}^{s} n_{i.}$ is sufficiently large that the vector $\hat{\underset{\sim}{\beta}}$ can be presumed approximately to have a multivariate normal distribution (which can result from either each of the $n_{i.}$ being large or from the extent to which the structure of $\underset{\sim}{X}$ links the data from the separate sub-populations together through $\underset{\sim}{\beta}$), then its further analysis with respect to various linear hypotheses and models of interest can be undertaken by the FARM approach described in Section 2.1. Moreover, the corresponding computations can be formulated in terms of the application of the standard one-stage GSK procedure for such models to the pseudo-contingency table involving the set of IPF (or otherwise determined) predicted frequencies $\hat{\underset{\sim}{m}}$. In this regard, the analysis is directed at the functions

$$\hat{\underset{\sim}{F}} = \begin{bmatrix} \hat{\underset{\sim}{F}}_1 \\ \hat{\underset{\sim}{F}}_2 \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}' \\ \underset{\sim C}{X}' \end{bmatrix} [\log_e(\hat{\underset{\sim}{\pi}})] \qquad (2.66)$$

where $\underset{\sim C}{X}$ is an (rs x u), where u=(rs-s-t), full rank matrix which is orthogonal to both $\underset{\sim}{X}$ and $[\underset{\sim rr}{1} \otimes \underset{\sim s}{I}]$. Then, the estimator $\hat{\underset{\sim}{\beta}}$ is obtained by using the GSK weighted least squares computational algorithms to fit the model

$$E_{\underset{\sim A}{}}\{\hat{\underset{\sim}{F}}\} = \begin{bmatrix} (\underset{\sim}{X}'\underset{\sim}{X}) \\ \underset{\sim ut}{0} \end{bmatrix} \underset{\sim}{\beta} \qquad (2.67)$$

to $\hat{\underset{\sim}{F}}$ with weights being based on the pseudo-variance matrix $\underset{\sim\sim}{V}_{PF}$ where

$$\underset{\sim\sim}{V}_{PF} = \begin{bmatrix} \underset{\sim}{X}' \\ \underset{\sim}{X}_C' \end{bmatrix} (\underset{\sim}{I} \times \underset{\sim}{D}_{\underset{\sim}{n}.})^{-1} \underset{\sim}{D}_{\underset{\sim}{p}}^{-1} [\underset{\sim}{X}, \underset{\sim}{X}_C] \quad . \tag{2.68}$$

Finally, the covariance matrix $\underset{\sim\sim}{V}_{\hat{\underset{\sim}{\pi}}}(\underset{\sim}{\pi})$ for the predicted proportions $\hat{\underset{\sim}{\pi}} = \underset{\sim}{\pi}(\hat{\underset{\sim}{\beta}})$ for this model (2.60) can be obtained by applying the δ-method as outlined in (2.45) - (2.46). Thus,

$$\underset{\sim\sim}{V}_{\hat{\underset{\sim}{\pi}}}(\underset{\sim}{\pi}) = [\underset{\sim\sim}{H}(\underset{\sim}{\pi})] \{ \sum_{i=1}^{s} \underset{\sim i}{n} \underset{\sim i}{X}_i' [\underset{\sim\sim}{D}_{\underset{\sim}{\pi}_i} - \underset{\sim i}{\pi} \underset{\sim i}{\pi}_i'] \underset{\sim i}{X}_i \}^{-1} [\underset{\sim\sim}{H}(\underset{\sim}{\pi})]' \tag{2.69}$$

where

$$\underset{\sim\sim}{H}(\underset{\sim}{\pi}) = \begin{bmatrix} (\underset{\sim\sim}{D}_{\underset{\sim}{\pi}_1} - \underset{\sim 1}{\pi} \underset{\sim 1}{\pi}_1') \underset{\sim 1}{X}_1 \\ (\underset{\sim\sim}{D}_{\underset{\sim}{\pi}_2} - \underset{\sim 2}{\pi} \underset{\sim 2}{\pi}_2') \underset{\sim 2}{X}_2 \\ \cdots \\ (\underset{\sim\sim}{D}_{\underset{\sim}{\pi}_s} - \underset{\sim s}{\pi} \underset{\sim s}{\pi}') \underset{\sim s}{X} \end{bmatrix} \tag{2.70}$$

Since $\hat{\underset{\sim}{\pi}}$ is a consistent estimator for $\underset{\sim}{\pi}$, a consistent estimator for the covariance matrix of $\hat{\underset{\sim}{\pi}}$ is

$$\underset{\sim\sim}{V}_{\hat{\underset{\sim}{\pi}}} = \underset{\sim\sim}{V}_{\hat{\underset{\sim}{\pi}}}(\hat{\underset{\sim}{\pi}}) = [\underset{\sim\sim}{H}(\hat{\underset{\sim}{\pi}})] \{ \sum_{i=1}^{s} \underset{\sim i}{n} \underset{\sim i}{X}_i' [\underset{\sim\sim}{D}_{\hat{\underset{\sim}{\pi}}_i} - \hat{\underset{\sim i}{\pi}} \hat{\underset{\sim i}{\pi}}_i'] \underset{\sim i}{X}_i \}^{-1} [\underset{\sim\sim}{H}(\hat{\underset{\sim}{\pi}})] \tag{2.71}$$

Other FARM analyses of interest can then be subsequently formulated in terms of $\hat{\underset{\sim}{\pi}}$ and $\underset{\sim\sim}{V}_{\hat{\underset{\sim}{\pi}}}$.

In concluding this discussion, it is worthwhile to note that the models like (2.60) considered here can also be fitted to data from the simple random sampling situation considered in Section 2.1 by directing attention at the conditional distribution of one set of attributes (which are the response profiles in (2.60)) given the respective fixed levels of another (which are the sub-populations in (2.60)). Since such conditional distributions have the form (2.54) if the overall joint unconditional distribution has the form (2.3),

it then follows that the results (2.61) - (2.71) can be applied to this conditional situation in exactly the same manner as if the sample had been stratified originally. Thus, with this point of view in mind, the models like (2.5) pertain conceptually to joint response profile probability distributions whereas the models like (2.60) pertain to conditional response profile probability distributions. When the sample is not restricted by stratification, both of these can be fitted by applying the methods described in Section 2.1 since conditional probabilities are simple ratio functions of joint probabilities for which the structure (2.60) is implied by (2.5). However, if the sample is stratified, then the observed data only provide information about within sub-population distributions; and hence the class of log-linear models which are applicable is correspondingly restricted to (2.60). This does not mean that models like (2.5) cannot be fitted in such situations, but rather that they cannot be fitted in the context outlined in Section 2.1. The critical issue is that if the relative distribution of the sub-populations is known in the sense of stratum weights $W_1, W_2, \ldots, W_s$, then the predicted proportion vector $\hat{\underset{\sim}{\pi}}$ can be transformed to an estimator of the joint probability distribution in the overall population by the matrix multiplication

$$\hat{\underset{\sim}{\pi}}_J = [\underset{\sim}{I}_r \otimes \underset{\sim}{D}_{\underset{\sim}{W}}] \hat{\underset{\sim}{\pi}} \qquad (2.72)$$

where $\underset{\sim}{W}' = (W_1, W_2, \ldots, W_s)$. Thus, the model (2.5) can be used to characterize joint probabilities in populations from which stratified samples have been selected to the extent that the structure of the corresponding stratum weight vector $\underset{\sim}{W}$ governs their application to $\hat{\underset{\sim}{\pi}}_J$ in (2.72). Other aspects of this type of analysis will be considered in the context of "raking" procedures in sub-section 2.5.

2.3 Complex Probability Random Sampling from a Population of Known Structure

Many applications where log-linear models are useful involve data obtained from complex stratified multi-stage cluster random samples. In these situations, contingency tables are constructed in terms of weighted frequencies that reflect the extent to which the sample contains elements with unequal probabilities of selection. As such, they cannot be assumed to follow either the multinomial distribution (2.1) or the product multinomial distribution (2.54). Moreover, whatever distribution does characterize these weighted frequencies may not permit straightforward maximum likelihood estimation. However, an alternative heuristic method for fitting log-linear models to such data is to use the same estimators for weighted contingency tables as would have been determined if the same identical table had arisen from a (possibly stratified) simple random sample. Although such estimators are not optimal, they may be satisfactory for most practical purposes because of their reasonable statistical properties (consistency and asymptotic normality) and their ease of computation. Thus, this section is concerned with the covariance structure of the log-linear model estimators which are the solutions to either the equations (2.9) or the equations (2.63) within the framework of complex probability random samples.

To be specific, let $i = 1,2,\ldots,s$ index a set of sub-populations of interest within which stratified random sampling has been undertaken. Let $\underset{\sim}{j} = (j_1,j_2,\ldots,j_d)$ index the response profiles as in Sections 2.1 and 2.2. Let $\ell = 1,2,\ldots,N_i$ index all of the elements in the i-th sub-population where N_i denotes the total number of such elements. Define element-wise response profile indicator random variables

$$N_{i,j_1j_2\cdots j_d,\ell} = \begin{cases} 1 & \text{if element } \ell \text{ from sub-population } i \text{ is classified as having response profile } \underset{\sim}{j} = (j_1,j_2,\ldots,j_d) \\ 0 & \text{otherwise} \end{cases} \tag{2.73}$$

which represent potential observations; and sample design indicator random variables

$$U_{i\ell} = \begin{cases} 1 & \text{if element } \ell \text{ from sub-population } i \text{ is in sample} \\ 0 & \text{otherwise} \end{cases} \tag{2.74}$$

which characterize the possibly multi-stage selection process including both the nature of any clustering as well as further stratification within the sub-populations according to other known partitions for either cost or other purposes. Thus, if $\phi_{i\ell} = E\{U_{i\ell}\}$ denotes the probability of selection for element ℓ in sub-population i, then the data pertaining to the multivariate relationships among the d attributes and their variation across the set of s sub-populations can be summarized in terms of the (d + 1)-th dimensional contingency table of weighted frequencies

$$\hat{N}_{i\underset{\sim}{j}} = \hat{N}_{i,j_1j_2\cdots j_d} = \sum_{\ell=1}^{N_i} \left(\frac{1}{\phi_{i\ell}}\right) U_{i\ell} N_{i,j_1j_2\cdots j_d,\ell} \; . \tag{2.75}$$

Although the $\{N_{i\underset{\sim}{j}}\}$ can be analyzed in their own right, attention here will be directed at their re-allocation to the respective within sub-population sample sizes $n_{1\cdot},n_{2\cdot},\ldots,n_{s\cdot}$ via the transformation

$$\tilde{n}_{i\underset{\sim}{j}} = \left(\frac{n_{i\cdot}}{N_i}\right) \hat{N}_{ij} \tag{2.76}$$

in order to maintain parallelism with the discussion in Section 2.2.

