



A General Methodology for the Analysis of Experiments with Repeated Measurement of Categorical Data

Gary G. Koch; J. Richard Landis; Jean L. Freeman; Daniel H. Freeman, Jr.; Robert G. Lehen

Biometrics, Vol. 33, No. 1 (Mar., 1977), 133-158.

Stable URL:

<http://links.jstor.org/sici?sici=0006-341X%28197703%2933%3A1%3C133%3AAGMFTA%3E2.0.CO%3B2-K>

Biometrics is currently published by International Biometric Society.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ibs.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

A General Methodology for the Analysis of Experiments with Repeated Measurement of Categorical Data

GARY G. KOCH

Department of Biostatistics, University of North Carolina,
Chapel Hill, North Carolina 27514 U.S.A.

J. RICHARD LANDIS

Department of Biostatistics, University of Michigan,
Ann Arbor, Michigan 48109 U.S.A.

JEAN L. FREEMAN and DANIEL H. FREEMAN, JR.

School of Medicine, Yale University, New Haven, Connecticut 06510 U.S.A.

ROBERT G. LEHNEN

Department of Political Science, University of Houston, Houston, Texas 77004 U.S.A.

Summary

This paper is concerned with the analysis of multivariate categorical data which are obtained from repeated measurement experiments. An expository discussion of pertinent hypotheses for such situations is given, and appropriate test statistics are developed through the application of weighted least squares methods. Special consideration is given to computational problems associated with the manipulation of large tables including the treatment of empty cells. Three applications of the methodology are provided.

1. Introduction

Koch and Reinfurt [1971] discussed the statistical analysis of multivariate categorical data from "mixed model" and "split plot" experiments. Such designs are classical prototypes of what are often called "repeated measurement designs," because they involve the collection of inherently multivariate data from the same subject for the same conceptual response variable under two or more experimental and/or observational conditions. The nature and range of application of these designs in the life and social sciences is illustrated by the following examples:

1. Bacteria samples are obtained from throats of each of n subjects with a particular infection. From each sample, three cultures are randomly assigned to three drugs. Each subject provides information about the effectiveness of the three drugs. For further discussion, see Section 3.1.
2. Three drugs are compared for relief of persistent symptoms of a chronic disease, e.g., asthma. Because of variation among patients, a two-period change-over design is used. This design involves the random assignment of paired sequences of drugs to patients who are then observed for two time periods with the first drug in the sequence being used in the first period and the other in the second period. For further discussion, see Section 3.2.

Key Words: Repeated measurement experiments; Multivariate categorical data; Weighted least squares.

3. A new drug and a standard drug are compared for patients with both mild and severe diagnoses of a disease, e.g., skin infection, mental depression, etc. Patients in each diagnosis group are randomly assigned to two drugs and their condition is evaluated as normal or abnormal at the end of one week, two weeks, and four weeks of continuous treatment. For further discussion, see Section 3.3. Analogous examples pertaining to behavioral development of children are given in Landis and Koch [1977c].
4. The agreement among two or more observers is evaluated with respect to the diagnosis of a disease like multiple sclerosis from patient record information, e.g., physical examinations, laboratory tests, family history, etc. Each subject is classified by each observer into a diagnostic category, e.g., certain multiple sclerosis; probable multiple sclerosis; possible multiple sclerosis; and doubtful, unlikely, or definitely not multiple sclerosis. For further discussion, see Landis [1975] and Landis and Koch [1975a, 1975b, 1977a, 1977b].
5. Multivariate paired comparison experiments are undertaken to obtain information on such phenomena as
 - a. the desired family size of women in an age and marital status group regarding their preferences for elements within pairs, e.g., two children vs. four children, one child vs. five children, no children vs. three children, etc.
 - b. the quality of soft drinks regarding preferences of subjects evaluating two pairs from four soft drinks according to two criteria—taste and fizziness. For further discussion, see Davidson and Bradley [1969], Koch, Abernathy and Imrey [1975], Fienberg and Larntz [1976], and Imrey, Johnson and Koch [1976].
6. A longitudinal panel survey is conducted to compare demographic groups regarding their political preferences for three presidential candidates at three time points prior to an election. Each respondent is interviewed at each time point. For further discussion, see Lehen and Koch [1974a] and Bishop, Fienberg and Holland [1975, Chapter 7].
7. A political survey is conducted to compare the level of trust demographic groups have for three political institutions—the Presidency, the Supreme Court and the Senate. Attitudes toward the three institutions are obtained from each respondent. For further discussion, see Lehen and Koch [1974b].
8. A political survey is conducted to compare preferences of demographic groups for seven tax expenditure policy alternatives. Each respondent is asked to rank the seven tax alternatives from “most desirable” to “least desirable.” For further discussion, see Koch, Freeman, and Lehen [1976].

For each of the preceding examples, subjects selected from s subpopulations are exposed to d different measurement conditions, e.g., drugs, time periods, stimuli like attitudes toward political candidates, governmental institutions, tax expenditure policies, etc., and are classified for each in terms of a response variable with L levels. More complex applications often involve incomplete research designs for which different subsets of the d conditions (or set of stimuli) are measured on different subjects. Consideration of this broader topic is given in Koch, Imrey and Reinfurt [1972], Lehen and Koch [1974a], Hocking and Oxspring [1974], and Chen and Fienberg [1974].

Questions of substantive interest for experiments (or surveys) with repeated measurements like (1)–(8) are as follows:

- i. Are there any differences among the s sub-populations regarding the average distribution of the responses to the d conditions?*

- ii. Are there any differences among the d conditions regarding the average distribution across the s sub-populations?
- iii. Are there any differences among the s sub-populations for differences among the distribution of responses to the d conditions?

These questions are analogous to hypotheses of "no whole-plot effects," "no split-plot effects," and "no whole-plot \times split-plot interaction" in split-plot experiments as described in Anderson and Bancroft [1952], Federer [1955] or Steel and Torrie [1960]. Here, however, such hypotheses must be formulated in terms of an underlying $(s \times r)$ contingency table where $r = L^d$ represents the number of possible multivariate response profiles. This is discussed in Section 2 where the set of univariate marginal distributions for the d measurement conditions within the s sub-populations contain information for dealing with the questions (i)–(iii). Test statistics for such hypotheses and the estimators of parameters for underlying linear regression models are obtained through weighted least squares computations by methods described in Grizzle, Starmer and Koch [1969] (hereafter GSK) as reviewed in Appendix 1.

The resulting methodology represents a categorical data analogue to more well-known counterparts for quantitative data like multivariate analysis of variance as described by Cole and Grizzle [1966] and Morrison [1967] in the parametric case, and multivariate rank analysis as described by Koch [1969, 1970] in the non-parametric case. Three examples which illustrate the scope of such analyses are discussed in Section 3.

2. Methodology

For repeated measurement experiments, each subject is observed under each of d different conditions, e.g., treatments, policy questions, time points, etc., and the corresponding responses are each classified in terms of L categories. Thus, there are $r = L^d$ possible multivariate response profiles. In accordance with the general framework in Appendix 1, these response profiles will be indexed by a vector subscript $\mathbf{j} = (j_1, j_2, \dots, j_d)$ where $j_g = 1, 2, \dots, L$ for $g = 1, 2, \dots, d$. As a result, $\pi_{i\mathbf{j}} = \pi_{i j_1, j_2, \dots, j_d}$ represents the joint probability of response profile \mathbf{j} for randomly selected subjects from the i th sub-population. If there are no differences among the sub-populations, then the $\{\pi_{i\mathbf{j}}\}$ satisfy the hypothesis,

$$H_{SJ} : \pi_{1\mathbf{j}} = \pi_{2\mathbf{j}} = \dots = \pi_{s\mathbf{j}} \quad \text{for all } \mathbf{j}; \quad (2.1)$$

similarly, if there are no differences among the d conditions, the $\{\pi_{i\mathbf{j}}\}$ satisfy the hypothesis of *total symmetry*

$$H_{CJ} : \pi_{i\mathbf{j}} = \pi_{i, \mathbf{z}(\mathbf{j})} \quad \text{for all } \mathbf{j} \quad \text{and} \quad i = 1, 2, \dots, s, \quad (2.2)$$

where $\mathbf{z}(\mathbf{j})$ is any permutation of \mathbf{j} . These hypotheses, with no interaction formulations (discussed in Bhapkar and Koch [1968a, 1968b]), can be tested by the general GSK methodology outlined in Appendix 1. However, these statistical tests are of limited interest since the quantities, which provide the most obvious indication of the relative effects associated with the respective measurement conditions and sub-populations, are the *first order marginal probabilities*

$$\phi_{i_g k} = \sum_{\mathbf{j} \text{ with } j_g = k} \dots \sum \pi_{i j_1, j_2, \dots, j_d} \quad \text{for} \quad \begin{matrix} i = 1, 2, \dots, s \\ g = 1, 2, \dots, d. \\ k = 1, 2, \dots, L \end{matrix} \quad (2.3)$$

Here, $\phi_{i_g k}$ represents the probability of the k th response category for the g th condition in

the i th sub-population. If there are no differences among the sub-populations (i.e., (2.1)), then the $\{\phi_{igk}\}$ satisfy the hypothesis,

$$H_{SM} : \phi_{1gk} = \phi_{2gk} = \cdots = \phi_{sgk} \quad \text{for} \quad \begin{matrix} g = 1, 2, \dots, d; \\ k = 1, 2, \dots, L; \end{matrix} \quad (2.4)$$

while if there are no differences among the conditions (i.e., 2.2) then the $\{\phi_{igk}\}$ satisfy the hypothesis of *first order marginal symmetry (homogeneity)*

$$H_{CM} : \phi_{i1k} = \phi_{i2k} = \cdots = \phi_{idk} \quad \text{for} \quad \begin{matrix} i = 1, 2, \dots, s; \\ k = 1, 2, \dots, L. \end{matrix} \quad (2.5)$$

Thus, it logically follows that

$$H_{SM} \text{ false} \rightarrow H_{SJ} \text{ false} \quad (2.6)$$

$$H_{CM} \text{ false} \rightarrow H_{CJ} \text{ false} \quad (2.7)$$

which means that H_{SJ} (or H_{CJ}) must be rejected whenever H_{SM} (or H_{CM}) is rejected. Alternatively, situations where H_{SM} (or H_{CM}) is true, regardless of whether H_{SJ} (or H_{CJ}) is true, would be interpreted as involving no gross (macro) differences among the sub-populations (conditions). Thus, the hypotheses H_{SM} and H_{CM} are of interest with respect to investigating the effects of sub-populations and conditions on the response distribution under study.