Given this framework, let $\underset{\sim}{\tilde{n}}$ denote the vector of weighted frequencies $\{n_{i\underset{\sim}{j}}\}$ in the same format as (2.51) and (2.53). In addition, let $\underset{\sim}{\pi}$ be the

corresponding vector of probabilities

$$\pi_{ij} = \pi_{i,j_1 j_2 \cdots j_d} = \frac{1}{N_i} E\{\hat{N}_{i,j_1 j_2 \cdots j_d}\} \qquad (2.77)$$

$$= \frac{1}{N_i} \sum_{\ell=1}^{N_i} E\{N_{i,j_1 j_2 \cdots j_d,\ell}\}$$

that reflect the average distribution of the various response profiles in the respective sub-populations. If the vector $\underset{\sim}{\pi}$ is assumed to be characterized by the log-linear model (2.60), then the pertinent vector of parameters $\underset{\sim}{\beta}$ can be estimated by solving the equations (2.63) with $\underset{\sim}{n}$ being replaced by $\underset{\sim}{\tilde{n}}$; i.e., the heuristic estimators $\underset{\sim}{\hat{\beta}}$ is obtained by solving the equations

$$\underset{\sim}{X}'[\underset{\sim}{m}(\underset{\sim}{\hat{\beta}})] = \underset{\sim}{X}'\underset{\sim}{\tilde{n}} \qquad (2.78)$$

Since the equations (2.78) implicitly define the estimators $\underset{\sim}{\hat{\beta}}$ as a function of the weighted frequencies $\underset{\sim}{\tilde{n}}$, its covariance matrix can be determined by applying the δ-method as outlined in (2.45) - (2.46). However, in this case, the required first derivative matrix must be obtained by applying implicit differentiation techniques. To be specific, if both sides of (2.78) are differentiated with respect to $\underset{\sim}{\tilde{n}}$ and the respective functions are evaluated at the point

$$\underset{\sim}{\mu} = \underset{\sim}{E}\{\underset{\sim}{\tilde{n}}\} = [\underset{\sim}{I}_r \otimes \underset{\sim}{n}_{\cdot}]\underset{\sim}{\pi} \quad , \qquad (2.79)$$

it follows that

$$\{\sum_{i=1}^{s} n_{i\cdot} \underset{\sim}{X}_i' [\underset{\sim}{D}_{\underset{\sim}{\pi}_i} - \underset{\sim}{\pi}_i \underset{\sim}{\pi}_i'] \underset{\sim}{X}_i\} \{\frac{d\underset{\sim}{\beta}}{d\underset{\sim}{y}'} \Big| \underset{\sim}{y}=\underset{\sim}{\mu}\} = \underset{\sim}{X}' \qquad (2.80)$$

Thus, the asymptotic (in the sense of large sample sizes $n = \sum_{i=1}^{s} n_{i\cdot}$ and extremely large sub-population sizes $\{N_i\}$) covariance matrix for $\underset{\sim}{\hat{\beta}}$

is given by

$$V_{\underset{\sim}{\beta}}(\underset{\sim}{\mu}) = \{\sum_{i=1}^{s} n_{i.}\underset{\sim}{X}_i'[D_{\underset{\sim}{\pi}_i} - \underset{\sim}{\pi}_i\,\underset{\sim}{\pi}_i']\underset{\sim}{X}_i\}^{-1}\,\underset{\sim}{X}'\,V_{\underset{\sim}{n}}\underset{\sim}{X}\{\sum_{i=1}^{s} n_{i.}\underset{\sim}{X}_i'[D_{\underset{\sim}{\pi}_i} - \underset{\sim}{\pi}_i\,\underset{\sim}{\pi}_i']\underset{\sim}{X}_i\}^{-1} \tag{2.81}$$

where $V_{\underset{\sim}{n}}$ is the sample survey design based covariance matrix of the linear sample statistics $\underset{\sim}{\tilde{n}}$. As a by-product of the discussion, it can be noted that (2.81) can be simplified to expression (2.64) for the special case of stratified simple random sampling.

Since the

$$\hat{\tilde{\pi}}_{\underset{\sim}{i}} = \{\exp(\underset{\sim}{X}_i\hat{\tilde{\beta}})\}/\{\underset{\sim}{1}_r'[\exp(\underset{\sim}{X}_i\hat{\tilde{\beta}})]\} \tag{2.82}$$

are satisfactory estimators of the $\underset{\sim}{\pi}_i$ (within the context of this discussion and the validity of the survey design for such purposes - i.e., absence of certain sources of non-sampling errors like interviewer variance, etc.), an appropriate estimator for the covariance matrix of $\hat{\tilde{\beta}}$ is

$$V_{\hat{\tilde{\beta}}} = V_{\hat{\tilde{\beta}}}(\tilde{n}) = \{\sum_{i=1}^{s} n_{i.}\underset{\sim}{X}_i'\,[D_{\hat{\underset{\sim}{\pi}}_i} - \hat{\underset{\sim}{\pi}}_i\,\hat{\underset{\sim}{\pi}}_i']\underset{\sim}{X}_i\}^{-1}\underset{\sim}{X}'\hat{V}_{\underset{\sim}{n}}\underset{\sim}{X}\{\sum_{i=1}^{s} n_{i.}\underset{\sim}{X}_i'[D_{\hat{\underset{\sim}{\pi}}_i} - \hat{\underset{\sim}{\pi}}_i\,\hat{\underset{\sim}{\pi}}_i']\underset{\sim}{X}_i\}^{-1} \tag{2.83}$$

where $\hat{V}_{\underset{\sim}{n}}$ is the sample survey design based estimated covariance matrix for $\underset{\sim}{\tilde{n}}$. With this framework in mind, additional analyses can be subsequently formulated in terms of $\hat{\tilde{\beta}}$ and $V_{\hat{\tilde{\beta}}}$ by applying FARM procedures as described in Sections 2.1 and 2.2.

Finally, it can be noted that Koch, Freeman, and Freeman [1975] and Brock, Freeman, Freeman, and Koch [1975] discuss the analysis of contingency tables based on complex sample survey data from a somewhat different point of view. In this regard, they consider the application of FAR methodology to certain general types of ratio estimators for situations where the pertinent covariance matrix is estimated by balanced repeated

replication procedures as described in Kish and Frankel [1970] and McCarthy [1969].

2.4 Sampling from Populations of Unknown Structure

In many situations where log-linear models are of interest, the specific nature of the corresponding $\underset{\sim}{X}$ matrix is not known a priori. When the sample sizes $n_{1\cdot}, n_{2\cdot}, \ldots, n_{s\cdot}$ from the respective sub-populations are very large in the sense that nearly all of the individual cell frequencies of the contingency table under study exceed 5 (in expectation), such structural ignorance causes no real difficulty because the interactive model building strategies described in reference to (2.28) - (2.33) can be applied de novo by formulating the first stage of analysis in terms of the constrained identity model

$$\underset{\sim}{X}_{CI} = \begin{bmatrix} \underset{\sim}{I}_{(r-1)} \\ -\underset{\sim}{1}'_{(r-1)} \end{bmatrix} \otimes \underset{\sim}{I}_{s} \qquad (2.84)$$

which in combination with (2.55) involves no reduction in dimensionality for the characterization of $\underset{\sim}{\pi}$ since both $\underset{\sim}{\pi}$ and $\underset{\sim}{\beta}$ are $s(r-1)$-dimensional. Thus, by definition, such models provide a perfect fit to the data with the corresponding implied vector of expected frequencies being identical to the observed frequencies; i.e.,

$$\hat{\underset{\sim}{m}}_{CI} = \underset{\sim}{m}(\hat{\underset{\sim}{\beta}}_{CI}) = \underset{\sim}{n} \; . \qquad (2.85)$$

Accordingly, various linear hypotheses of the type (2.28) can be tested via (2.29) to identify appropriate reduced models $\underset{\sim}{X}_{R}$ which can be fitted via (2.30) - (2.33) and can then be refined to an overall final model $\underset{\sim}{X}_{FM}$ by the appropriate continuation of the successive reduction process. Once this has been done, then the corresponding WLS estimators based on the standard GSK procedures can be used either in their own right or as the starting values for the determination of corresponding maximum

likelihood estimators by successive approximation numerical methods (e.g., those in Kaplan and Elston [1972]). However, as indicated previously (and also in Koch and Tolley [1975]), the choice between these two methods of estimation in the context of such very large sample sizes is largely a matter of personal tastes and computational convenience because their respective results are suitably similar for most practical purposes in the sense that

1. statistical tests of significance based on these two methods of estimation yield similar conclusions at the $\alpha=0.05$ level
2. for models which provide satisfactory fits, the specific estimators from the two procedures tend to agree within the limits of accuracy implied by their standard errors (i.e., the difference between them does not exceed either of their separate estimated standard errors)

On the other hand, for those situations where many of the cell frequencies for the contingency table under study are small (i.e., less than 5 in expectation) even though the overall sample size $n=\sum_{i=1}^{s} n_{i.}$ is large, the log-linear model $\underset{\sim}{X}_{CI}$ cannot be validly analyzed within the scope of the standard GSK methodology because neither the observed frequency vector $\underset{\sim}{n}$ nor its logarithmic transform $\underset{\sim}{f} = \log_e \underset{\sim}{n}$ (ignoring all the difficulties associated with 0-frequencies) can be assumed approximately to have multivariate normal distributions. Thus, an alternative approach is required. For this purpose, one potentially effective strategy is to analyze the data in terms of reasonable pseudo-models $\underset{\sim}{X}_{PM}$ which are based on the "most important" effects with all other effects like higher order interactions and/or gradient irregularities in the categorical scales for the respective attributes (e.g., the non-

linear effects associated with quantitative attributes) being excluded. Of course, the definition of "most important" is necessarily unclear because of the unknown structure of the population. However, either substantive considerations or variable screening procedures (e.g., as described by Clarke and Koch [1975] in terms of statistics based on marginal tables and/or the combination of sub-tables) can often be used to identify those variables which must be included simply because they are variables of practical interest. Moreover, other variables like certain lower order interaction effects may also be included provided that once a reasonable pseudo-model $\underset{\sim}{X}_{PM}$ has been formulated, the sample size n is sufficiently large to justify the assumption that the corresponding estimated parameter vector $\hat{\underset{\sim}{\beta}}_{PM}$ approximately has a multivariate normal distribution. Thus, the sample size places a definite limitation on the number of effects that can be put into $\underset{\sim}{X}_{PM}$ in the sense that the inclusion of too many effects restricts such analysis to descriptive purposes only by contradicting the multivariate normality assumption on which inferences regarding the corresponding estimators are based. In summary, $\underset{\sim}{X}_{PM}$ should include as many "most important" effects as possible within the scope of these sample size considerations.

Although such pseudo-models $\underset{\sim}{X}_{PM}$ do not in general characterize the vector of probability parameters $\underset{\sim}{\pi}$, they may nevertheless be fitted to the observed frequency vector $\underset{\sim}{n}$ by solving the equations (2.63) to estimate the corresponding parameters $\underset{\sim}{\beta}_{PM}$. However, the resulting estimators $\hat{\underset{\sim}{\beta}}_{PM}$, are not maximum likelihood estimators unless $\underset{\sim}{X}_{PM}$ fortuitously happens to be identical in structure to the true model $\underset{\sim}{X}_{TM}$. For this reason, the covariance matrix $\underset{\sim}{V}_{\underset{\sim}{\beta}_{PM}}(\underset{\sim}{\pi})$ for $\hat{\beta}_{PM}$ cannot be obtained by the Fisher Information Matrix approach (2.64), but instead must be

derived by a direct application of the δ-method similar to that outlined in (2.78) - (2.80). As a result, it follows that

$$\underset{\sim}{V}_{\underset{\sim}{\beta}_{PM}}(\underset{\sim}{\pi}) = [\underset{\sim}{H}\{\underset{\sim}{\pi}^*(\underset{\sim}{\beta}_{PM})\}]\ \underset{\sim}{V}_{\underset{\sim}{n}}[\underset{\sim}{H}\{\underset{\sim}{\pi}^*(\underset{\sim}{\beta}_{PM})\}]' \qquad (2.86)$$

where

$$\underset{\sim}{H}\{\underset{\sim}{\pi}^*(\underset{\sim}{\beta}_{PM})\} = \{\sum_{i=1}^{s} n_i \cdot \underset{\sim}{X}_i'[\underset{\sim}{D}_{\underset{\sim}{\pi}_i^*(\underset{\sim}{\beta}_{PM})} - \{\underset{\sim}{\pi}_i^*(\underset{\sim}{\beta}_{PM})\}\{\underset{\sim}{\pi}_i^*(\underset{\sim}{\beta}_{PM})\}']\underset{\sim}{X}_i\}^{-1}\underset{\sim}{X}', \qquad (2.87)$$

$\underset{\sim}{V}_{\underset{\sim}{n}}$ is the covariance matrix of $\underset{\sim}{n}$, and $\underset{\sim}{\pi}^*(\underset{\sim}{\beta}_{PM}) = \underset{\sim}{\pi}^*[\underset{\sim}{\beta}_{PM}(\underset{\sim}{\pi})]$ is the vector of pseudo probability parameters corresponding to the pseudo-model $\underset{\sim}{X}_{PM}$. Thus, if the true model $\underset{\sim}{X}_{TM}$ is contained within $\underset{\sim}{X}_{PM}$ (i.e., the columns of $\underset{\sim}{X}_{TM}$ are linear combinations of the columns of $\underset{\sim}{X}_{PM}$), then $\underset{\sim}{\pi}^*[\underset{\sim}{\beta}_{PM}(\underset{\sim}{\pi})] = \underset{\sim}{\pi}$ so that (2.87) becomes identical to (2.81) for general samples and to (2.64) for stratified simple random samples. Otherwise, as long as $\underset{\sim}{X}_{PM}$ includes the "most important" effects, $\underset{\sim}{\pi}^*[\underset{\sim}{\beta}_{PM}(\underset{\sim}{\pi})]$ will be sufficiently close to $\underset{\sim}{\pi}$ that their corresponding estimators $\hat{\underset{\sim}{\pi}}^*_{PM} = \underset{\sim}{\pi}^*[\hat{\underset{\sim}{\beta}}_{PM}]$ are also reasonable estimators of $\underset{\sim}{\pi}$ in moderate size samples. In this sense, $\hat{\underset{\sim}{\beta}}_{PM}$ can provide an effective basis for the further analysis of the data in such situations by FARM methodology with $\underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}_{PM}}(\underset{\sim}{\pi})$ being estimated by replacing $\underset{\sim}{\pi}^*[\underset{\sim}{\beta}_{PM}(\underset{\sim}{\pi})]$ with $\hat{\underset{\sim}{\pi}}^*_{PM}$ and replacing $\underset{\sim}{V}_{\underset{\sim}{n}}$ by a similarly appropriate estimator $\hat{\underset{\sim}{V}}_{\underset{\sim}{n}}$. These considerations as well as other statistical issues pertaining to pseudo-models are discussed in a somewhat different context pertaining to generalized measures of location or association in Koch, Tolley, and Freeman [1976].

2.5 Marginal Adjustment (Raking) of Contingency Tables

Whenever a sample is selected from a specific population, two types of information are obtained:

A. the marginal distributions of certain subsets of attributes WITHIN the respective sub-populations,

B. higher order measures of association which reflect the relationships among the attributes in the sense of interactions ACROSS the marginal subsets in (A).

In this context, Type (A) information will be called "allocation structure" while Type (B) information will be called "association structure." With these considerations in mind, the sample can be adjusted to provide estimators for other target populations of interest if the following assumptions hold:

1. the target population has KNOWN "allocation structure" via census or other sample survey data,
2. the target population has the SAME "association structure" as the sampled population.

Examples of such target populations include

a. various local (county or state) sub-divisions of a nationally sampled population for which local data do not provide sufficiently reliable estimators, if any at all;

b. other local, national, or international target populations which may or may not partially overlap a sampled local population.

More specifically, let $\hat{\underset{\sim}{\pi}}$ denote the vector of log-linear model based predicted probabilities for the respective response profiles $\underset{\sim}{j}$ within the respective sub-populations $i = 1,2,\ldots,s$. Let $\hat{\underset{\sim}{\pi}}_T$ denote the corresponding vector to be determined for the target population. Let $\underset{\sim}{A}_T$ denote a matrix of coefficients whose columns generate the pertinent marginal distributions comprising the known "allocation structure," and let $\underset{\sim}{\xi}_T$ denote their corresponding known values. Thus, assumption (1) means that $\hat{\underset{\sim}{\pi}}_T$ satisfies

$$\underset{\sim}{A}_T' \hat{\underset{\sim}{\pi}}_T = \underset{\sim}{\xi}_T \tag{2.88}$$

where, without loss of generality, $\underset{\sim}{A}_T$ will be regarded as having full rank

by deletion of unnecessary rows. Let $\underset{\sim}{X}$ denote the log-linear model on which $\hat{\underset{\sim}{\pi}}$ is based and let $\underset{\sim}{X}_C$ denote an ortho-complement matrix to $[1_{rr} \times I_s), \underset{\sim}{X}]$ which is also necessarily orthogonal to $\underset{\sim}{A}_T$ because of assumption (2) and the nature of the definitions (A) and (B) of "allocation structure" and "association structure." Finally, let $\underset{\sim}{K}$ denote an ortho-complement matrix to $[\underset{\sim}{A}_T, \underset{\sim}{X}]$. Then, assumption (2) via the formulation (2.66) - (2.67) implies that $\hat{\underset{\sim}{\pi}}_T$ satisfies

$$\begin{bmatrix} \underset{\sim}{K}' \\ \underset{\sim}{X}_C' \end{bmatrix} \{\log_e(\hat{\underset{\sim}{\pi}}_T)\} = \begin{bmatrix} \underset{\sim}{K}' \\ \underset{\sim}{X}_C' \end{bmatrix} \{\log_e(\hat{\underset{\sim}{\pi}})\} = \begin{bmatrix} (\underset{\sim}{K}'\underset{\sim}{X})\hat{\underset{\sim}{\beta}} \\ \underset{\sim}{0} \end{bmatrix} \qquad (2.89)$$

where $\hat{\underset{\sim}{\beta}}$ is the vector of estimated parameters which corresponds to the model $\underset{\sim}{X}$ and $\hat{\underset{\sim}{\pi}}$. Given this formulation, the estimators $\hat{\underset{\sim}{\pi}}_T$ may be determined (provided assumption (2) is true) by applying IPF to adjust the initial within sub-population conditional probability estimator $\hat{\underset{\sim}{\pi}}$ or its joint probability analogue $\hat{\underset{\sim}{\pi}}_J$ via (2.72) to conform successively with the marginal configurations corresponding to (2.88) since such operations preserve the "association structure" required by (2.89). Such marginal adjustment uses of IPF are sometimes referred to as "raking" procedures.

To obtain the covariance matrix of the raking estimators $\hat{\underset{\sim}{\pi}}_T$, the δ-method as outlined in (2.45) - (2.46) will be used with the required first derivative matrix being obtained by implicit techniques. In this regard, if both sides of the equations (2.88) - (2.89) are differentiated with respect to $\hat{\underset{\sim}{\beta}}$, it follows that

$$\begin{bmatrix} \underset{\sim}{A}_T' & \\ \underset{\sim}{K}' \, D^{-1}_{\hat{\underset{\sim}{\pi}}_T} & \\ \underset{\sim}{X}_C' \, D^{-1}_{\hat{\underset{\sim}{\pi}}_T} & \end{bmatrix} \begin{bmatrix} d\hat{\underset{\sim}{\pi}}_T \\ \hline d\hat{\underset{\sim}{\beta}} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{0} \\ \underset{\sim}{K}' \, \underset{\sim}{X} \\ \underset{\sim}{0} \end{bmatrix}. \qquad (2.90)$$

The equations (2.90) may be solved to yield

$$\begin{bmatrix} \frac{\partial \hat{\pi}_T}{\partial \hat{\beta}} \\ \sim \quad \sim \end{bmatrix} = \left[\underset{\sim}{D}_{\hat{\pi}_T} \underset{\sim}{A}_T [\underset{\sim}{A}_T' \underset{\sim}{D}_{\hat{\pi}_T}^{-1} \underset{\sim}{A}_T]^{-1}, \; [\underset{\sim}{K}, \underset{\sim}{X}_C] \left\{ \begin{bmatrix} \underset{\sim}{K}' \underset{\sim}{D}_{\hat{\pi}_T}^{-1} \underset{\sim}{K}, & \underset{\sim}{K}' \underset{\sim}{D}_{\hat{\pi}_T}^{-1} \underset{\sim}{X}_C \\ \underset{\sim}{X}_C' \underset{\sim}{D}_{\hat{\pi}_T}^{-1} \underset{\sim}{K}, & \underset{\sim}{X}_C' \underset{\sim}{D}_{\hat{\pi}_T}^{-1} \underset{\sim}{X}_C \end{bmatrix}^{-1} \right\} \begin{bmatrix} \underset{\sim}{0} \\ \underset{\sim}{K}' \underset{\sim}{X} \\ \underset{\sim}{0} \end{bmatrix} \right] \quad (2.91)$$

$$= [\underset{\sim}{K}, \; \underset{\sim}{X}_C] \begin{bmatrix} \underset{\sim}{K}' \underset{\sim}{D}_{\hat{\pi}_T}^{-1} \underset{\sim}{K} & \underset{\sim}{K}' \underset{\sim}{D}_{\hat{\pi}_T}^{-1} \underset{\sim}{X}_C \\ \underset{\sim}{X}_C' \underset{\sim}{D}_{\hat{\pi}_T}^{-1} \underset{\sim}{K} & \underset{\sim}{X}_C' \underset{\sim}{D}_{\hat{\pi}_T}^{-1} \underset{\sim}{X}_C \end{bmatrix}^{-1} \begin{bmatrix} \underset{\sim}{K}' \underset{\sim}{X} \\ \underset{\sim}{0} \end{bmatrix}$$

Thus, if $\underset{\sim}{\pi}_T$ denotes the target population probability parameter vector, then the covariance matrix for its estimator $\hat{\underset{\sim}{\pi}}_T$ is given by

$$\underset{\sim}{V}_{\hat{\pi}_T}(\underset{\sim}{\pi}_T; \underset{\sim}{\pi}) = [\underset{\sim}{K}, \underset{\sim}{X}_C] \begin{bmatrix} \underset{\sim}{K}' \underset{\sim}{D}_{\pi_T}^{-1} \underset{\sim}{K}, & \underset{\sim}{K}' \underset{\sim}{D}_{\pi_T}^{-1} \underset{\sim}{X}_C \\ \underset{\sim}{X}_C' \underset{\sim}{D}_{\pi_T}^{-1} \underset{\sim}{K}, & \underset{\sim}{X}_C' \underset{\sim}{D}_{\pi_T}^{-1} \underset{\sim}{X}_C \end{bmatrix}^{-1} \begin{bmatrix} \underset{\sim}{K}' \underset{\sim}{X} \\ \underset{\sim}{0} \end{bmatrix} [\underset{\sim}{V}_{\hat{\beta}}(\underset{\sim}{\pi})] [\underset{\sim}{X}' \underset{\sim}{K}, \; \underset{\sim}{0}] \begin{bmatrix} \underset{\sim}{K}' \underset{\sim}{D}_{\pi_T}^{-1} \underset{\sim}{K}, & \underset{\sim}{K}' \underset{\sim}{D}_{\pi_T}^{-1} \underset{\sim}{X}_C \\ \underset{\sim}{X}_C' \underset{\sim}{D}_{\pi_T}^{-1} \underset{\sim}{K}, & \underset{\sim}{X}_C' \underset{\sim}{D}_{\pi_T}^{-1} \underset{\sim}{X}_C \end{bmatrix}^{-1} \begin{bmatrix} \underset{\sim}{K}' \\ \underset{\sim}{X}_C' \end{bmatrix} \quad (2.92)$$

where $\underset{\sim}{V}_{\hat{\beta}}(\underset{\sim}{\pi})$ is the covariance matrix of the estimator $\hat{\underset{\sim}{\beta}}$ pertaining to the sampled population probability vector $\underset{\sim}{\pi}$ as given for general samples in (2.81) and for stratified simple random samples in (2.64). As in the previous sub-sections, a reasonable estimator for the covariance matrix $\underset{\sim}{V}_{\hat{\pi}_T}(\underset{\sim}{\pi}_T; \underset{\sim}{\pi})$ is obtained by replacing $\underset{\sim}{\pi}_T$ with $\hat{\underset{\sim}{\pi}}_T$ and $\underset{\sim}{V}_{\hat{\beta}}(\underset{\sim}{\pi})$ with the appropriate $\underset{\sim}{V}_{\hat{\beta}}$ in (2.92). Finally, this discussion has assumed that $\underset{\sim}{\xi}_T$ is a vector of known constants. However, these results may be extended to account for the case where $\underset{\sim}{\xi}_T$ is replaced by an estimator $\hat{\underset{\sim}{\xi}}_T$ by applying implicit differentiation to (2.88) - (2.89) with respect to $\hat{\underset{\sim}{\xi}}_T$, augmenting (2.91) accordingly, and substituting the overall first derivative matrix into (2.92) with $\underset{\sim}{V}_{\hat{\beta}}(\underset{\sim}{\pi})$ being replaced by the joint covariance matrix of $\hat{\underset{\sim}{\beta}}$ and $\hat{\underset{\sim}{\xi}}_T$. Otherwise, for an alternative formulation of the covariance matrix $\underset{\sim}{V}_{\hat{\pi}_T}(\underset{\sim}{\pi}_T; \underset{\sim}{\pi})$ from a somewhat different point of view, see Causey [1972].