Moreover, H_{SM} (or H_{CM}) have greater logical relevance for this purpose than H_{SJ} (or H_{CJ}). This view can be justified by noting that the additional constraints on the π_{ij} , associated with H_{SJ} (or H_{CJ}) but not H_{SM} (or H_{CM}), involve relatively complicated equations which do not have a straightforward interpretation. Also, these constraints are partially confounded with those pertaining to the patterns of association, i.e., lack of independence, among the responses to the respective measurement conditions, since the hypothesis,

$$H_{JI} : \pi_{i_1, i_2, \dots, i_d} = \prod_{g=1}^d \phi_{igj_g} \quad \text{for all } j \text{ and } i = 1, 2, \dots, s, \quad (2.8)$$

of joint independence in conjunction with H_{SM} (or H_{CM}) necessarily imply H_{SJ} (or H_{CJ}). In repeated measurement experiments, such association is present to the extent that underlying subjects effects cause measurements under two different conditions on the same subject to be more similar than corresponding measurements on different subjects. This type of dependence among responses may be further complicated by factors pertaining to the adjacency of measurements in time and space, e.g., in some longitudinal studies, measurements from adjacent time points are often more related to one another than those for distant time points. Since such association can be present without having any bearing on the interpretation of sub-population and condition effects, it becomes appropriate to remove the corresponding constraints from H_{SJ} and H_{CJ} ; but this, of course, yields H_{SM} and H_{CM} . For most practical purposes, consideration can often be restricted to H_{SM} and H_{CM} in repeated measurement experiments (or surveys). Statistical tests directed at H_{SM} and H_{CM} tend to have better asymptotic convergence properties (regarding the valid application of Central Limit Theory (CLT) arguments) in moderately large samples than those directed at H_{SJ} and H_{CJ} since they involve fewer constraint equations, i.e., degrees of freedom (d.f.). This aspect of the analysis of repeated measurement experiments is further discussed in Appendix 2.

The hypothesis of no interaction between conditions and sub-populations can be formulated for the $\{\phi_{igk}\}$ in terms of the model

$$H_{AM} : \phi_{igk} = \mu_k + \xi_{i* k} + \tau_{*gk} \quad \text{for} \quad \begin{matrix} i = 1, 2, \dots, s \\ g = 1, 2, \dots, d, \\ k = 1, 2, \dots, L \end{matrix} \quad (2.9)$$

where μ_k is an overall mean associated with the k th response category, $\xi_{i* k}$ is an effect due to the i th sub-population, and τ_{*gk} is an effect due to the g th condition and where it is usually understood that the $\{\mu_k\}$, $\{\xi_{i* k}\}$ and $\{\tau_{*gk}\}$ satisfy the following types of constraints

$$\begin{aligned} \sum_{k=1}^L \mu_k &= 1, & \sum_{k=1}^L \xi_{i* k} &= 0, & \sum_{k=1}^L \tau_{*gk} &= 0, \\ \sum_{i=1}^s \xi_{i* k} &= 0, & \sum_{g=1}^d \tau_{*gk} &= 0, \end{aligned} \quad (2.10)$$

as a consequence of (A.2) and certain parameter identifiability considerations, i.e., removal of logical redundancies among parameters. If the model (2.9) is appropriate, then the hypothesis H_{SM} in (2.4) implies

$$H_{SM|AM} : \xi_{1* k} = \xi_{2* k} = \dots = \xi_{s* k} = 0 \quad \text{for} \quad k = 1, 2, \dots, L, \quad (2.11)$$

and the hypothesis H_{CM} in (2.5) implies

$$H_{CM|AM} : \tau_{*1k} = \tau_{*2k} = \dots = \tau_{*dk} = 0 \quad \text{for} \quad k = 1, 2, \dots, L. \quad (2.12)$$

Other hypotheses can be formulated to account for any inherent structure associated with the s sub-populations or the d conditions, e.g., trends or separate roles of two or more underlying factors. In some cases, it may be more appropriate to work with multiplicative no interaction models in which the additivity relationships like (2.9) pertain to the $\{\log \phi_{igk}\}$ (see Bhapkar and Koch [1968a, 1968b]). An example of this is given in Section 3.3.

These considerations can be extended further if the response categories $k = 1, 2, \dots, L$ are ordinally scaled with progressively larger intensities. In this event, the effects of the respective sub-populations and measurement conditions can be compared in one or more of the following cumulative marginal probability functions:

$$\theta_{igk} = \sum_{l=(k+1)}^L \phi_{igl} \quad \text{for} \quad \begin{matrix} i = 1, 2, \dots, s \\ g = 1, 2, \dots, d \end{matrix} \quad (2.13)$$

where $k = 1, 2, \dots, (L-1)$. Here θ_{igk} represents the probability that the response intensity is strictly greater than the k th category for the g th condition and the i th population. If the relative substantive importance of the θ_{igk} can be characterized by non-negative weights $W_1, W_2, \dots, W_{(L-1)}$ (at least one of which is non-zero), then attention can be directed at summary indexes:

$$\eta_{ig} = \sum_{k=1}^{(L-1)} W_k \theta_{igk} = \sum_{k=1}^L a_k \phi_{igk} \quad \text{for} \quad \begin{matrix} i = 1, 2, \dots, s \\ g = 1, 2, \dots, d \end{matrix} \quad (2.14)$$

where

$$a_1 = 0 \quad \text{and} \quad a_k = \sum_{l=1}^{(k-1)} W_l \quad \text{for} \quad k = 2, \dots, L.$$

Thus, η_{ig} can also be regarded as a *mean score* for the g th condition and the i th sub-population

with respect to an underlying numerical scaling a_1, a_2, \dots, a_L of the L response categories. In this context, the $\{\eta_{ig}\}$ are equivalent to *mean scores* derived from quantitatively scaled response variables (discussed by Bhapkar [1968]). If there are no differences among the sub-populations like H_{SM} in (2.4), then the $\{\eta_{ig}\}$ satisfy the hypothesis:

$$H_{SAM} : \eta_{1g} = \eta_{2g} = \dots = \eta_{sg} \quad \text{for } g = 1, 2, \dots, d; \quad (2.15)$$

while if there are no differences among the conditions like H_{CM} in (2.5), then the $\{\eta_{ig}\}$ satisfy the hypothesis

$$H_{CAM} : \eta_{i1} = \eta_{i2} = \dots = \eta_{id} \quad \text{for } i = 1, 2, \dots, s. \quad (2.16)$$

The logical relationship between H_{SAM} (or H_{CAM}) and H_{SM} (or H_{CM}) is directly analogous to the relationship previously discussed between H_{SM} (or H_{CM}) and H_{SJ} (or H_{CJ}); i.e., H_{SM} (or H_{CM}) must be rejected whenever H_{SAM} (or H_{CAM}) is rejected. Alternatively, situations where H_{SAM} (or H_{CAM}) is true, regardless of whether H_{SM} (or H_{CM}) is true, would be interpreted as involving no average gross (macro) differences among the sub-populations (conditions). Thus, the hypotheses H_{SAM} and H_{CAM} are of interest with respect to investigating the effects of sub-populations and conditions on the response distribution under study. Moreover, H_{SAM} (or H_{CAM}) have greater logical relevance for this purpose than H_{SM} (or H_{CM}), particularly if the scores a_1, a_2, \dots, a_L provide a meaningful and valid quantitative measure of the response's intensity. Further justification for this view can be derived by noting that the additional constraints on the $\{\phi_{igk}\}$, which are associated with H_{SM} (or H_{CM}) but not H_{SAM} (or H_{CAM}), involve complicated equations which are difficult to interpret because they pertain more to the "shape" of the response distribution than its location.

Other hypotheses can be formulated in terms of the $\{\eta_{ig}\}$. For example, if there is no interaction between conditions and sub-populations (i.e., (2.9)), the $\{\eta_{ig}\}$ are characterized by a similarly structured additive model with an overall mean, average effects for the sub-populations and for the conditions. Thus, the hypotheses pertaining to the goodness-of-fit of this model, the differences among sub-populations and among conditions are directly analogous to the hypotheses of "no whole-plot \times split-plot interaction," "no whole-plot main effects," and "no split-plot main effects" in standard split-plot (or repeated measurement) experiments which involve a univariate quantitative scaled response. From this view, all the models and hypotheses considered in this section represent straight-forward extensions of split-plot (repeated measurement) analysis of variance to multivariate categorical data.