Thus, on the basis of these considerations "raking" estimators can also be further analyzed by FARM procedures. In this regard, the following special case is of interest both for purposes of simplification as well as clarification of the nature of a potentially typical application.

a. the target population is identical to the sampled population which is the case when the "allocation structure" of the population under study is known a priori (e.g., samples of subjects from registration systems like licensed drivers)
b. simple random sampling is used
c. the log-linear model for $\hat{\underset{\sim}{\pi}}$ involves no reduction; i.e., $\underset{\sim}{X}$ is any matrix which is an ortho-complement to $\underset{\sim r}{1}$; and thus $\underset{\sim C}{X}$ is not defined.

Given the conditions (a) - (c), it can be verified that the covariance matrix in (2.92) becomes

$$V_{\underset{\sim T}{\hat{\pi}}}(\underset{\sim}{\pi}) = \frac{1}{n} \underset{\sim}{K}[\underset{\sim}{K}' \underset{\sim \pi}{D}^{-1} \underset{\sim}{K}]^{-1} \underset{\sim}{K}' \qquad (2.93)$$

3. EXAMPLES

In this section, four different types of examples involving log-linear models are presented. Section 3.1 is concerned with data from a single population, but pertaining to a repeated measurement experiment. Log-linear models are used to characterize the association structure among the responses to three drugs, prior to the comparison of their equivalence. Section 3.2 deals with a multiple population problem involving the relationship between driver injury in automobile accidents and selected variables characterizing the accident environment. Attention here is also directed at philosophical issues pertaining to the interpretation of model parameters. In Section 3.3, an incomplete contingency table involving paired comparison data is considered with emphasis being given to the problem of model formulation for such situations. Finally, Section 3.4 illustrates the reduction in variance which can be achieved by using "raking" procedures to estimate cell probabilities.

3.1 A Single Population Drug Comparison Example

The hypothetical data in Table 1 have been previously analysed in several references including BFH (pages 308-309), GSK, Koch and Reinfurt [1971], and Koch et al. [1974] to illustrate the construction of test statistics for the comparison of first order marginal probabilities in the setting of repeated measurement experiments. They are being reanalyzed here to demonstrate how such tests can be undertaken in terms of log-linear model based predicted values for cell probabilities which have been obtained via (2.50) by excluding certain "unimportant" higher order interaction.

The experimental design for this example involves $n = 46$ subjects from $s = 1$ population, each of whom is observed with respect to the occurrence of a favorable response for each of three treatments (e.g., having a positive effect on a bacteria culture for some specific type of throat infection). Thus, there are $d = 3$ attributes which represent the three drugs (Drug A, Drug B, and Drug C); and each of these has $L = L_g = 2$ response categories so that there are $r = L^d = 2^3 = 8$ possible multivariate response profiles.

Since an appropriate log-linear model for these data was not known a priori, the first stage of analysis is formulated in terms of the complete (or saturated) hierarchical model

$$\underset{\sim}{X}_H = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}, \tag{3.1}$$

whose respective columns are defined by the indicator functions (2.16) - (2.17)

1. TABULATION OF RESPONSES TO DRUGS A, B, AND C

	Response profile for Drug A vs Drug B vs Drug C							
Drug A	F	F	F	F	U	U	U	U
Drug B	F	F	U	U	F	F	U	U
Drug C	F	U	F	U	F	U	F	U
Overall group observed cell frequency	6	16	2	4	2	4	6	6
Observed proportions	0.130	0.348	0.043	0.087	0.043	0.087	0.130	0.130
Estimated s.e.	0.050	0.070	0.030	0.042	0.030	0.042	0.050	0.050

F denotes favorable response; U denotes unfavorable response.

2. TEST STATISTICS FOR LOG-LINEAR MODEL EFFECTS

Source of variation	D.F.	Complete model $X_{\sim H}$ GSK test statistic	Reduced model $X_{\sim R2}$ GSK test statistic	Reduced model $X_{\sim R2}$ FARM test statistic
Drug A main effect	1	0.47	0.71	0.79
Drug B main effect	1	0.47	0.71	0.79
A x B interaction	1	7.71**	9.08**	8.59**
Drug C main effect	1	2.72	3.74	4.12*
A x C interaction	1	0.47	(--)	(--)
B x C interaction	1	0.47	(--)	(--)
A x B x C interaction	1	0.08	(--)	(--)
Residual lack of fit	3	(--)	1.73	1.75†

* means significant at $\alpha = .05$;
** means significant at $\alpha = .01$;
(--) corresponds to sources of variation which are not defined.

† This is Log-likelihood Ratio Chi-Square Statistic from ECTA

and pertain to the corresponding effects shown in the rows of Table 2. Before proceeding further, it should be noted that several of the observed cell frequencies in Table 1 are very small (i.e., less than 5); and this consideration implies that the overall sample size n = 46 is not really large enough to ensure the statistical validity of all the results based on $\underset{\sim}{X}_H$. Nevertheless, such analysis is of interest as a "screening procedure" for identifying "unimportant" sources of variation for elimination from the model in terms of their corresponding $\underset{\sim}{Q}_C$-statistics. These quantities, which are given in the third column of Table 2, suggest that the A x B x C interaction is clearly "unimportant" and can be excluded from the model. Although the small sample size for these data does not permit a valid probability statement to be attached to this decision, it can be justified from a practical point of view by arguing that such test statistics increase linearly as a function of the sample size n; and thus, this source of variation would still have been "unimportant" in a relative sense, even if the sample size had been sufficiently large that its test statistic could have been regarded as approximately having a chi-square distribution under the corresponding null hypothesis; e.g., if the cell frequencies in Table 1 were all multiplied by 3, then the sample size would have been large enough to exclude the A x B x C interaction effect from the model on the basis of the non-significance (α = .25) of its test statistic (which in this case, would be 3 times as large as the corresponding result in Table 2). In summary, the primary objective of this first stage of analysis is to indicate that a reasonable reduced model for the data in Table 1 is

$$\underset{\sim}{X}_{R1} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & -1 & 1 & -1 \\ -1 & -1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 & 1 & 1 \end{bmatrix}, \qquad (3.2)$$

which is obtained from $\underset{\sim}{X}_H$ by excluding the A x B x C interaction effect. In view of the previous discussion, $\underset{\sim}{X}_{R1}$ will be regarded as a legitimate model as opposed to a pseudo-model in the sense of Section 2.4.

Since the log-linear model $\underset{\sim}{X}_{R1}$ has the hierarchical family structure (2.18), it can be fitted to the data in a maximum likelihood framework by using IPF in terms of the A vs B, A vs C, and B vs C two-way marginal configurations. More specifically, the University of Chicago computer program ECTA is used for this purpose, and the corresponding Log-Likelihood Ratio Chi Square Statistic for the goodness of fit of this model is $Q_L(\underset{\sim}{X}_{R1}) = 0.08$ with D.F. = 1, which is comparable to the GSK complete model result in Table 2, which is based on the linearized Neyman Chi-Square criterion. However, it should also be interpreted with caution because the sample size is not large enough for it to be regarded as approximately having a chi-square distribtuion because its asymptotic behavior is linked to the individual cell frequencies. On the other hand, the estimated parameters which are obtained on the basis of $\underset{\sim}{X}_R$ can reasonably be presumed as approximately having a multivariate normal distribution because their asymptotic behavior is linked to the cell frequencies for the pertinent marginal tables which are fitted by IPF and all of these can be noted to exceed 5.

Aside from these considerations, the model $\underset{\sim}{X}_{R1}$ is, for the most part, of only intermediate interest; and thus, estimated parameters are not shown. Alternatively, more specific attention is directed at the reduced model

$$\underset{\sim}{X}_{R2} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 \end{bmatrix}, \qquad (3.3)$$

which is obtained from $\underset{\sim}{X}_H$ by excluding the A x C and B x C interaction

effects as well as the A x B x C interaction. The use of this model can be initially motivated by appealing to the same types of arguments which led to $\underset{\sim}{X}_{R1}$. However, more rigorous justification can be based on the Log-Likelihood Ratio Chi-Square Statistic for the fit of $\underset{\sim}{X}_{R2}$ relative to $\underset{\sim}{X}_{R1}$. In this regard, $\underset{\sim}{X}_{R2}$ is fitted to the data in a maximum likelihood framework by using IPF in terms of the A vs B and C marginal configurations, for which the corresponding Log-Likelihood Ratio Chi-Square Statistic for goodness of fit is $Q_L(\underset{\sim}{X}_{R2}) = 1.75$ with D.F. = 3. Hence, the Log-Likelihood Ratio Chi-Square Statistic for the fit of $\underset{\sim}{X}_{R2}$ relative to $\underset{\sim}{X}_{R1}$ (i.e., the exclusion of the A x C and B x C interaction effects from $\underset{\sim}{X}_{R1}$) is

$$Q_L(\underset{\sim}{X}_{R2} | \underset{\sim}{X}_{R1}) = Q_L(\underset{\sim}{X}_{R2}) - Q_L(\underset{\sim}{X}_{R1}) = 1.75 - 0.08 = 1.67 \qquad (3.4)$$

with D.F. = 2, which is non-significant ($\alpha = .25$). Thus, the log-linear model $\underset{\sim}{X}_{R2}$ provides a satisfactory characterization for the distribution of the data in Table 1.

The IPF predicted frequency contingency table corresponding to the model $\underset{\sim}{X}_{R2}$ is given in Table 3. As indicated in Section 2.1, the maximum likelihood estimator $\hat{\underset{\sim}{\beta}}_{R2}$ for the parameters associated with $\underset{\sim}{X}_{R2}$ and its corresponding estimated covariance matrix $\underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}_{R2}}$ can be determined by applying the GSK weighted least squares computational procedures (2.36) - (2.44) to these IPF predicted frequencies. Thus, it follows that

$$\hat{\underset{\sim}{\beta}}_{R2} = \begin{bmatrix} 0.152 \\ 0.152 \\ 0.498 \\ -0.314 \end{bmatrix}, \quad \underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}_{R2}} = \begin{bmatrix} 2.8883 & & \text{Symmetric} & \\ -1.2784 & 2.8883 & & \\ -0.2367 & -0.2367 & 2.8883 & \\ 0.0000 & 0.0000 & 0.0002 & 2.3958 \end{bmatrix}. \qquad (3.5)$$

The estimators $\hat{\underset{\sim}{\beta}}_{R2}$ and their respective standard errors are also displayed in a more formal fashion in Table 4 under the heading "FARM Analysis." Corresponding test statistics for the respective components of $\hat{\underset{\sim}{\beta}}_{R2}$, which

3. IPF LOG-LINEAR MODEL PREDICTED CONTINGENCY TABLE

	Response profile for Drug A vs Drug B vs Drug C							
Drug A	F	F	F	F	U	U	U	U
Drug B	F	F	U	U	F	F	U	U
Drug C	F	U	F	U	F	U	F	U
Overall group IPF log-linear model predicted cell frequency	7.65	14.35	2.09	3.91	2.09	3.91	4.17	7.83

F denotes favorable response; U denotes unfavorable response.