Although it is preferable for the weights $W_1, W_2, \dots, W_{(L-1)}$ (or alternatively, the scores a_1, a_2, \dots, a_L) to have a strong substantive scaling basis, these same conclusions still apply if the weights are equal; i.e.,

$$W_1 = W_2 = \dots = W_{(L-1)} = 1, \quad (2.17)$$

so that

$$a_1 = 0, \quad a_2 = 1, \quad a_3 = 2, \dots, a_L = (L - 1) \quad (2.18)$$

represent an equally spaced scale. For most practical purposes, consideration can often be restricted to some appropriate formulation of H_{SAM} and H_{CAM} in repeated measurement experiments (or surveys) involving ordinally scaled response categories. This methodological strategy may be particularly effective for moderate sample sizes since statistical tests directed at H_{SAM} and H_{CAM} are based on fewer constraint equations, i.e., degrees of

freedom, and have better asymptotic convergence properties with respect to CLT than those directed at H_{SM} and H_{CM} . This aspect of the analysis of repeated measurement experiments is further discussed in Appendix 2.

Statistical tests for the hypotheses discussed in this section as well as the estimation of corresponding model parameters can be undertaken within the general framework in Appendix 1 by specifying the appropriate set of functions \mathbf{F} which represent estimators for the parameters under analysis and the corresponding operator matrices (e.g., \mathbf{A} 's, \mathbf{X} , and \mathbf{C} 's)

1. The \mathbf{A}_1 matrix specifies additive operations and can be used to generate linear functions of the observed compound proportion vector \mathbf{p} which are unbiased estimators for certain sets of *cell probabilities* like the $\{\pi_{ij}\}$, *marginal probabilities* like the $\{\phi_{igk}\}$ and/or *mean scores* like the $\{\eta_{io}\}$. The rows of \mathbf{A}_1 represent the respective functions while the columns are the corresponding coefficients of the elements of the compound vector \mathbf{p} which produce such functions.
2. The \mathbf{A}_2 matrix specifies multiplicative operations, i.e., additive on the \log_e scale, and in combination with the appropriate \mathbf{A}_1 matrix can be used to generate asymptotically unbiased estimators for functions involving the $\{\log_e \pi_{ij}\}$ or the $\{\log_e \phi_{igk}\}$.
3. The \mathbf{X} matrix has the same role here as it does in standard multiple regression; i.e., it indicates the manner in which the variation among a certain set of functions $\mathbf{F}(\boldsymbol{\pi})$ of cell probabilities can be characterized as linear functions of a smaller set of unknown parameters. Thus, the columns of \mathbf{X} correspond to the components of $\boldsymbol{\beta}$, and the rows of \mathbf{X} specify the appropriate linear combination of the components of the parameter vector $\boldsymbol{\beta}$ which apply to the respective elements of $\mathbf{F}(\boldsymbol{\pi})$.
4. The \mathbf{C} matrix also has the same role as in standard multiple regression; i.e., it indicates which linear functions of the parameters are equal to zero in accordance with a particular hypothesis which is to be tested.

More complete discussion of the construction of these operator matrices is given, with respect to the examples in Section 3, in Landis [1975] and in Landis and Koch [1977a].

3. Examples

Three examples are presented to illustrate the methodology outlined in Section 2. First, a simple experiment for the comparison of three drugs is considered in Section 3.1 to indicate the full range of hypotheses which can be tested and the relationships between the corresponding fitted models. Sections 3.2 and 3.3 are concerned with two typical, but relatively complex, medical experiments in which interaction is present among the effects of factors associated with the sub-populations and measurement conditions.

3.1. A Single Population Drug Comparison Example

The hypothetical data in Table 1 have been analyzed in GSK, Koch and Reinfurt [1971], and Bishop, Fienberg and Holland [1975, Chapter 8], to illustrate the construction of test statistics for the hypothesis H_{CM} of first order marginal symmetry (homogeneity) in (2.5). They are reanalyzed here to demonstrate the relationships between the test statistics for the hypotheses H_{CJ} and H_{CM} in Section 2 and the respective predicted values from corresponding fitted models.

A practical situation where this type of example arises is described in (1) of Section 1. In this case the experimental design involves $s = 1$ subpopulation, $d = 3$ conditions which represent three drugs (Drug A, Drug B, and Drug C), and $L = 2$ response categories

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

Sub-population	Response profile for Drug A vs Drug B vs Drug C							
	F	F	F	F	U	U	U	U
	F	F	U	U	F	F	U	U
	F	U	F	U	F	U	F	U
Overall group	6	16	2	4	2	4	6	6
Observed proportions	0.13	0.35	0.04	0.09	0.04	0.09	0.13	0.13
Estimated s.e.	0.05	0.07	0.03	0.04	0.03	0.04	0.05	0.05
Joint model pred. prop.	0.13	0.35	0.04	0.10	0.04	0.10	0.10	0.13
Estimated s.e.	0.05	0.07	0.02	0.02	0.02	0.02	0.02	0.05
Marg. model pred. prop.	0.13	0.35	0.04	0.09	0.04	0.04	0.13	0.13
Estimated s.e.	0.05	0.07	0.03	0.03	0.03	0.03	0.05	0.05

F - Favorable, U - Unfavorable.

(favorable F and unfavorable U). Thus, there are $r = L^d = 2^3 = 8$ possible multivariate response profiles. If $\pi_{i_1 i_2 i_3}$ represents the probability of response profile (j_1, j_2, j_3) for the three drugs, then the hypothesis H_{CJ} of *total symmetry* in (2.2) here corresponds to

$$H_{CJ} : \begin{matrix} \pi_{FFU} = \pi_{FUF} = \pi_{UFF} \\ \pi_{FUU} = \pi_{UFU} = \pi_{UUF} \end{matrix} \tag{3.1}$$

The functions required to test this hypothesis can be generated (with proper account given to (A.2)) via

$$\mathbf{F} = \mathbf{A}_1 \mathbf{p} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.13 \\ 0.35 \\ 0.04 \\ 0.09 \\ 0.04 \\ 0.09 \\ 0.13 \\ 0.13 \end{bmatrix} = \begin{bmatrix} 0.13 \\ 0.35 \\ 0.04 \\ 0.09 \\ 0.04 \\ 0.09 \\ 0.13 \\ 0.13 \end{bmatrix} \tag{3.2}$$

If $\mathbf{X} = \mathbf{I}_7$ is used, then a test statistic for H_{CJ} in (3.1) is obtained via (A.11) with

$$\mathbf{C}_{CJ} = \begin{matrix} 4 \times 7 \\ \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \end{matrix} \tag{3.3}$$

The resulting $Q_c = 16.29$ with d.f. = 4 which implies that there are significant ($\alpha = .01$) differences among drug effects. The nature of these differences can be attributed to the large

magnitude of p_{FFU} , the observed proportion with favorable responses to Drug A and Drug B but an unfavorable response to Drug C . This conclusion can be justified by fitting the model associated with

$$\mathbf{X}_J = \begin{matrix} 7 \times 4 \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix} . \quad (3.4)$$

From (A.8), the goodness-of-fit statistic for (3.4) is $Q = 0.51$ with d.f. = 3 which is non-significant ($\alpha = .25$). Thus, this model provides a satisfactory characterization of the distribution of the joint probabilities $\{\pi_{i_1 i_2 i_3}\}$. Predicted values for these quantities, which are obtained for the model (3.4) via (A.12), are given in Table 1 with corresponding estimated standard errors based on (A.13) in the "joint model" row.

The hypothesis H_{CM} of *first order marginal symmetry (homogeneity)* in (2.5) is formulated for this example as

$$H_{CM} : \phi_{AF} = \phi_{BF} = \phi_{CF} \quad (3.5)$$

where

$$\begin{aligned} \phi_{AF} &= \pi_{FFF} + \pi_{FFU} + \pi_{UFU} + \pi_{FUU} \\ \phi_{BF} &= \pi_{FFF} + \pi_{FFU} + \pi_{UFF} + \pi_{UFU} . \\ \phi_{CF} &= \pi_{FFF} + \pi_{FUF} + \pi_{UFF} + \pi_{UUF} \end{aligned} \quad (3.6)$$

The functions required to test H_{CM} can be generated (with proper account given to (A.2)) via

$$\mathbf{F} = \mathbf{A}_1 \mathbf{p} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.13 \\ 0.35 \\ 0.04 \\ 0.09 \\ 0.04 \\ 0.09 \\ 0.13 \\ 0.13 \end{bmatrix} = \begin{bmatrix} 0.61 \\ 0.61 \\ 0.34 \end{bmatrix} . \quad (3.7)$$

If $\mathbf{X} = \mathbf{I}_3$ is used, then a test statistic for H_{CM} in (3.5) is obtained via (A.11) with

$$C_{CM} = \begin{matrix} 2 \times 3 \\ \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \end{matrix} . \quad (3.8)$$

The resulting $Q_c = 6.58$ with d.f. = 2, which implies that there are significant ($\alpha = .05$) gross (macro) differences among the drug effects. In particular, the proportion with favorable responses for Drug *C* is less than those associated with Drug *A* and Drug *B* which are essentially the same. A model which reflects this conclusion is

$$\mathbf{X}_M = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (3.9)$$

for which the goodness-of-fit statistic from (A.8) is $Q = 0.00$ with d.f. = 1. However, this model only pertains to the *first order marginal probabilities* $\{\phi_{gk}\}$. For this reason, there is occasionally some interest in fitting a model directly to the joint probabilities $\{\pi_{i_1 i_2 i_3}\}$ which induces (3.9) on the $\{\phi_{gk}\}$. This type of analysis can be undertaken by using an augmented \mathbf{A}_1 -matrix

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad (3.10)$$

7×8

whose rows include the \mathbf{A}_1 -matrix in (3.7) on the one hand and constitute a basis for the same vector space as the rows of the \mathbf{A}_1 -matrix in (3.2). These functions are then analyzed in terms of an augmented \mathbf{X} matrix

$$\mathbf{X}_{MJ} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_M & \mathbf{0}_{3,4} \\ \mathbf{0}_{4,2} & \mathbf{I}_4 \end{bmatrix} \quad (3.11)$$

7×6

which applies (3.9) to the (3.7) functions and a non-restrictive identity matrix \mathbf{I}_4 to the others. The goodness-of-fit test statistic (A.8) for this model is $Q = 0.00$ with d.f. = 1 which is identical to that associated with (3.9) by construction. Predicted values for the joint probabilities, which are obtained for the model (3.11) via (A.12), are given in Table 1 with corresponding estimated standard errors based on (A.13) in the "marginal model" row.

3.2. A Complex Split Plot Contingency Table

The following hypothetical example arose from a two-period change-over design clinical trial as described in (2) of Section 1. Patients were sub-divided into two groups according

Table 2
TABULATION OF RESPONSES FOR TWO-PERIOD-CHANGE-OVER DESIGN

Age	Sequence	Response profile at time 1 vs time 2				Total
		FF	FU	UF	UU	
0	A:B	12	12	6	20	50
0	B:P	8	5	6	31	50
0	P:A	5	3	22	20	50
Y	B:A	19	3	25	3	50
Y	A:P	25	6	6	13	50
Y	P:B	13	5	21	11	50

F - Favorable, U - Unfavorable.

to age with 50 patients being assigned to each of three sequence sub-groups in each age group. The resulting data are shown in Table 2 where 0 denotes the older age group and Y denotes the younger age group.

This experiment involves $s = 6$ sub-populations ($2 \text{ ages} \times 3 \text{ sequence groups}$), $d = 2$ conditions (time 1 and time 2), and $L = 2$ response categories (favorable F and unfavorable U). There are $r = L^d = 2^2 = 4$ possible multivariate response profiles. The effects of age, drug and time on a patient's response are investigated by using the \mathbf{A}_1 -matrix

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \otimes \mathbf{I}_6 \tag{3.12}$$

to generate estimates for the *first order marginal probabilities* of patients with a favorable response and its corresponding estimated covariance matrix shown in Table 3. In this regard, the estimated covariances for the Period 1 and Period 2 estimators are of particular importance because their presence in \mathbf{V}_F from (A.5) represents the basic difference between the analysis of this experiment from a repeated measurement view as opposed to a standard multi-way factorial classification view (as would be the case if subjects were observed during only one period so that Period 1 and Period 2 would represent distinct sub-populations rather than distinct measurement conditions for the same sub-population).

Table 3
ESTIMATES OF FIRST ORDER MARGINAL PROBABILITIES OF A FAVORABLE RESPONSE

Age	Sequence	Period 1		Period 2		Covariance estimate
		Prob. fav. estimate	Variance estimate	Prob. fav. estimate	Variance estimate	
0	A:B	0.48	0.0050	0.36	0.0046	0.0013
0	B:P	0.26	0.0038	0.28	0.0040	0.0017
0	P:A	0.16	0.0027	0.54	0.0050	0.0003
Y	B:A	0.44	0.0049	0.88	0.0021	0.0001
Y	A:P	0.62	0.0047	0.62	0.0047	0.0023
Y	P:B	0.36	0.0046	0.68	0.0044	0.0003

Let $\phi_{i_1 i_2 g}$ denote the probability of a favorable response for the i_1 th age group, i_2 th treatment and g th period. If there are no carry-over effects of the drug in Period 1 to the response in Period 2 as described in Grizzle [1965] or Koch [1972], then the following model is of interest

$$\phi_{i_1 i_2 g} = \mu_{i_1} + \xi_{i_1 i_2} + \tau_{i_1 g} \tag{3.13}$$

where μ_{i_1} is an overall mean for the i_1 th age group, $\xi_{i_1 i_2}$ is an effect due to i_2 th treatment in i_1 th age group, and $\tau_{i_1 g}$ is an effect due to g th period for i_1 th age group. This model can be fitted, via the regression model in (3.14), with the repeated measurement aspect of the experiment

$$\mathbf{E}\{\mathbf{F}\} = \mathbf{X}_1 \boldsymbol{\beta}_1 = \begin{bmatrix} 1 & 1 & 0 & -1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & 1 & 0 & 1 \\ & & & 1 & 0 & 1 & -1 \\ & & & 1 & 1 & 0 & 1 \\ & & & 1 & 1 & 0 & -1 \\ & & & 1 & -1 & -1 & 1 \\ & & & 1 & -1 & -1 & -1 \\ & & & 1 & 0 & 1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} \mu_0 \\ \xi_{0A} \\ \xi_{0B} \\ \tau_{02} \\ \mu_Y \\ \xi_{YA} \\ \xi_{YB} \\ \tau_{Y2} \end{bmatrix}$$

(3.14)

being taken into account in (A.6)–(A.13) through the structure of \mathbf{V}_F . Since the goodness-of-fit statistic $Q = 0.24$ with d.f. = 4 is non-significant ($\alpha = 0.25$), the model provides a satisfactory characterization of the $\{\phi_{i_1 i_2 g}\}$. Statistical tests for certain hypotheses are given in Table 4. These results suggest that the model can be simplified to reflect no age \times treatment interaction as indicated in (3.15).

Table 4
STATISTICAL TESTS FOR \mathbf{X}_1 MODEL

Hypothesis	D.F.	Q_C
$\xi_{YA} = \xi_{0A}, \xi_{YB} = \xi_{0B}$	2	0.16
$\tau_{02} = \tau_{Y2}$	1	6.08
$\mu_0 = \mu_Y$	1	40.34

$$\mathbf{E}\{\mathbf{F}\} = \mathbf{X}_2\boldsymbol{\beta}_2 = \begin{bmatrix} 1 & 0 & 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & -1 & 0 \\ 1 & 0 & -1 & -1 & 1 & 0 \\ 1 & 0 & -1 & -1 & -1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & -1 & 0 & 1 \\ 0 & 1 & -1 & -1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_0 \\ \mu_Y \\ \xi_A \\ \xi_B \\ \tau_{02} \\ \tau_{Y2} \end{bmatrix}. \quad (3.15)$$

For this model, the goodness-of-fit statistic is $Q = 0.41$ with d.f. = 6; the resulting estimated parameters and predicted values are shown in (3.16).

$$\begin{bmatrix} 1 & 0 & 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & -1 & 0 \\ 1 & 0 & -1 & -1 & 1 & 0 \\ 1 & 0 & -1 & -1 & -1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & -1 & 0 & 1 \\ 0 & 1 & -1 & -1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.34 \\ 0.60 \\ 0.16 \\ -0.04 \\ 0.05 \\ 0.13 \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.35 \\ 0.26 \\ 0.27 \\ 0.18 \\ 0.55 \\ 0.43 \\ 0.88 \\ 0.63 \\ 0.61 \\ 0.35 \\ 0.69 \end{bmatrix}. \quad (3.16)$$

Statistical tests for hypotheses pertaining to this model are given in Table 5.

Thus, this analysis suggests that Drug *A* is significantly different from Drug *B* and Drug *P* which are similar, period effects are significant for each group and interact with age (in the sense of being of greater magnitude for the younger age group), and the overall mean for the younger age group is significantly greater than that for the older age group.

3.3. A Longitudinal Growth Curve Model Contingency Table

The following hypothetical example arose from a longitudinal study to compare a new drug and a standard drug as described in (3) of Section 1. The resulting data are shown

Table 5
STATISTICAL TESTS FOR \mathbf{X}_2 MODEL

Hypothesis	D.F.	$Q_{\tilde{C}}$
$\xi_A = \xi_B = 0$	2	49.21
$\mu_0 = \mu_Y$	1	40.37
$\tau_{02} = 0$	1	4.81
$\tau_{Y2} = 0$	1	31.16
$\tau_{02} = \tau_{Y2}$	1	6.37
$2\xi_A + \xi_B = 0 \ (\xi_A = \xi_P)$	1	47.55
$\xi_A + 2\xi_B = 0 \ (\xi_B = \xi_P)$	1	3.68
$\xi_A = \xi_B$	1	21.17

in Table 6. This experiment involves $s = 4$ sub-populations (2 diagnoses \times 2 treatments), $d = 3$ conditions (week 1, week 2, week 4), and $L = 2$ response categories (normal N and abnormal A). Thus, there are $r = L^d = 2^3 = 8$ possible multivariate response profiles. The effects of diagnosis and treatment on a patient's condition over time is investigated by using the \mathbf{A}_1 -matrix

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \otimes \mathbf{I}_4 \tag{3.17}$$

to generate estimates for the *first order marginal probabilities* of normal (N) and abnormal (A) for each week vs. diagnosis vs. treatment combination and the \mathbf{A}_2 matrix $\mathbf{A}_2 = [1 \ - \ 1]$

Table 6
TABULATION OF RESPONSES FOR LONGITUDINAL STUDY

Diagnosis	Treatment	Response profile at week 1 vs week 2 vs week 4								Total
		NNN	NNA	NAN	NAA	ANN	ANA	AAN	AAA	
Mild	Standard	16	13	9	3	14	4	15	6	80
Mild	New drug	31	0	6	0	22	2	9	0	70
Severe	Standard	2	2	8	9	9	15	27	28	100
Severe	New drug	7	2	5	2	31	5	32	6	90

$\otimes \mathbf{I}_{12}$ to generate their respective log ratios (or logits) as shown in Table 7 together with the corresponding estimated covariance matrix as shown in (3.18).