4. ESTIMATED PARAMETERS FOR REDUCED LOG-LINEAR MODEL $X_{\sim R2}$

Parameter	GSK Analysis		FARM Analysis	
	Estimated parameter	Estimated s.e.	Estimated parameter	Estimated s.e.
Drug A main effect	0.144	0.171	0.152	0.170
Drug B main effect	0.144	0.171	0.152	0.170
A x B interaction	0.513	0.170	0.498	0.170
Drug C main effect	-0.305	0.158	-0.314	0.155

5. LOG-LINEAR MODEL PREDICTED PROPORTIONS BASED ON $X_{\sim R2}$

	Response profile for Drug A vs Drug B vs Drug C							
Drug A	F	F	F	F	U	U	U	U
Drug B	F	F	U	U	F	F	U	U
Drug C	F	U	F	U	F	U	F	U
MLE estimates	0.166	0.312	0.045	0.085	0.045	0.085	0.091	0.170
Estimated s.e.	0.042	0.059	0.020	0.034	0.020	0.034	0.029	0.046
GSK estimates	0.167	0.308	0.045	0.083	0.045	0.083	0.094	0.174
Estimated s.e.	0.046	0.055	0.020	0.033	0.020	0.033	0.027	0.051

F denotes favorable response; U denotes unfavorable response.

are based on the FARM statistic (2.29), are given in the last column of Table 2. These latter results imply that no further model reduction is appropriate from the hierarchical family point of view since the significance ($\alpha = .01$) of the A x B interaction implies that both this source of variation and the underlying main effects for Drug A and Drug B be maintained in the model. Otherwise, predicted values $\hat{\underset{\sim}{\pi}} = \underset{\sim}{\pi}(\hat{\underset{\sim}{\beta}})$ for the probabilities of the respective response profiles and their corresponding estimated covariance matrix $\underset{\sim}{V}_{\hat{\underset{\sim}{\pi}}}$ based on (2.48) are obtained by using the compounded function methods of Forthofer and Koch [1973]. For this purpose, $\hat{\underset{\sim}{\pi}}$ is computed according to the matrix formulation

$$\hat{\underset{\sim}{\pi}} = \underset{\sim}{A}_4\{\underset{\sim\sim\sim}{\exp}[\underset{\sim}{A}_3(\log\{\underset{\sim}{A}_2[\underset{\sim\sim\sim}{\exp}(\underset{\sim}{A}_1\hat{\underset{\sim}{\beta}})]\})]\} , \tag{3.6}$$

where

$$\underset{\sim}{A}_1 = \underset{\sim}{X}_{R2}, \quad \underset{\sim}{A}_2 = \begin{bmatrix} \underset{\sim}{I}_8 \\ \underset{\sim}{1}_8' \end{bmatrix}, \quad \underset{\sim}{A}_3 = [\underset{\sim}{I}_8, -\underset{\sim}{1}_8], \quad \underset{\sim}{A}_4 = \underset{\sim}{I}_8 \tag{3.7}$$

so that $\underset{\sim}{V}_{\hat{\underset{\sim}{\pi}}}$ can be determined as the matrix product

$$\underset{\sim}{V}_{\hat{\underset{\sim}{\pi}}} = \underset{\sim}{A}_4\underset{\sim}{D}_{\underset{\sim}{y}_3}\underset{\sim}{A}_3\underset{\sim}{D}^{-1}_{\underset{\sim}{a}_2}\underset{\sim}{A}_2\underset{\sim}{D}_{\underset{\sim}{y}_1}\underset{\sim}{A}_1\underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}}\underset{\sim}{A}_1'\underset{\sim}{D}_{\underset{\sim}{y}_1}\underset{\sim}{A}_2'\underset{\sim}{D}^{-1}_{\underset{\sim}{a}_2}\underset{\sim}{A}_3'\underset{\sim}{D}_{\underset{\sim}{y}_3}\underset{\sim}{A}_4 \tag{3.8}$$

where

$$\underset{\sim}{y}_1 = \underset{\sim\sim\sim}{\exp}(\underset{\sim}{A}_1\hat{\underset{\sim}{\beta}}), \quad \underset{\sim}{a}_2 = \underset{\sim}{A}_2\underset{\sim}{y}_1, \quad \underset{\sim}{y}_3 = \underset{\sim\sim\sim}{\exp}\{\underset{\sim}{A}_3[\underset{\sim\sim\sim}{\log}(\underset{\sim}{a}_2)]\}. \tag{3.9}$$

These computations can be performed by the computer program GENCAT which is documented in Landis, Stanish, and Koch [1975], and specific results for the predicted proportions $\hat{\underset{\sim}{\pi}}$ and their estimated standard errors are given in Table 5. Corresponding first order marginal probability estimators based on (2.49) - (2.50) with

$$\underset{\sim}{A} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \tag{3.10}$$

are given in Table 6 together with an appropriate $Q_{\underset{\sim}{C}}$-statistic. Thus, it follows that there are significant ($\alpha = .05$) differences among the three drugs, with the probability of a favorable response being less for Drug C than for Drug A and Drug B which are essentially the same.

For purposes of comparison, Table 2 and Tables 4-6 also contain analogous results based on the application of the standard one-stage GSK approach (2.36) - (2.44) to the actual observed data in Table 1. For the most part, these quantities are very similar to their maximum likelihood based counterparts. Thus, even though the sample size is not large enough to support the validity of the standard GSK approach for fitting the log-linear model $\underset{\sim}{X}_{R2}$ from a rigorous point of view, it still may be regarded as being of practical interest either in its own right or as a procedure for obtaining reasonable approximations for maximum likelihood results.

Finally, Table 6 contains estimates and test statistics based on applying the GSK procedure directly to the observed first order marginal proportions without the use of any log-linear model. This type of analysis is described in more detail in GSK, Koch and Reinfurt [1971], and Koch et al. [1974]. Its principal advantage is that no underlying log-linear model assumptions or preliminary testing are required for its use. Moreover, the sample size is large enough to support the validity of such statistics for the data in Table 1 since their asymptotic behavior is linked directly to the first order marginal configurations on which they are based. On the other hand, the test statistics associated with this direct approach may not be as powerful as those based on a preliminary log-linear model like $\underset{\sim}{X}_{R2}$. This type of conclusion is suggested by the results given in Table 6, but additional research is required before a definitive statement can be made with respect to this issue.

6. ESTIMATES AND TEST STATISTICS FOR COMPARING DRUGS

Drug	Marginal Model		Reduced Model X_{R2}			
	GSK Analysis		GSK Analysis		FARM Analysis	
	Estimated proportion favorable	Estimated s.e.	Estimated proportion favorable	Estimated s.e.	Estimated proportion favorable	Estimated s.e.
Drug A	0.609	0.072	0.604	0.073	0.609	0.072
Drug B	0.609	0.072	0.604	0.073	0.609	0.072
Drug C	0.348	0.070	0.352	0.072	0.348	0.070
Q_C statistic (D.F. = 2) for comparing drugs	6.58		8.54		7.83	

7. TABULATION OF DRIVER INJURY
BY WEATHER, TIME OF DAY, AND MODEL YEAR
FOR 1966, 1968-1972 NORTH CAROLINA, SINGLE VEHICLE ACCIDENTS
INVOLVING NON-DRINKING MALES
AND OCCURRING AT MEDIUM SPEED IN AN OPEN COUNTRY LOCATION

Sub-population			Observed frequencies for driver injury		IPF predicted frequencies for driver injury	
Weather	Time	Model year	Not severe	Severe	Not severe	Severe
Good	Day	-1966	5633	898	5621.65	909.35
Good	Day	1967-1969	2371	259	2371.18	258.82
Good	Day	1970-1973	1022	100	1022.05	99.95
Good	Night	-1966	7583	1526	7584.64	1524.36
Good	Night	1967-1969	3314	451	3315.38	449.62
Good	Night	1970-1973	1308	168	1316.09	159.90
Bad	Day	-1966	3915	428	3924.01	418.99
Bad	Day	1967-1969	2006	149	2010.16	144.84
Bad	Day	1970-1973	700	43	697.95	45.05
Bad	Night	-1966	3793	504	3793.70	503.30
Bad	Night	1967-1969	1924	166	1918.27	171.73
Bad	Night	1970-1973	718	51	711.90	57.10

3.2 A Multiple Population Investigation of Driver Injury in Automobile Accidents

This example is based on a research project undertaken at the University of North Carolina Highway Safety Research Center by Stewart [1975] for the purpose of studying the relationship between the severity of driver injury in automobile accidents and selected variables characterizing the accident environment with respect to crash configurations, location, time, and weather conditions, automobile type, and driver demographic status. In this regard, the data in Table 7 are from a specific, isolated modular component of that investigation which involved the accident (sub-)population with

$$\begin{aligned} \text{Crash Configuration} &= \text{Single Vehicle, Medium Speed} \\ \text{Location} &= \text{Open Country} \\ \text{Driver Demographic Status} &= \text{Non-Drinking (When Accident Occurred), Male} \\ \text{Calendar Year of Occurrence} &= 1966 \text{ or } 1968 - 1972 \end{aligned} \tag{3.11}$$

and its further partition into more refined sub-populations corresponding to the cross-classification of Weather (Good vs Bad), Time (Day vs Night), and Model Year (Before 1966 vs 1967-1969 vs 1970-1973). The attribute under study is whether or not the driver experienced "severe" injury where "severe" means either an "A"-injury (serious visible injury - a bleeding wound, distorted member, or any injury that requires the victim to be carried from the scene) or a "Fatal"-injury (an injury that results in death within 12 months of the accident). Given this framework, the questions of primary statistical interest pertain to the relationship between the conditional probability of "severe" injury and the "Weather," "Time," and "Model Year" characteristics of the accident. Thus, the analysis will be formulated in terms of the multiple population log-linear models in Section 2.2. For this purpose, the data in Table 7 are regarded as coming from a "stratified simple random

sample" from a hypothetical super-population of accidents (which might have been) within which driver injury severity is a stochastic outcome variable (which might have been different from what it actually was with respect to a series of conceptually repeated trials for the same accident environment) even though they actually have come from an observational (or convenience sample) of all North Carolina police reported accidents for the years 1966 and 1968-1972. The validity of this basic assumption does not really affect the specific nature of the analysis to be given subsequently, but rather its interpretation either in an inferential context with respect to the hypothetical super-population of accidents (if this point of view is realistic) or in a limited descriptive context restricted to the observational population under study. Other aspects of these philosophical considerations with respect to the super-population interpretation of observational data are discussed in Koch et al. [1975].

As was the case with the example in Section 3.1, an appropriate log-linear model for the data in Table 7 was not known a priori. For this reason, the first stage of analysis was formulated in terms of the complete (or saturated) hierarchical model

$$\underset{\sim}{X}_H = \begin{bmatrix} +0.5 \\ -0.5 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 \\ 1 & 1 & -1 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 1 & -1 & 1 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 1 & -1 & -1 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 \\ 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} +0.5 \\ -0.5 \end{bmatrix} \otimes \underset{\sim}{X}_{HL} \qquad (3.12)$$

whose respective columns pertain to the corresponding effects shown in the rows of Table 8. Tests of significance for these effects are undertaken by

fitting the model $\underset{\sim}{X}_H$ via the standard one-stage GSK approach (which is identical to the maximum likelihood approach since $\underset{\sim}{X}_H$ is complete) and using Q_C-statistics. However, instead of using the formulation outlined in (2.66) - (2.68), this analysis was directed at the logit functions

$$\underset{\sim}{F} = \underset{\sim}{F}(\underset{\sim}{p}) = \underset{\sim}{K}[\log_e(\underset{\sim}{A}\ \underset{\sim}{p})] \tag{3.13}$$

where

$$\underset{\sim}{A} = \underset{\sim}{I}_{24} \qquad \underset{\sim}{K} = [+1 \quad -1] \otimes \underset{\sim}{I}_{12}$$

because the model (3.12) for $\underset{\sim}{\pi}$ induces the asymptotic linear regression model

$$\underset{\sim}{F}(\underset{\sim}{\pi}) = \underset{\sim}{E}_{\underset{\sim}{A}}\{\underset{\sim}{F}(\underset{\sim}{p})\} = \underset{\sim}{X}_{HL}\underset{\sim}{\beta}_H \tag{3.14}$$

onto the observed logit functions $\underset{\sim}{F}$. Hence, the GSK estimated parameter vector $\underset{\sim}{b}$ for $\underset{\sim}{\beta}$ may be determined by fitting the model $\underset{\sim}{X}_{HL}$ to the logit function vector $\underset{\sim}{F}$ by weighted least squares. The Q_C-statistics for testing various hypotheses pertaining to $\underset{\sim}{\beta}$ within the framework of this model are given in the third column of Table 8. These results suggest that only the main effects for Weather, Time, and Model Year are significant ($\alpha = .01$). Thus, the analysis is directed at the reduced model

$$\underset{\sim}{X}_R = \begin{bmatrix} +0.5 \\ -0.5 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & 0 \\ 1 & 1 & -1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 & 0 \\ 1 & -1 & 1 & 0 & 1 \\ 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 0 \\ 1 & -1 & -1 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} +0.5 \\ -0.5 \end{bmatrix} \otimes \underset{\sim}{X}_{RL} \tag{3.15}$$

and its logit analogue

$$\underset{\sim}{F}(\underset{\sim}{\pi}) = \underset{\sim}{E}_{\underset{\sim}{A}}\{\underset{\sim}{F}(\underset{\sim}{p})\} = \underset{\sim}{X}_{RL}\underset{\sim}{\beta}_R \ . \tag{3.16}$$

8. TEST STATISTICS FOR LOG-LINEAR MODEL EFFECTS

Source of variation	D.F.	Complete model $X_{\sim H}$	Reduced model $X_{\sim R}$	
		GSK test statistic	GSK test statistic	FARM test statistic
Weather	1	79.74**	146.10**	146.67**
Time	1	16.25**	46.52**	46.53**
Weather x time	1	0.61	(--)	(--)
Model year	2	135.90**	157.79**	158.26**
Weather x model year	2	1.02	(--)	(--)
Time x model year	2	0.12	(--)	(--)
Weather x time x model year	2	0.13	(--)	(--)
Model year: -1966 vs 1967-1969	1	91.73**	107.91**	107.93**
Model year: -1966 vs 1970-1973	1	65.14**	74.12**	74.74**
Model year: 1967-1969 vs 1970-1973	1	3.63	2.84	2.93
Residual lack of fit	7	(--)	1.98	1.99†

* means significant at $\alpha = .05$
** means significant at $\alpha = .01$
(--) corresponds to sources of variation which are not defined

† This is Log-likelihood Ratio Chi-Square Statistic from ECTA.