$$\mathbf{V_F} = \left[\begin{array}{rrrrrrrrrr} 5.00 & 1.27 & -0.76 & & & & & & & \\ 1.27 & 5.16 & -0.51 & & & & & & & \\ -0.76 & -0.51 & 5.70 & & & & & & & \\ & & & 5.73 & 0.94 & 3.12 & & & & \\ & & & 0.94 & 8.48 & -1.87 & & & & \\ & & & 3.12 & -1.87 & 51.47 & & & & \\ & & & & & & 6.03 & -0.56 & 0.08 & \\ & & & & & & -0.56 & 4.96 & -0.38 & \\ & & & & & & 0.08 & -0.38 & 4.03 & \\ & & & & & & & & & 7.60 & 0.34 & -0.81 \\ & & & & & & & & & 0.34 & 4.44 & 0.18 \\ & & & & & & & & & -0.81 & 0.18 & 8.00 \end{array} \right] \times 10^{-2} \tag{3.18}$$

Let $\lambda_{i_1 i_2 g}$ denote the large sample expected value of the logit corresponding to the i_1 th diagnosis, i_2 th treatment and g th week. If time is assumed to represent a metric which reflects dosage of the drugs under study, then the linear logistic model with respect to log time represents a reasonable model by analogy to well-known results from existing methodology for quantal bio-assays as discussed by Berkson [1944, 1953, 1955] or Finney [1964]. More specifically, we consider the model (3.19)

$$\lambda_{i_1 i_2 g} = \mu_{i_1 i_2} + \gamma_{i_1 i_2} x_{i_1 i_2 g}$$

$i_1 = 1: \text{Mild}, \quad 2: \text{Severe}$
 $i_2 = 1: \text{Standard}, \quad 2: \text{New Drug}$
 $g = 1: \text{Week 1}, \quad 2: \text{Week 2}, \quad 3: \text{Week 4}$

(3.19)

where $\mu_{i_1 i_2}$ represents an intercept parameter in reference to week 1 which is associated with i_1 th diagnosis and i_2 th treatment, $\gamma_{i_1 i_2}$ represents a corresponding continuous slope effect over time, and $x_{i_1 i_2 g}$ is the log to the base 2 of week, i.e., $x_{i_1 i_2 g} = 0, 1, 2$. In matrix notation, this model can be fitted via the regression model (3.20).

Table 7

OBSERVED AND PREDICTED ESTIMATES FOR FIRST ORDER MARGINAL
PROBABILITIES FOR NORMAL RESPONSE AND CORRESPONDING LOGITS

Diagnosis	Treatment	Week	Observed	Est.	Observed	Est.	Predicted	Est.	Predicted	Est.
			est. prob. normal		est. s.e.		est. s.e.		est. prob. normal	
Mild	Standard	1	0.51	0.06	0.05	0.22	-0.07	0.13	0.48	0.03
Mild	Standard	2	0.59	0.06	0.35	0.23	0.42	0.11	0.60	0.03
Mild	Standard	4	0.68	0.05	0.73	0.24	0.92	0.16	0.71	0.03
Mild	New drug	1	0.53	0.06	0.11	0.24	-0.07	0.13	0.48	0.03
Mild	New drug	2	0.79	0.05	1.30	0.29	1.38	0.15	0.80	0.02
Mild	New drug	4	0.97	0.02	3.53	0.71	2.84	0.25	0.94	0.01
Severe	Standard	1	0.21	0.04	-1.32	0.25	-1.35	0.13	0.21	0.02
Severe	Standard	2	0.28	0.04	-0.94	0.22	-0.86	0.10	0.30	0.02
Severe	Standard	4	0.46	0.05	-0.16	0.20	-0.36	0.15	0.41	0.04
Severe	New drug	1	0.18	0.04	-1.53	0.27	-1.35	0.13	0.21	0.02
Severe	New drug	2	0.50	0.05	0.00	0.21	0.10	0.12	0.53	0.03
Severe	New drug	4	0.83	0.04	1.61	0.27	1.56	0.21	0.82	0.03

$$\mathbf{E}_A\{\mathbf{f}\} = \mathbf{X}_1\boldsymbol{\beta}_1 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ & 1 & 0 \\ & 1 & 1 \\ & 1 & 2 \\ & & 1 & 0 \\ & & & 1 & 1 \\ & & & 1 & 2 \\ & & & & 1 & 0 \\ & & & & & 1 & 1 \\ & & & & & & 1 & 2 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \gamma_{11} \\ \mu_{12} \\ \gamma_{12} \\ \mu_{21} \\ \gamma_{21} \\ \mu_{22} \\ \gamma_{22} \end{bmatrix}. \quad (3.20)$$

for which the goodness-of-fit statistic is $Q = 1.60$ with d.f. = 4. The hypotheses and test statistics in Table 8 suggest differences exist among the respective diagnosis \times treatment patient groups with respect to the intercept and slope parameters; and that such differences among the intercepts can be explained in terms of a diagnosis effect while such differences among the slopes can be explained in terms of a treatment effect. On the basis of these results, the original model can be simplified to the following

$$\mathbf{E}_A\{\mathbf{f}\} = \mathbf{X}_2\boldsymbol{\beta}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} \mu_{1*} \\ \mu_{2*} \\ \gamma_{*1} \\ \gamma_{*2} \end{bmatrix} \quad (3.21)$$

where μ_{i*} is an intercept parameter associated with the i_1 th diagnosis and γ_{*i_2} is a slope effect associated with the i_2 th treatment. For this model, the goodness-of-fit statistic is $Q = 4.20$ with d.f. = 8. The corresponding estimated parameter vector \mathbf{b}_2 and its estimated covariance matrix \mathbf{V}_{b_2} are given in (3.22).

Table 8
STATISTICAL TESTS FOR \mathbf{X}_1 MODEL

Hypothesis	D.F.	Q_C
$\mu_{11} = \mu_{12} = \mu_{21} = \mu_{22}$	3	44.47
$\gamma_{11} = \gamma_{12} = \gamma_{21} = \gamma_{22}$	3	29.46
$\mu_{11} = \mu_{12}$	1	0.01
$\mu_{21} = \mu_{22}$	1	0.17
$\gamma_{11} = \gamma_{21}$	1	1.29
$\gamma_{12} = \gamma_{22}$	1	0.16

$$\mathbf{b}_2 = \begin{bmatrix} -0.07 \\ -1.35 \\ 0.49 \\ 1.46 \end{bmatrix}, \quad \mathbf{V}_{\mathbf{b}_2} = \begin{bmatrix} 1.82 & 0.75 & -0.74 & -0.64 \\ 0.75 & 1.82 & -0.83 & -1.04 \\ -0.74 & -0.83 & 0.93 & 0.53 \\ -0.64 & -1.04 & 0.53 & 1.69 \end{bmatrix} \times 10^{-2}. \quad (3.22)$$

From these results, predicted logits as shown in Table 8 can be determined via (A.12). These can then be used to obtain the predicted values for the *first order marginal probabilities* of normal (N) responses by reverse transformation. These quantities are also shown in Table 7. Estimated standard errors for these predicted values obtained through suitable manipulations of (A.13) are substantially smaller than those for the corresponding observed estimates, and thus reflect the extent to which the fitted model \mathbf{X}_2 enhances statistical efficiency. Finally, the hypotheses and test statistics in Table 9 justify the conclusions that the effects of diagnosis are significant but do not interact with time. Drug effects are also significant but are modulated in terms of different linear logistic trends over time. In other words, for both mild and severe diagnoses, a patient's condition becomes graded (N) sooner with the new drug than the standard drug even though there is essentially no difference between the treatments at the end of one week.

Table 9
STATISTICAL TESTS FOR \mathbf{X}_2 MODEL

Hypothesis	D.F.	Q_C
$\mu_{1*} = \mu_{2*}$	1	77.02
$\gamma_{*1} = \gamma_{*2}$	1	59.12
$\gamma_{*1} = 0$	1	26.35
$\gamma_{*2} = 0$	1	125.08

Acknowledgments

This research was supported in part by National Institutes of Health, Institute of General Medical Sciences through Grant GM-70004-05, by the U. S. Bureau of the Census through Joint Statistical Agreement JSA 75-2, and by Burroughs-Wellcome Co. through Joint Research Agreement 75-1.

The authors would like to thank Ms. Karen McKee and Ms. Rebecca Wesson for their cheerful and conscientious typing of the manuscript. The authors also wish to thank Ms. Linda L. Blakley and Ms. Connie Massey for their competent typing of the tables and other manuscript revisions.

Une Methodologie Generale pour L'Analyse d'Experiences avec Mesures Repetees de Donnees en Categories

Résumé

Cet article s'attache à l'analyse de données multivariates en catégories, données obtenues à partir de mesures sur des expériences répétées. On fournit dans un exposé introductif, les hypothèses pertinentes pour de telles situations, et on développe les tests statistiques adaptés en appliquant la méthode des moindres carrés pondérés. On s'attarde tout particulièrement sur certains problèmes de calcul liés à la manipulation de grands tableaux incluant le traitement de cellules vides. On fournit trois applications de cette méthodologie.

References

- Anderson, R. L. and Bancroft, T. A. [1952]. *Statistical Theory in Research*. McGraw-Hill, New York, Chapter 23.
- Berkson, J. [1944]. Application of the logistic function to bioassay. *Journal of the American Statistical Association* 39, 357-365.
- Berkson, J. [1953]. A statistically precise and relatively simple method of estimating the bioassay with quantal response based on the logistic function. *Journal of the American Statistical Association* 48, 565-599.
- Berkson, J. [1955]. Maximum likelihood and minimum χ^2 estimates of the logistic function. *Journal of the American Statistical Association* 50, 130-162.
- Bhapkar, V. P. [1966]. A note on the equivalence of two test criteria for hypotheses in categorical data. *Journal of the American Statistical Association* 61, 228-235.
- Bhapkar, V. P. [1968]. On the analysis of contingency tables with a quantitative response. *Biometrics* 24, 329-338.
- Bhapkar, V. P. and Koch, G. G. [1968a]. Hypotheses of 'no interaction' in multidimensional contingency tables. *Technometrics* 10, 107-123.
- Bhapkar, V. P. and Koch, G. G. [1968b]. On the hypotheses of 'no interaction' in multidimensional contingency tables. *Biometrics* 24, 567-594.
- Bishop, Y. M., Fienberg, S. E. and Holland, P. W. [1975]. *Discrete Multivariate Analysis: Theory and Practice*. M.I.T. Press, Cambridge, Massachusetts.
- Bowker, A. H. [1948]. A test for symmetry in contingency tables. *Journal of the American Statistical Association* 43, 572-574.
- Chen, T. and Fienberg, S. E. [1974]. Two-dimensional contingency tables with both completely and partially cross-classified data. *Biometrics* 30, 629-642.
- Cochran, W. [1950]. The comparison of percentages in matched samples. *Biometrika* 37, 256-266.
- Cole, J. W. L. and Grizzle, J. E. [1966]. Application of multivariate analysis of variance to repeated measurements experiments. *Biometrics* 22, 810-827.
- Davidson, R. R. and Bradley, R. A. [1969]. Multivariate paired comparisons: the extension of a univariate model and associated estimation and test procedures. *Biometrika* 56, 81-95.

- Federer, W. T. [1955]. *Experimental Design: Theory and Application*. Macmillan, New York, Chapter 10.
- Fienberg, S. E. and Larntz, K. [1975]. Loglinear representation for paired and multiple comparisons models. *Biometrika* 63, 245–254.
- Finney, D. J. [1964]. *Statistical Method in Biological Assay*, 2nd ed., Griffin, Ltd., London.
- Forthofer, R. N. and Koch, G. G. [1973]. An analysis for compounded functions of categorical data. *Biometrics* 29, 143–157.
- Grizzle, J. E. [1965]. The two-period change-over design and its use in clinical trials. *Biometrics* 21, 467–480.
- Grizzle, J. E., Starmer, C. F. and Koch, G. G. [1969]. Analysis of categorical data by linear models. *Biometrics* 25, 489–504.
- Hocking, R. R. and Oxspring, H. H. [1974]. The analysis of partially categorized contingency data. *Biometrics* 30, 469–483.
- Imrey, P. B., Johnson, W. D. and Koch, G. G. [1975]. An incomplete contingency table approach to paired-comparison experiments. *Journal of the American Statistical Association* 71, 614–623.
- Koch, G. G. [1969]. Some aspects of the statistical analysis of ‘split plot’ experiments in completely randomized layouts. *Journal of the American Statistical Association* 64, 485–505.
- Koch, G. G. [1970]. The use of non-parametric methods in the statistical analysis of a complex split plot experiment. *Biometrics* 26, 105–128.
- Koch, G. G. [1972]. The use of non-parametric methods in the statistical analysis of the two-period change-over design. *Biometrics* 28, 577–584.
- Koch, G. G., Abernathy, J. A and Imrey, P. B. [1975]. On a method for studying family size preferences. *Demography* 12, 57–66.
- Koch, G. G., Freeman, J. L., Freeman, D. H., Jr. and Lehnén, 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.
- Koch, G. G., Freeman, J. L. and Lehnén, R. G. [1976]. A general methodology for the analysis of ranked policy preference data. *International Statistical Review* 44, 1–28.
- Koch, G. G., Imrey, P. B., Freeman, D. H., Jr. and Tolley, H. D. [1976]. The asymptotic covariance structure of estimated parameters from contingency table log-linear models. *Proceedings of the 9th International Biometric Conference*, 317–336.
- Koch, G. G., Imrey, P. B. and Reinfurt, D. W. [1972]. Linear model analysis of categorical data with incomplete response vectors. *Biometrics* 28, 663–692.
- Koch, G. G. and Reinfurt, D. W. [1971]. The analysis of categorical data from mixed models. *Biometrics* 27, 157–173.
- Koch, G. G., Johnson, W. D., and Tolley, H. D. [1972]. A linear models approach to the analysis of survival and extent of disease in multidimensional contingency tables. *Journal of the American Statistical Association* 67, 783–796.
- Koch, G. G. and Tolley, H. D. [1975]. A generalized modified χ^2 analysis of data from a complex serial dilution experiment. *Biometrics* 31, 59–92.
- Kullback, S. [1971]. Marginal homogeneity of multidimensional contingency tables. *Annals of Math Statistics* 42, 594–606.
- Landis, J. R. [1975]. A general methodology for the measurement of observer agreement when the data are categorical. Ph.D. Dissertation, University of North Carolina Institute of Statistics Mimeo Series, No. 1022.
- Landis, J. R. and Koch, G. G. [1975a]. A review of statistical methods in the analysis of data arising from observer reliability studies (Part I). *Statistica Neerlandica* 29, 101–123.
- Landis, J. R. and Koch, G. G. [1975b]. A review of statistical methods in the analysis of data arising from observer reliability studies (Part II). *Statistica Neerlandica* 29, 151–161.
- Landis, J. R. and Koch, G. G. [1977a]. The measurement of observer agreement for categorical data. *Biometrics* 33, 159–175.
- Landis, J. R. and Koch, G. G. [1977b]. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. University of Michigan Biostatistics Technical Report No. 13. To appear in *Biometrics*.
- Landis, J. R. and Koch, G. G. [1977c]. The analysis of categorical data in longitudinal studies of behavioral development. To appear in *Manual on Longitudinal Research Methodology*, a technical report, J. R. Nesselroade and P. B. Baltes, (eds.), for the National Institute of Education.

- Landis, J. R., Stanish, W. M., Freeman, J. L. and Koch, G. G. [1976]. A computer program for the generalized chi-square analysis of categorical data using weighted least squares (GENCAT). University of Michigan Biostatistics Technical Report No. 8. To appear in *Computer Programs in Biomedicine*.
- Lehnen, R. G. and Koch, G. G. [1974a]. Analyzing panel data with uncontrolled attrition. *The Public Opinion Quarterly* 38, 40–56.
- Lehnen, R. G. and Koch, G. G. [1974b]. The analysis of categorical data from repeated measurement research designs. *Political Methodology* 1, 103–123.
- Morrison, D. F. [1967]. *Multivariate Statistical Methods*. McGraw-Hill, New York, Chapters 4 and 5.
- Neyman, J. [1949]. Contribution to the theory of the χ^2 test. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley and Los Angeles: University of California Press, Berkeley and Los Angeles, (239–272).
- Steel, R. G. D. and Torrie, J. H. [1960]. *Principles and Procedures of Statistics*. McGraw-Hill, New York, Chapter 12.
- Wald, A. Tests of statistical hypotheses concerning general parameters when the number of observations is large. *Transactions of the American Mathematical Society* 54, 426–482.

Received July 1975, Revised May 1976

Appendix 1

General Framework

The purpose of this appendix is to restate the general framework for the methodology originally outlined in Grizzle, Starmer and Koch [1969] with emphasis directed at:

- i. additional clarification of the relationship between model fitting and hypothesis testing in terms of the relationship between goodness-of-fit statistics and contrast matrix test statistics;
- ii. additional clarification of the Central Limit Theory justification for weighted least square model fitting procedures;
- iii. additional aspects of analysis like predicted functions and their corresponding estimated covariance matrix;
- iv. coverage of a broader class of functions beyond linear and log-linear functions.

This appendix also expresses why marginal probability functions and/or mean scores can be effectively analyzed for moderate size repeated measurement experiments which involve large tables with sparse data within the scope of weighted least squares methods but not necessarily maximum likelihood methods (which require attention to be directed at the set of joint cell probabilities in the overall table).

Let $g = 1, 2, \dots, d$ index the d conditions under which measurements on the same basic response with L categories are observed. Let $j = 1, 2, \dots, r$ index the set of categories which correspond to the $r = L^d$ response profiles associated with the simultaneous classification for the d responses of interest. Similarly, let $i = 1, 2, \dots, s$ index a set of categories which correspond to distinct sub-populations as defined in terms of pertinent independent variables. If samples of size n_i where $i = 1, 2, \dots, s$ are independently selected from the respective sub-populations, then the resulting data can be summarized in an $(s \times r)$ contingency table as shown in Table 10, where n_{ij} denotes the frequency of response category j in the sample from the i th sub-population.

The vector \mathbf{n}_i where $\mathbf{n}_i' = (n_{i1}, n_{i2}, \dots, n_{ir})$ will be assumed to follow the multinomial distribution with parameters n_i and $\boldsymbol{\pi}_i' = (\pi_{i1}, \pi_{i2}, \dots, \pi_{ir})$, where π_{ij} represents the probability that a randomly selected element from the i th population is classified in the j th response category. Thus, the relevant product multinomial model is

$$\phi = \prod_{i=1}^s \left\{ n_i! \prod_{j=1}^r [\pi_{ij}^{n_{ij}} / n_{ij}!] \right\} \quad (\text{A.1})$$

with the constraints

$$\sum_{j=1}^r \pi_{ij} = 1 \quad \text{for } i = 1, 2, \dots, s. \quad (\text{A.2})$$

Table 10
OBSERVED CONTINGENCY TABLE

Sub-population	Response profile categories				
	1	2	...	r	Total
1	n_{11}	n_{12}	...	n_{1r}	n_1
2	n_{21}	n_{22}	...	n_{2r}	n_2
...
s	n_{s1}	n_{s2}	...	n_{sr}	n_s

Let $\mathbf{p}_i = (\mathbf{n}_i/n_i)$ be the $(r \times 1)$ vector of observed proportions associated with the sample from the i th sub-population and let \mathbf{p} be the $(sr \times 1)$ compound vector defined by $\mathbf{p}' = (\mathbf{p}_1', \mathbf{p}_2', \dots, \mathbf{p}_s')$. Thus, the vector \mathbf{p} is the unrestricted maximum likelihood estimator (MLE) of $\boldsymbol{\pi}$ where $\boldsymbol{\pi}' = (\boldsymbol{\pi}_1', \boldsymbol{\pi}_2', \dots, \boldsymbol{\pi}_s')$. A consistent estimator for the covariance matrix of \mathbf{p} is given by the $(sr \times sr)$ block diagonal matrix $\mathbf{V}(\mathbf{p})$ with the matrices

$$\mathbf{V}_i(\mathbf{p}_i) = \frac{1}{n_i} [\mathbf{D}_{\mathbf{p}_i} - \mathbf{p}_i \mathbf{p}_i'] \quad (\text{A.3})$$

$(r \times r)$

for $i = 1, 2, \dots, s$ on the main diagonal, where $\mathbf{D}_{\mathbf{p}_i}$ is an $(r \times r)$ diagonal matrix with elements of the vector \mathbf{p}_i on the main diagonal.