The standard one-stage GSK estimators $\bar{b}_R$ for β_R and its corresponding estimated covariance matrix $\underset{\sim}{V}_{\bar{b}_R}$ can be determined by fitting the model X_{RL} directly to F by weighted least squares. Thus,

$$\bar{b} = \begin{bmatrix} 2.2190 \\ -0.2075 \\ 0.1086 \\ -0.2983 \\ 0.0949 \end{bmatrix}, \quad \underset{\sim}{V}_{\bar{b}} = \begin{bmatrix} 5.5506 & & & & \\ -1.1858 & 2.9461 & & \text{Symmetric} & \\ 0.3979 & 0.2279 & 2.5355 & & \\ -3.8139 & -0.0141 & -0.0049 & 6.1373 & \\ -1.4226 & 0.1466 & 0.0067 & 0.2028 & 8.6008 \end{bmatrix} \times 10^{-4}. \tag{3.17}$$

Corresponding Q_C-statistics for testing hypotheses pertaining to β_R are given in the fourth column of Table 8 and indicate that no further reduction is necessary (except for possibly re-structuring the model year effects to show no difference between the 1967-1969 vehicles and the 1970-1973 vehicles). Finally, predicted values $\bar{\pi}_S$ for the conditional probabilities of severe injury based on $\bar{b}$ can be formulated in the compounded function framework of Forthofer and Koch [1973] as

$$\bar{\pi}_S = \pi_S(\bar{b}) = A_4\{\exp[A_3(\log_e\{A_2[\exp(A_1\bar{\beta})]\})]\} \tag{3.18}$$

where

$$A_1 = X_R, \quad A_2 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \otimes I_{12}, \quad A_3 = [1 \ \ -1] \otimes I_{12}, \quad A_4 = I_{12} \tag{3.19}$$

with the corresponding estimated covariance matrix being obtained via matrix product operations analogous to (3.8). The estimators $\bar{\pi}_S$ obtained by this approach and their corresponding estimated standard errors are given in the sixth and seventh columns of Table 9. Similarly, the fourth and fifth columns of Table 9 contain the original observed proportions (which pertain to the model X_H) and their estimated standard errors. Thus, it can be noted that the predicted proportions $\bar{\pi}_S$ are very similar to the original observed proportions (as would be anticipated on the basis of the acceptable goodness of fit statistic in Table 8), but have substantially

9. OBSERVED AND LOG-LINEAR MODEL PREDICTED PROPORTIONS OF DRIVERS WITH SEVERE INJURY FOR NORTH CAROLINA DATA AND CORRESPONDING STANDARD ERRORS

Weather	Time	Model year	Observed proportion severe injury	Estimated s.e.	GSK log-linear predicted proportion severe injury	Estimated s.e.	IPF-MLE log-linear predicted proportion severe injury	Estimated s.e.
Good	Day	-1966	0.1375	0.0043	0.1392	0.0035	0.1392	0.0035
Good	Day	1967-1969	0.0985	0.0058	0.0984	0.0035	0.0984	0.0035
Good	Day	1970-1973	0.0891	0.0085	0.0892	0.0048	0.0891	0.0048
Good	Night	-1966	0.1675	0.0039	0.1673	0.0034	0.1674	0.0034
Good	Night	1967-1969	0.1198	0.0053	0.1194	0.0039	0.1194	0.0039
Good	Night	1970-1973	0.1138	0.0083	0.1085	0.0055	0.1083	0.0055
Bad	Day	-1966	0.0985	0.0045	0.0965	0.0031	0.0965	0.0031
Bad	Day	1967-1969	0.0691	0.0055	0.0672	0.0027	0.0672	0.0028
Bad	Day	1970-1973	0.0579	0.0086	0.0607	0.0036	0.0606	0.0036
Bad	Night	-1966	0.1173	0.0049	0.1172	0.0035	0.1171	0.0035
Bad	Night	1967-1969	0.0794	0.0059	0.0822	0.0032	0.0822	0.0032
Bad	Night	1970-1973	0.0663	0.0090	0.0744	0.0043	0.0742	0.0043

smaller estimated standard errors. This gain in statistical efficiency is one of the major advantages of the log-linear modeling process.

Alternatively, the model $\underset{\sim}{X}_R$ in (3.15) can be fitted by maximum likelihood methods. For this purpose, the observed data are initially transformed to IPF predicted frequencies by fitting the Weather vs Time vs Model Year, Weather vs Severity, Time vs Severity, and Model Year vs Severity marginal configurations. Then the maximum likelihood estimators $\hat{\underset{\sim}{\beta}}_R$ and its corresponding estimated covariance matrix $\underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}_R}$ are obtained by applying the same GSK procedures used to determine $\underset{\sim}{\bar{b}}$ and its companion results to the contingency table based on these predicted frequencies. Thus, it follows that

$$\hat{\underset{\sim}{\beta}}_R = \begin{bmatrix} 2.2197 \\ -0.2077 \\ 0.1086 \\ -0.2989 \\ 0.0945 \end{bmatrix}, \quad \underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}_R} = \begin{bmatrix} 5.5157 & & & & \\ -1.1506 & 2.9408 & & \text{Symmetric} & \\ 0.4095 & 0.2014 & 2.5322 & & \\ -3.7857 & -0.0679 & -0.0091 & 6.1319 & \\ -1.3982 & 0.1176 & 0.0254 & 0.1933 & 8.5911 \end{bmatrix} \times 10^{-4}. \tag{3.20}$$

Corresponding $\underset{\sim}{Q}_C$-statistics based on $\hat{\underset{\sim}{\beta}}_R$ are given in the last column of Table 8; predicted values for the conditional probabilities of severe injury based on $\hat{\underset{\sim}{\beta}}_R$ and their corresponding estimated standard errors are given in the last two columns of Table 9. Thus, it can be seen that the maximum likelihood approach and the standard one-stage GSK approach yield very similar results for this example. However, this conclusion could have been anticipated in view of the asymptotic equivalence of these two procedures and the very large sample involved with this example. As stated previously, when the sample size is sufficiently large to support the validity of the standard GSK approach, the choice between it and maximum likelihood is more a matter of personal taste and computational convenience than statistical efficiency.

A more critical issue is the interpretation of the log-linear model itself and its justification for specific problems. For these data, a log-linear model can be justified by regarding the relationship between Injury Severity and the accident environment variables in the context of a generalized "dose-response" relationship analogous to those used in quantal bioassay models. This point of view gains further support from the fact that no interaction is detected among the variables Weather, Time, and Model Year with respect to their effects on Injury Severity in this framework. Thus, the conceptual "dose" is an additive function of the pertinent main effect parameters for the respective sub-populations. Moreover, these parameters can be interpreted as measures of relative risk which are associated with the specific effects of one of the accident environment variables after controlling for the others. On the other hand, the major disadvantage of the use of log-linear models for the analysis of conditional distributions like those under study here is that researchers often prefer to interpret interaction directly with respect to the quantities being analyzed rather than some transformation of it. For this reason, the actual analysis given in Stewart [1975] uses straight linear models which are fitted by the GSK approach directly to the observed proportions of severe injury. As with the log-linear model, only the main effects for Weather, Time, and Model Year are important so that higher order interaction effects can be excluded. However, in this framework, such main effects can be interpreted directly as increments (or decrements) in the estimated probability of severe injury corresponding to the respective accident environment variables. Moreover, if interaction is present, it can be readily handled in this framework by partitioning the interacting variables into components which correspond to specific combinations of other non-interacting variables and then linking

them back together again. On the other hand, interaction with respect to the log-linear model tends to cloud the relative risk interpretation of the parameters and, as a consequence, also contradicts its "dose-response" justification. Thus, caution should be used in interpreting the results based on log-linear models in such situations.

In summary, both log-linear models and straight linear models may be regarded as appropriate analytical strategies for many situations involving the relationship between conditional probabilities and their corresponding control variables. The major advantage of the straight linear model is ease of interpretation for its parameters; the major disadvantage is its possibly poor or uncertain mathematical properties for small samples for which many of the cell frequencies are less than 5 (e.g., predicted values outside the 0-1 range can be obtained). Conversely, the major advantages of the log-linear model are its robust and stable mathematical properties while its major disadvantage is the possibly unclear interpretation of its parameters. Thus, for any specific application, the researcher should choose the model which is most appropriate with respect to these considerations as opposed to having an unconditional commitment to one or the other.

3.3 An Incomplete Contingency Table Involving Paired Comparison Data

The data in Table 10 represent the responses of 213 white North Carolina women (between 31 - 44 years of age, with less than 12 years of education, and married to their first husband) to the following question concerning "ideal family size":

> "Let's suppose, for a moment, that you have just been married and that you are given a choice of having, during your entire life-time, either $\underline{x}$ or $\underline{y}$ children. Which would you choose, $\underline{x}$ or $\underline{y}$?" (3.21)

Each woman was queried only with respect to one pair $(\underline{x},\underline{y})$; and all 42 possible

10. DESIRED FAMILY SIZE PAIRED CHOICES OF 213 WHITE NORTH CAROLINA WOMEN 31-44 YEARS OLD WITH LESS THAN 12 YEARS OF EDUCATION AND MARRIED TO THEIR FIRST HUSBAND

Pair		Choice							
x	y	0	1	2	3	4	5	6+	Total
0	1	0	7	--	--	--	--	--	7
0	2	1	--	6	--	--	--	--	7
0	3	1	--	--	6	--	--	--	7
0	4	1	--	--	--	12	--	--	13
0	5	1	--	--	--	--	3	--	4
0	6+	3	--	--	--	--	--	8	11
1	2	--	1	12	--	--	--	--	13
1	3	--	1	--	15	--	--	--	16
1	4	--	2	--	--	9	--	--	11
1	5	--	0	--	--	--	7	--	7
1	6+	--	5	--	--	--	--	6	11
2	3	--	--	4	9	--	--	--	13
2	4	--	--	7	--	0	--	--	7
2	5	--	--	11	--	--	5	--	16
2	6+	--	--	8	--	--	--	4	12
3	4	--	--	--	7	3	--	--	10
3	5	--	--	--	5	--	2	--	7
3	6+	--	--	--	12	--	--	2	14
4	5	--	--	--	--	9	2	--	11
4	6+	--	--	--	--	4	--	3	7
5	6+	--	--	--	--	--	6	3	9
Total		7	16	48	54	37	25	26	213

ordered pairs $(\underline{x},\underline{y})$ from the set $S = \{0,1,2,3,4,5,6+\}$ were randomly assigned to the women, although responses to the pair $(\underline{x},\underline{y})$ and the pair $(\underline{y},\underline{x})$ are pooled here since order of presentation is assumed to be unimportant. These data were gathered during the 1968 North Carolina Abortion Survey as described in Abernathy, Greenburg, and Horvitz [1970].

For purposes of simplicity, the data in Table 10 are assumed to constitute a simple random sample of the indicated (sub-)population of women throughout North Carolina. However, they are actually based on a complex probability sample. Thus, the results of the analyses presented in this section should be interpreted with some caution. On the other hand, such analysis can be justified by using the same types of super-population arguments outlined in Section 3.2. Moreover, the relatively refined partition of the women into sub-groups on the basis of demographic variables and assigned pair $(\underline{x},\underline{y})$ and the multiple regression flavor of the model fitting procedures which are to be used, both tend to minimize the complex sample survey design effect. For this reason, the results obtained here are regarded as reasonable approximations to those which might have been obtained in a complex sample survey framework (this not being done because an estimator for $\underset{\sim}{V}_{\underset{\sim}{n}}$ in (2.81) is not currently available).

The data array in Table 10 is called an incomplete contingency table because certain of its cells, by definition, correspond to impossible responses and hence are empty. Such incomplete contingency tables are discussed extensively in BFH (Chapter 5) with respect to log-linear models. In addition, they are discussed with respect to a broad class of multivariate paired comparison experiments in Imrey, Johnson, and Koch [1975]; and with respect to the "ideal family size" type of data presented in Table 10, in Koch, Abernathy, and Imrey [1975]. For the most part, log-linear models may be fitted to such incomplete tables in the same spirit as described for complete tables in Sections 2.1-2.2 because from an analytical point of view (as opposed to a conceptual one), such

tables are really not incomplete at all since the empty cells simply do not exist. Thus, corresponding to the incomplete contingency table in Table 10 is a transformed complete contingency table analogue in Table 11 for which the choice dimension is defined as "lower" vs. "upper." Although this complete contingency table is more difficult to interpret than its incomplete counterpart, it does nevertheless represent the data array to which the log-linear model is fitted and the context in terms of which the corresponding parameters must be defined. For this reason, the principal conceptual problem associated with the analysis of incomplete contingency tables is <u>model choice</u> rather than <u>model fitting</u>.

One useful framework for the analysis of paired comparison data is the Bradley-Terry [1952] model. If $\pi_{xy,x}$ represents the probability that a randomly selected subject chooses $\underline{x}$ when presented the pair $(\underline{x},y)$, then the Bradley-Terry model is formulated as

$$\pi_{xy,x} = \frac{\lambda_x}{\lambda_x + \lambda_y} \tag{3.22}$$

where $\lambda_0,\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5,\lambda_{6+}$ are "preference parameters" which provide an indication of the relative strengths of the respective choices for desired family size. However, the model (3.22) implies that the logit functions

$$F_{xy}(\pi_{xy,x}) = \log_e\{\pi_{xy,x}/\pi_{xy,y}\} = \log_e(\lambda_x) - \log_e(\lambda_y) \ . \tag{3.23}$$

Thus, if parameters β_y are defined by

$$\beta_y = \log_e(\lambda_0) - \log_e(\lambda_y) \tag{3.24}$$

for y = 1,2,3,4,5,6+, it follows that the observed logit functions for the complete contingency table data in Table 11

$$\underset{\sim}{F} = \underset{\sim}{F}(\underset{\sim}{p}) = \underset{\sim}{K}[\log_e(\underset{\sim}{A}\ \underset{\sim}{p})] \tag{3.25}$$

11. OBSERVED AND LOG-LINEAR MODEL PREDICTED COMPLETE CONTINGENCY TABLE ANALOGUE FOR PAIRED CHOICE DATA

Pair		Observed choice		IPF - MLE predicted choice	
x	y	Lower	Upper	Lower	Upper
0	1	0	7	2.16	4.84
0	2	1	6	0.44	6.56
0	3	1	6	0.29	6.71
0	4	1	12	1.33	11.67
0	5	1	3	0.55	3.45
0	6+	3	8	2.22	8.78
1	2	1	12	1.71	11.29
1	3	1	15	1.40	14.60
1	4	2	9	2.24	8.76
1	5	0	7	1.84	5.16
1	6+	5	6	3.98	7.02
2	3	4	9	5.05	7.95
2	4	7	0	4.40	2.60
2	5	11	5	11.23	4.77
2	6+	8	4	9.47	2.53
3	4	7	3	7.26	2.74
3	5	5	2	5.51	1.49
3	6+	12	2	11.97	2.03
4	5	9	2	6.40	4.60
4	6+	4	3	4.83	2.17
5	6+	6	3	5.53	3.47

12. ESTIMATED BRADLEY-TERRY PREFERENCE PARAMETERS FOR PAIRED CHOICE DATA

Choice	GSK Analysis		IPF-MLE Analysis	
	Estimated parameter	Estimated s.e.	Estimated parameter	Estimated s.e.
0	0.025	0.010	0.017	0.007
1	0.040	0.015	0.037	0.012
2	0.213	0.054	0.245	0.056
3	0.379	0.076	0.386	0.077
4	0.170	0.051	0.145	0.041
5	0.102	0.031	0.104	0.030
6+	0.071	0.020	0.065	0.019

where

$$\underset{\sim}{A} = \underset{\sim}{I}_{42} , \quad \underset{\sim}{K} = [1 \quad -1] \otimes \underset{\sim}{I}_{21} \tag{3.26}$$

are characterized by the model

$$\underset{\sim}{E}_{\underset{\sim}{A}}\{\underset{\sim}{F}\} = \underset{\sim}{F}(\underset{\sim}{\pi}) = \underset{\sim}{X}_{BTL}\underset{\sim}{\beta} \tag{3.27}$$

where

$$\underset{\sim}{X}_{BTL} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}, \quad \underset{\sim}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_{6+} \end{bmatrix} . \tag{3.28}$$

Given this framework, the one-stage standard GSK estimated parameter vector $\underset{\sim}{\bar{b}}$ for $\underset{\sim}{\beta}$ and its corresponding estimated covariance matrix $\underset{\sim}{V}_{\underset{\sim}{\bar{b}}}$ can be determined by fitting the model $\underset{\sim}{X}_{BTL}$ directly to $\underset{\sim}{F}$ (where 0-frequencies are replaced by (1/2) as described in Berkson [1955] in order to avoid "$\log_e(0)$" computations) by weighted least squares. This approach yields

$$\underset{\sim}{\bar{b}} = \begin{bmatrix} -0.48 \\ -2.16 \\ -2.73 \\ -1.93 \\ -1.42 \\ -1.06 \end{bmatrix}, \quad \underset{\sim}{V}_{\underset{\sim}{\bar{b}}} = \begin{bmatrix} 0.2861 & & & & & \\ 0.1528 & 0.2376 & & \text{Symmetric} & & \\ 0.1544 & 0.1636 & 0.2450 & & & \\ 0.1624 & 0.1459 & 0.1589 & 0.2555 & & \\ 0.1500 & 0.1704 & 0.1576 & 0.1551 & 0.2497 & \\ 0.1562 & 0.1478 & 0.1434 & 0.1425 & 0.1460 & 0.2001 \end{bmatrix} . \tag{3.29}$$

In addition, this model apparently provides a reasonable fit to the data since the goodness of fit statistic from (2.33), Q = 11.99 with D.F. = 15, is non-significant (α = .25). However, all of these results should be interpreted cautiously because of the relatively small sample size on which this example is based and the presence in Table 11 of many cell frequencies which are less than 5.

Alternatively, the parameters $\underset{\sim}{\beta}$ can be estimated by using a maximum likelihood approach. For this purpose, it is initially noted that the Bradley-Terry model (3.21) is equivalent to the model of quasi-independence of the rows and columns of the incomplete contingency table data in Table 10. However, as indicated in BFH, maximum likelihood predicted frequencies based on the quasi-independence model for the incomplete contingency table data in Table 10 can be obtained by using IPF in terms of its separate row and column marginal configurations with cells corresponding to impossible responses being constrained to 0 (by being assigned 0-initial values,. The resulting predicted frequencies are shown in the last two columns of Table 11 in the complete contingency table analogue format. The logit model (3.27) is then fitted to these predicted frequencies by the same weighted least squares procedures used for the GSK estimator $\underset{\sim}{\bar{b}}$ in order to compute the maximum likelihood estimator $\underset{\sim}{\hat{\beta}}$ and its corresponding estimated covariance matrix $\underset{\sim}{V}_{\underset{\sim}{\hat{\beta}}}$. Thus, it follows that

$$\underset{\sim}{\hat{\beta}} = \begin{bmatrix} -0.80 \\ -2.69 \\ -3.15 \\ -2.17 \\ -1.84 \\ -1.37 \end{bmatrix}, \quad \underset{\sim}{V}_{\underset{\sim}{\hat{\beta}}} = \begin{bmatrix} 0.2355 & & & & & \\ 0.1638 & 0.2595 & & \text{Symmetric} & & \\ 0.1648 & 0.1959 & 0.2801 & & & \\ 0.1556 & 0.1734 & 0.1773 & 0.2429 & & \\ 0.1617 & 0.1917 & 0.1827 & 0.1771 & 0.2601 & \\ 0.1532 & 0.1664 & 0.1671 & 0.1534 & 0.1653 & 0.2221 \end{bmatrix} \tag{3.30}$$

Maximum likelihood estimators $\underset{\sim}{\hat{\lambda}}$ (which are normalized to add to 1) for the preference parameters of the Bradley-Terry model can be formulated in the

compounded function framework of Forthofer and Koch [1973] as

$$\hat{\underset{\sim}{\lambda}} = \underset{\sim}{A}_4\{\exp[\underset{\sim}{A}_3(\log_e\{\underset{\sim}{A}_2[\exp(\underset{\sim}{A}_1\hat{\underset{\sim}{\beta}})]\})]\} \tag{3.31}$$

where

$$\underset{7x6}{\underset{\sim}{A}_4} = \begin{bmatrix} \underset{\sim}{1}_6' \\ (\underset{\sim}{1}_{66} - \underset{\sim}{I}_6) \end{bmatrix}, \quad \underset{8x7}{\underset{\sim}{A}_2} = \begin{bmatrix} \underset{\sim}{I}_7 \\ \underset{\sim}{1}_7' \end{bmatrix}, \quad \underset{7x8}{\underset{\sim}{A}_3} = [\underset{\sim}{I}_7, -\underset{\sim}{1}_7], \quad \underset{7x7}{\underset{\sim}{A}_4} = \underset{\sim}{I}_7 \tag{3.32}$$

with the corresponding estimated covariance matrix being determined via matrix operations analogous to (3.8). The estimators $\hat{\underset{\sim}{\lambda}}$ obtained by this approach and their corresponding estimated standard errors are given in the last two columns of Table 12. Analogous results based on the GSK estimator $\bar{\underset{\sim}{b}}$ are also given in Table 12 and are relatively similar to the maximum likelihood estimators (i.e., the differences between them tend to be less than the corresponding estimated standard errors). Thus, as in Section 3.1, the standard GSK approach provides reasonable results even though the sample size is not large enough to support its validity from a rigorous point of view.

For the sake of completeness, it should be noted that the Log-Likelihood Ratio Chi-Square Statistic (from the IPF computer program ECTA) for the goodness of fit of the Bradley-Terry model is $Q_{L,BT} = 24.31$ with D.F. = 15. However, this result should be interpreted with the same caution as its much smaller GSK counterpart because the asymptotic behavior of each of these statistics is linked to the individual cell frequencies, and thus the sample size is not large enough for either to be regarded as approximately having a chi-square distribution. Since neither the GSK approach nor the maximum likelihood approach provide a valid goodness of fit statistic for evaluating the suitability of the Bradley-Terry model for this example, alternative descriptive criteria will be used. In this regard, it

can be noted that the observed frequencies in Table 11 are relatively similar to the corresponding IPF predicted frequencies for the Bradley-Terry model with no difference between them exceeding 3 and two-thirds of such differences being less than 1. Alternatively, it can be noted that the $\underset{\sim}{Q}_C$-statistic for testing the hypothesis

$$H_0: \quad \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0 \tag{3.33}$$

of equal preference for all choices in terms of the maximum likelihood estimator $\hat{\underset{\sim}{\beta}}$ is $\underset{\sim}{Q}_C(\hat{\underset{\sim}{\beta}}) = 53.29$. Thus, it can be argued that the Bradley-Terry model is suitable for this example since its estimated parameters account for considerably more variation than its lack of fit. This same conclusion applies if the hypothesis in (3.33) is tested in terms of the GSK estimator $\overline{\underset{\sim}{b}}$ for which $\underset{\sim}{Q}_C(\overline{\underset{\sim}{b}}) = 46.59$.