Let $F_1(\mathbf{p}), F_2(\mathbf{p}), \dots, F_u(\mathbf{p})$ be a set of u functions of \mathbf{p} which pertain to some aspect of the relationship between the distribution of the response profiles and the nature of the sub-populations. Each of these functions is assumed to have continuous partial derivatives through order two with respect to the elements of \mathbf{p} within an open region containing $\boldsymbol{\pi} = \mathbf{E}\{\mathbf{p}\}$. If $\mathbf{F} \equiv \mathbf{F}(\mathbf{p})$ is defined by

$$\mathbf{F}' = [\mathbf{F}(\mathbf{p})]' = [F_1(\mathbf{p}), F_2(\mathbf{p}), \dots, F_u(\mathbf{p})], \quad (\text{A.4})$$

then a consistent estimator for the covariance matrix of \mathbf{F} is the $(u \times u)$ matrix

$$\mathbf{V}_{\mathbf{F}} = \mathbf{H}[\mathbf{V}(\mathbf{p})]\mathbf{H}', \quad (\text{A.5})$$

where $\mathbf{H} = [d\mathbf{F}(\mathbf{x})/d\mathbf{x} \mid \mathbf{x} = \mathbf{p}]$ is the $(u \times sr)$ matrix of first partial derivatives of the functions \mathbf{F} evaluated at \mathbf{p} . In all applications, the functions comprising \mathbf{F} are chosen so that $\mathbf{V}_{\mathbf{F}}$ is asymptotically nonsingular.

The function vector \mathbf{F} is a consistent estimator of $\mathbf{F}(\boldsymbol{\pi})$. Hence, the variation among the elements of $\mathbf{F}(\boldsymbol{\pi})$ can be investigated by fitting linear regression models by the method of weighted least squares. This phase of the analysis can be characterized by writing

$$\mathbf{E}_{\mathbf{A}}\{\mathbf{F}\} \equiv \mathbf{E}_{\mathbf{A}}\{\mathbf{F}(\mathbf{p})\} = \mathbf{F}(\boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\beta}, \quad (\text{A.6})$$

where \mathbf{X} is a pre-specified $(u \times t)$ design (or independent variable) matrix of known coefficients with full rank $t \leq u$, $\boldsymbol{\beta}$ is an unknown $(t \times 1)$ vector of parameters, and “ $\mathbf{E}_{\mathbf{A}}$ ” means “asymptotic expectation”.

An appropriate test statistic for the goodness-of-fit of the model (A.6) is

$$Q = Q(\mathbf{X}, \mathbf{F}) = (\mathbf{R}\mathbf{F})'[\mathbf{R}\mathbf{V}_{\mathbf{F}}\mathbf{R}']^{-1}\mathbf{R}\mathbf{F}, \quad (\text{A.7})$$

where \mathbf{R} is any full rank $[(u - t) \times u]$ matrix orthogonal to \mathbf{X} . Here Q is approximately distributed according to the χ^2 distribution with d.f. = $(u - t)$ if the sample sizes $\{n_i\}$ are sufficiently large that the elements of the vector \mathbf{F} have an approximate multivariate normal distribution as a consequence of Central Limit Theory (CLT). Test statistics such as Q are known as generalized Wald [1943] statistics and various aspects of their application to a broad range of problems involving the analysis of multivariate categorical data are discussed in Bhapkar and Koch [1968a, 1968b] and in GSK [1969].

However, these test statistics like (A.7) are obtained in actual practice by using weighted least squares

as a computational algorithm which is justified on the basis of the fact that Q of (A.7) is identically equal to

$$Q = (\mathbf{F} - \mathbf{X}\mathbf{b})'\mathbf{V}_F^{-1}(\mathbf{F} - \mathbf{X}\mathbf{b}) \quad (\text{A.8})$$

where $\mathbf{b} = (\mathbf{X}'\mathbf{V}_F^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_F^{-1}\mathbf{F}$ is a BAN estimator for $\boldsymbol{\beta}$ based on the linearized modified χ_1^2 -statistic of Neyman [1949]. In view of this identity demonstrated in Bhapkar [1966], both Q and \mathbf{b} are regarded as having reasonable statistical properties in samples which are sufficiently large for applying CLT to the functions \mathbf{F} . As a result, a consistent estimator for the covariance matrix of \mathbf{b} is given by

$$\mathbf{V}_b = (\mathbf{X}'\mathbf{V}_F^{-1}\mathbf{X})^{-1}. \quad (\text{A.9})$$

If the model (A.6) does adequately characterize the vector $\mathbf{F}(\boldsymbol{\pi})$, tests of linear hypotheses pertaining to the parameters $\boldsymbol{\beta}$ can be undertaken by standard multiple regression procedures. In particular, for a general hypothesis of the form,

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}, \quad (\text{A.10})$$

where \mathbf{C} is a known $(c \times t)$ matrix of full rank $c \leq t$ and $\mathbf{0}$ is a $(c \times 1)$ vector of 0's, a suitable test statistic is

$$Q_c = (\mathbf{C}\mathbf{b})'[\mathbf{C}(\mathbf{X}'\mathbf{V}_F^{-1}\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\mathbf{b} \quad (\text{A.11})$$

which has approximately a χ^2 -distribution with d.f. = c in large samples under H_0 in (A.10).

In this framework, the test statistic Q_c reflects the amount by which the goodness-of-fit statistic (A.8) would increase if the model (A.6) were simplified (or reduced) by substitutions based on the additional constraints implied by (A.10). Thus, these methods permit the total variation within $\mathbf{F}(\boldsymbol{\pi})$ to be partitioned into specific sources and hence represent a statistically valid analysis of variance for the corresponding estimator functions \mathbf{F} .

Predicted values for $\mathbf{F}(\boldsymbol{\pi})$ based on the model (A.6) can be calculated from

$$\hat{\mathbf{F}} = \mathbf{X}\mathbf{b} = \mathbf{X}(\mathbf{X}'\mathbf{V}_F^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_F^{-1}\mathbf{F}. \quad (\text{A.12})$$

Thus, consistent estimators for the variances of the elements of $\hat{\mathbf{F}}$ can be obtained from the diagonal elements of

$$\mathbf{V}_{\hat{\mathbf{F}}} = \mathbf{X}(\mathbf{X}'\mathbf{V}_F^{-1}\mathbf{X})^{-1}\mathbf{X}'. \quad (\text{A.13})$$

The predicted values $\hat{\mathbf{F}}$ not only have the advantage of characterizing essentially all the important features of the variation in $\mathbf{F}(\boldsymbol{\pi})$, but also represent better estimators than the original function statistics \mathbf{F} since they are based on the data from the entire sample as opposed to its component parts. Moreover, they are descriptively advantageous in the sense that they make trends more apparent and permit a clearer interpretation of the relationship between $\mathbf{F}(\boldsymbol{\pi})$ and the variables comprising the columns of \mathbf{X} .

Although the formulation of $\mathbf{F}(\mathbf{p})$ can be quite general, GSK [1969] and Forthofer and Koch [1973] demonstrated that a wide range of problems in categorical data analysis could be considered within the framework of a few specified classes of compounded logarithmic, exponential and linear functions of the observed proportions. However, these functions are all special cases of a broad class of functions which can be expressed in terms of repeated applications of any sequence of the following matrix operations:

(i) *Linear transformations of the type*

$$\mathbf{F}_1(\mathbf{p}) = \mathbf{A}_1\mathbf{p} = \mathbf{a}_1, \quad (\text{A.14})$$

where \mathbf{A}_1 is a matrix of known constants;

(ii) *Logarithmic transformations of the type*

$$\mathbf{F}_2(\mathbf{p}) = \mathbf{log}_e(\mathbf{p}) = \mathbf{a}_2, \quad (\text{A.15})$$

where \mathbf{log}_e transforms a vector to the corresponding vector of natural logarithms;

(iii) *Exponential transformations of the type*

$$\mathbf{F}_3(\mathbf{p}) = \mathbf{exp}(\mathbf{p}) = \mathbf{a}_3, \quad (\text{A.16})$$

where \mathbf{exp} transforms a vector to the corresponding vector of exponential functions, i.e., of antilogarithms.

Then the linearized Taylor-series-based estimate of the covariance matrix of \mathbf{F}_k for $k = 1, 2, 3$, is given by (A.5), where the corresponding \mathbf{H}_k matrix operator is

$$\mathbf{H}_1 = \mathbf{A}_1 ; \quad (\text{A.17})$$

$$\mathbf{H}_2 = \mathbf{D}_p^{-1} ; \quad (\text{A.18})$$

$$\mathbf{H}_3 = \mathbf{D}_{a_3} , \quad (\text{A.19})$$

where \mathbf{D}_y is a diagonal matrix with elements of the vector \mathbf{y} on the main diagonal.