Finally, all of the previous discussion is concerned with estimating the parameters $\hat{\underset{\sim}{\lambda}}$ by fitting logistic models rather than fitting log-linear models to the probability vector $\underset{\sim}{\pi}$ itself as described in Sections 2.1 - 2.2. This latter method of analysis is more difficult to formulate because it requires attention to be directed at the structure of the incomplete contingency table data array in Table 10 rather than its complete contingency table analogue in Table 11. For this reason, models with the hierarchical family structure (2.18) are not necessarily applicable since their definition is linked to the structure of complete contingency tables. Thus, other appropriate models must be formulated in their own right, and this can present substantial conceptual problems for certain types of incomplete contingency tables as discussed in BFH as well as computational difficulties. In the case of Bradley-Terry models for paired comparison data, an appropriate incomplete contingency table formulation can be based on the indicator functions

$$x_k(y) = \begin{cases} 1 & \text{if } y = k \text{ is chosen} \\ -1 & \text{if } y = 0 \text{ is chosen} \\ 0 & \text{otherwise} \end{cases} \qquad \text{where } y = 0,1,2,3,4,5,6+ \quad (3.34)$$

for k = 1,2,3,4,5,6+. The corresponding matrix formulation for this model is

$$\underset{\sim}{X}_{BT} = \begin{bmatrix}
-1 & -1 & -1 & -1 & -1 & -1 \\
1 & 0 & 0 & 0 & 0 & 0 \\
-1 & -1 & -1 & -1 & -1 & -1 \\
0 & 1 & 0 & 0 & 0 & 0 \\
-1 & -1 & -1 & -1 & -1 & -1 \\
0 & 0 & 1 & 0 & 0 & 0 \\
-1 & -1 & -1 & -1 & -1 & -1 \\
0 & 0 & 0 & 1 & 0 & 0 \\
-1 & -1 & -1 & -1 & -1 & -1 \\
0 & 0 & 0 & 0 & 1 & 0 \\
-1 & -1 & -1 & -1 & -1 & -1 \\
0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}, \qquad (3.35)$$

which can be fitted to either the observed frequencies of the IPF predicted frequencies by the methods outlined in (2.66) - (2.68). However, it should

be noted that the parameters $\underset{\sim}{\gamma}$ associated with this model are different from the parameters $\underset{\sim}{\beta}$ for the logit model $\underset{\sim BTL}{X}$, although it can be verified that they are related to each other by the equation

$$\underset{\sim}{\gamma} = \underset{\sim}{\beta} + (\underset{\sim 6}{1}' \underset{\sim}{\beta})\underset{\sim 6}{1} \tag{3.36}$$

by transforming $\underset{\sim BT}{X}$ to its logit analogue by operations similar to (3.13). In summary, the logit formulation of Bradley-Terry models provides the most straightforward framework for analysis; but the incomplete contingency table nature of the observed paired comparison data is of considerable interest because it provides the justification by which IPF can be used to obtain maximum likelihood estimators for the predicted frequencies.

3.4 A "Raked" Contingency Table

The data in Table 13 have been used by Ireland and Kullback [1968] to illustrate the application of IPF for the adjustment of a contingency table to a known marginal "allocation structure." They are being reanalyzed here to indicate the reduction in variance which is achieved by using such "raking" procedures to estimate the cell probabilities $\underset{\sim}{\pi}$.

These data originally come from a study undertaken by Roberts et al. [1939]. The experimental design involves n = 3734 mice from s = 1 population, each of which is classified with respect to the presence or absence of the attributes A, B, and D. The "allocation structure" of interest is defined in terms of the hypothesis that the probability of the presence (or absence) of each separate attribute is (1/2). Thus, with respect to the matrix notation in (2.88), it follows that

$$\underset{\sim T}{A}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}, \quad \underset{\sim T}{\xi} = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}. \tag{3.37}$$

13. TABULATION OF MICE ACCORDING TO ATTRIBUTES A, B, AND D

	Response profile for attributes A vs B vs D							
Attribute A	Y	Y	Y	Y	N	N	N	N
Attribute B	Y	Y	N	N	Y	Y	N	N
Attribute D	Y	N	Y	N	Y	N	Y	N
Overall group observed cell frequency	475	460	462	509	467	440	494	427
Observed proportions	0.1272	0.1232	0.1237	0.1363	0.1251	0.1178	0.1323	0.1144
Estimated s.e.	0.0055	0.0054	0.0054	0.0056	0.0054	0.0053	0.0055	0.0052

Y denotes presence of the attribute; N denotes absence.

14. "RAKED" PREDICTED CONTINGENCY TABLE FOR ATTRIBUTES A, B, AND D

	Response profile for attributes A vs B vs D							
Attribute A	Y	Y	Y	Y	N	N	N	N
Attribute B	Y	Y	N	N	Y	Y	N	N
Attribute D	Y	N	Y	N	Y	N	Y	N
Overall group "raked" predicted cell frequency	463.3	464.5	438.7	500.5	475.4	463.8	489.6	438.2
Predicted Proportions	0.1241	0.1244	0.1175	0.1340	0.1273	0.1242	0.1311	0.1174
Estimated s.e.	0.0041	0.0041	0.0041	0.0041	0.0041	0.0041	0.0041	0.0041

Y denotes presence of the attribute; N denotes absence.

The "association structure" which is to be preserved in the sense of (2.89) corresponds to the log-linear functions

$$\underset{\sim}{F}(\underset{\sim}{p}) = \underset{\sim}{K}'[\underset{\sim}{\log_e}(\underset{\sim}{p})] \tag{3.38}$$

where

$$\underset{\sim}{K}' = \begin{bmatrix} 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix} . \tag{3.39}$$

By using IPF to adjust the observed frequencies in Table 13 to the "allocation structure" specified by (3.37), Ireland and Kullback [1968] obtain the "raked" predicted cell frequencies shown in Table 14. The corresponding predicted proportions $\hat{\underset{\sim}{\pi}}_T$ are also given there together with their respective standard errors based on (2.93). Thus, by comparing these results with their counterparts in Table 13, it can be noted that the predicted proportions $\hat{\underset{\sim}{\pi}}_T$ are very similar to the original observed proportions, but have substantially smaller estimated standard errors.

Finally, since the "allocation structure" (3.37) corresponded to an hypothesis rather than a priori known constraints, Ireland and Kullback [1968] indicate that its acceptability for these data is supported by a non-significant ($\alpha = .25$) Minimum Discrimination Information Chi-Square Statistic for goodness of fit $Q_{MDI}(\hat{\underset{\sim}{\pi}}_T | \underset{\sim}{p}) = 3.42$ with D.F. = 3.

ACKNOWLEDGMENTS

The authors would like to thank Stephen E. Fienberg for several comments which were helpful to us during the initial stages of this research. Additional insights were gained as part of general discussions of statistical methodology during the Social Science Research Council, 1975 Summer Workshop on Crime and Criminal Justice Statistics. Finally, we would like to thank Alcinda Lewis and Jane Beth Markley for computational assistance, and Karen McKee and Rebecca Wesson for their cheerful and conscientious typing of the manuscript.

REFERENCES

Abernathy, James R., Greenberg, Bernard G., and Horvitz, Daniel G. [1970]. Estimates of induced abortion in urban North Carolina. Demography 7, 19-29.

Allen, D. M. and Grizzle, J. E. [1969]. Analysis of growth and dose response curves. Biometrics 25, 357-82.

Berkson, Joseph. [1955]. Maximum likelihood and minimum χ^2 estimates of the logistic function. Journal of the American Statistical Association 50, 130-62.

Birch, M. W. [1963]. Maximum likelihood in three-way contingency tables. Journal of the Royal Statistical Society 25B, 220-33.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. [1975]. Discrete Multivariate Analysis:: Theory and Practice, The MIT Press, Cambridge, Massachusetts.

Bradley, Ralph A. and Terry, Milton E. [1952]. Rank analysis of incomplete block designs: I. The method of paired comparisons. Biometrika 39, 324-45.

Brock, D. B., Freeman, D. H., Jr., Freeman, J. L., and Koch, G. G. [1975]. An application of categorical data analysis to the National Health Interview Survey. To appear in 1975 Proceedings of the Social Statistics Section of the ASA.

Causey, B. D. [1972]. Sensitivity of raked contingency table totals to changes in problem conditions. The Annals of Mathematical Statistics 43, 656-58.

Clarke, S. H. and Koch, G. G. [1974]. What determines whether persons arrested for burglary and larceny go to prison? Submitted to Journal of Criminal Law and Criminology.

ECTA. [1974]. University of Chicago Computer Program for fitting log-linear models to contingency tables.

Forthofer, R. N. and Koch, G. G. [1973]. An analysis for compounded functions of categorical data. Biometrics 29, 143-57.

Goodman, L. A. [1970]. The multivariate analysis of qualitative data: interactions among multiple classifications. Journal of the American Statistical Association 55, 226-56.

Grizzle, J. E., Starmer, C. F., and Koch, G. G. [1969]. Analysis of categorical data by linear models. Biometrics 25, 489-504.

Grizzle, J. E. and Williams, O. D. [1972]. Log linear models and tests of independence for contingency tables. Biometrics 28, 137-56.

Imrey, P. B., Johnson, W. D., and Koch, G. G. [1976]. An incomplete contingency table approach to paired comparison experiments, Institute of Statistics Mimeo Series, No. 1004. Submitted to the Journal of the American Statistical Association.

Ireland, C. T and Kullback, C. T. [1968]. Minimum discrimination information estimation. Biometrics 24, 707-13.

Kaplan, E. B. and Elston, R. C. [1973]. A subroutine package for maximum likelihood estimation (MAXLIK). North Carolina Institute of Statistics Mimeo Series, No. 823.

Kish, L. and Frankel, M. R. [1970]. Balanced repeated replications for standard errors. Journal of the American Statistical Association 65, 1071-94.

Koch, G. G., Abernathy, J. R., and Imrey, P. B. [1975]. On a method for studying family size preferences. Demography 12, 57-66.

Koch, G. G., Freeman, D. H., Jr., and Freeman, J. L. [1975]. Strategies in the multivariate analysis of data from complex surveys. International Statistical Review 43, 59-78.

Koch, G. G., Freeman, J. L., Freeman, D. H., Jr., and Lehnen, R. G. [1974]. A general methodology for the analysis of experiments with repeated measurement of categorical data. University of North Carolina Institute of Statistics Mimeo Series No. 961. Submitted to Biometrics.

Koch, G. G. and Greenberg, B. G. [1971]. The growth curve model approach to the statistical analysis of large data files. University of North Carolina Institute of Statistics Mimeo Series No. 786.

Koch, G. G. and Reinfurt, D. W. [1971]. The analysis of categorical data from mixed models. Biometrics 27, 157-73.

Koch, G. G. and Tolley, H. D. [1975]. A generalized modified χ^2 analysis of categorical data from a complex dilution experiment. Biometrics 31, 59-92.

Koch, G. G., Tolley, H. D. and Freeman, J. L. [1976]. An application of the clumped binomial model to the analysis of clustered attribute data. To appear in Biometrics.

Ku, H. H. and Kullback, S. [1974]. Log-linear models in contingency table analysis. The American Statistician 28, 115-22.

Landis, J. R. and Koch, G. G. [1976]. Compound function measures of observer agreement for multivariate categorical data. To appear in Biometrics.

Landis, J. R., Stanish, W. M., and Koch, G. G. [1975]. A computer program for the generalized chi-square analysis of categorical data using weighted least squares to generate Wald statistics. Documentation in the Department of Biostatistics, University of Michigan, Ann Arbor, Michigan. To be submitted to Computer Programs in Biomedicine.

McCarthy, P. J. [1969]. Pseudo-replication: half samples. Rev. Int. Statist. Inst. 37, 239-64.

Potthoff, R. F. and Roy, S. N. [1964]. A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika 51, 122-27.

Roberts, E., Dawson, W. M., and Madden, Margaret [1939]. Observed and theoretical ratios in Mendelian inheritance. Biometrika 31, 56-66.

Stewart, J. R. [1975]. An analysis of automobile accidents to determine which variables are most strongly associated with driver injury: relationships between driver injury and vehicle model year. University of North Carolina Highway Safety Research Center Technical Report.

Tolley, H. D. and Koch, G. G. [1974]. A two-stage approach to the analysis of longitudinal type categorical data. University of North Carolina Institute of Statistics Mimeo Series, No. 962.

Wald, A. [1943]. Tests of statistical hypotheses concerning general parameters when the number of observations is large. Transactions of the American Mathematical Society 54, 426-82.