The hypotheses involving marginal distributions can all be tested in terms of linear functions of the form given in (A.14). Furthermore, compounded logarithmic-exponential-linear functions of the form

$$\mathbf{F}(\mathbf{p}) = \exp [\mathbf{A}_4(\log_e \{ \mathbf{A}_3[\exp \{ \mathbf{A}_2 \{ \log_e [\mathbf{A}_1 \mathbf{p}] \} \} \})] \quad (\text{A.20})$$

can be used to generate complex ratio estimates such as generalized kappa-type statistics described in Landis and Koch [1977a, 1977b]. As a result, the linearized Taylor-series-based estimate of the covariance matrix associated with $\mathbf{F}(\mathbf{p})$ in (A.20) can be obtained by repeated application of the chain rule for matrix differentiation. In particular, let

$$\mathbf{a}_1 = \mathbf{A}_1 \mathbf{p} ; \quad (\text{A.21})$$

$$\mathbf{a}_2 = \exp \{ \mathbf{A}_2 [\log_e (\mathbf{a}_1)] \} ; \quad (\text{A.22})$$

$$\mathbf{a}_3 = \mathbf{A}_3 \mathbf{a}_2 ; \quad (\text{A.23})$$

$$\mathbf{a}_4 = \exp \{ \mathbf{A}_4 [\log_e (\mathbf{a}_3)] \} . \quad (\text{A.24})$$

Then the results in (A.17)-(A.19) can be used to provide a consistent estimate of the covariance matrix via (A.5) by using

$$\mathbf{H} = \mathbf{D}_{a_4} \mathbf{A}_4 \mathbf{D}_{a_3}^{-1} \mathbf{A}_3 \mathbf{D}_{a_2} \mathbf{A}_2 \mathbf{D}_{a_1}^{-1} \mathbf{A}_1 . \quad (\text{A.25})$$

Since the methods of analysis described here are based on asymptotic considerations, the functions $\mathbf{F}(\mathbf{p})$ must be chosen carefully so as to support the inherent assumption that the statistical behavior (i.e., expected values and covariance structure) of such functions is approximately characterized by their linearized Taylor series counterparts. However, by virtue of the uniqueness of the Taylor series with respect to \mathbf{p} , the reader should realize that this issue pertains to the inherent nature of such functions rather than to the complexity of their compounded function representation via (A.20). As stated previously, the primary purpose of (A.2) is the computational convenience achieved for (A.25) by such simplistic operations as matrix multiplication, diagonal matrix inversion, \log_e and \exp . Thus, although the complexity of (A.20) should be a source of concern regarding the use of asymptotic methods, it should not dominate such decisions, but rather should call attention to the necessity of alternative arguments for justifying the applicability of the linearized Taylor series approximations. Examples where such justification can be given for moderately large sample sizes include the Goodman-Kruskal rank correlation coefficients (see Forthofer and Koch [1973]), life-table based survival rates (see Koch, Johnson and Tolley [1972]), and generalized kappa type measures of agreement (see Landis and Koch [1977a]). On the other hand, for those situations where the linear Taylor series approximation is not valid but Central Limit Theory arguments are still appropriate with respect to $\mathbf{F}(\mathbf{p})$, then other procedures (e.g., Jack-knife methods) should be used to obtain the required consistent estimator for the covariance structure of \mathbf{F} , and the weighted least squares aspects of the analysis (A.6)-(A.13) can proceed as previously described with no loss of generality.

Koch, Imrey, Freeman and Tolley [1976] discuss the application of this general approach to implicitly defined functions of \mathbf{p} in the context of estimated parameters from fitted log-linear models. All aspects of this methodology can be directed at implicit functions which are based on maximum likelihood estimation equations corresponding to preliminary or intermediate (as opposed to final) models with a priori assumed validity; in other words, models in which the likelihood (A.1) initially, i.e., prior to any data analysis, satisfies both (A.2) as well as certain other constraints analogous to (A.6).

For purposes of completeness, it should be noted that other statistical procedures for the analysis of categorical data from repeated measurement experiments are available in the literature. Bishop, Fienberg and Holland [1975, Chapter 8] discuss the application of maximum likelihood methods to test hypotheses of total symmetry and marginal symmetry as well as certain other hypotheses of interest. They also provide

a complete literature review of other papers dealing with similar questions including the early work of Bowker [1948] and Cochran [1950] as well as the minimum discrimination information approach of Kullback [1971]. However, all of these methods are intimately connected to the product multinomial model (A.1) which involves the joint probability π and the observed frequency vector \mathbf{n} . Thus, for applications involving large tables, e.g., $d \geq 4$, $L \geq 4$, so that $r = L^d \geq 256$, the sample sizes required for their statistical validity may be considerably larger than those required for the weighted least squares estimators which are directed at the first order marginal probabilities $\{\phi_{igk}\}$ or mean scores $\{\eta_{ig}\}$. Moreover, they may also require a prohibitively greater amount of computational effort in such situations because they involve the manipulation of the full $(s \times r)$ contingency table. On the other hand, this difficulty can be bypassed with the approach presented in this paper by constructing the estimators \mathbf{F} for the first order marginal probabilities $\{\phi_{igk}\}$ and mean scores $\{\eta_{ig}\}$ directly from the observed raw data for each subject (as opposed to the corresponding contingency table) by the indicator function operations described in Appendix 2. For these reasons, weighted least squares oriented methods may often provide more effective analyses than maximum likelihood oriented methods for repeated measurement experiments with categorical data, particularly for those involving large underlying contingency tables with sparsely distributed samples.

Appendix 2

Special Considerations for Large Tables and Zero Cells

Unless both the number of conditions d and the number of response categories L are small, e.g., $d \leq 3$, $L \leq 3$, the number of possible multivariate response profiles $r = L^d$ can become very large. In this event, there are two sources of potential difficulty:

- i. The manipulation of very large contingency tables and operator matrices can involve expensive computations.
- ii. For each sub-population $i = 1, 2, \dots, s$, some of the r possible response profiles \mathbf{j} will not necessarily be observed so that the corresponding cell frequencies $n_{i\mathbf{j}}$ are zero.

As noted in Section 2, the hypotheses of primary interest in repeated measurement experiments (or surveys) involve the first order marginal probabilities $\{\phi_{igk}\}$. Unbiased estimators for these quantities can be obtained by applying \mathbf{A}_i matrix operations (A.14) to form the appropriate linear functions of the observed proportion vector \mathbf{p} . Alternatively, these same estimators can be generated by forming appropriate functions of the raw data from each subject and then computing the corresponding across subject arithmetic means within each of the respective sub-populations. In particular, let

$$y_{igkm} = \begin{cases} 1 & \text{if the } m\text{th subject in the sample from the } i\text{th sub-population has the response for the } g\text{th} \\ & \text{condition classified into the } k\text{th category} \\ 0 & \text{otherwise} \end{cases} \quad (\text{A.26})$$

where $i = 1, 2, \dots, s$; $g = 1, 2, \dots, d$; $k = 1, 2, \dots, (L - 1)$; and $m = 1, 2, \dots, n_i$. If

$$\bar{y}_{igk} = \frac{1}{n_i} \sum_{m=1}^{n_i} y_{igkm}, \quad (\text{A.27})$$

$$\bar{\mathbf{y}}_i' = (\bar{y}_{i11}, \dots, \bar{y}_{i1,(L-1)}, \dots, \bar{y}_{id1}, \dots, \bar{y}_{id,(L-1)}), \quad (\text{A.28})$$

$$\bar{\mathbf{y}} = (\bar{\mathbf{y}}_1', \bar{\mathbf{y}}_2', \dots, \bar{\mathbf{y}}_s'), \quad (\text{A.29})$$

then $\bar{\mathbf{y}} = \mathbf{a}$ where $\mathbf{a} = \mathbf{A}_i \mathbf{p}$ as defined in (A.14). Moreover, the covariance matrix $\mathbf{V}_{\bar{\mathbf{y}}}$ of $\bar{\mathbf{y}}$ is block diagonal with the matrices

$$\mathbf{V}_{\bar{\mathbf{y}}_i} = \frac{1}{n_i^2} \sum_{m=1}^{n_i} (\mathbf{y}_{im} - \bar{\mathbf{y}}_i)(\mathbf{y}_{im} - \bar{\mathbf{y}}_i)' \quad (\text{A.30})$$

where

$$\mathbf{y}_{im}' = (y_{i11m}, \dots, y_{i1,(L-1),m}, \dots, y_{id1m}, \dots, y_{id,(L-1),m}) \quad (\text{A.31})$$

on the main diagonal; and

$$\mathbf{V}_{\bar{\mathbf{y}}} \equiv \mathbf{V}_a = \mathbf{A}_i \mathbf{V}[(\mathbf{p})] \mathbf{A}_i' \quad (\text{A.32})$$

as given in (A.5) with (A.17). Since $sd(L - 1)$ is usually moderate in size, this method of computation of the estimators for the *first order marginal probabilities* is reasonably straightforward and efficient. Moreover, it can be similarly applied to compute *mean scores* based on quantitative scales a_1, a_2, \dots, a_L for ordinal responses. Finally, as indicated in Landis, Stanish, Freeman and Koch [1976], these operations can be readily linked with other algorithms for performing logarithmic and exponential operations as shown in (A.15)-(A.20) as well as weighted least squares regression analysis. Thus, this computational strategy represents an effective way of dealing with (i). Examples which illustrate its application are given in Koch, Freeman and Lehen [1976] and Landis and Koch [1977b].

For repeated measurement experiments, the potential tendency for some of the n_{ij} to be zero does not cause any real problems except when such zero frequencies induce singularities in the estimated covariance matrix $\mathbf{V}\mathbf{F}$ in (A.5) for the function vector \mathbf{F} which is to be analyzed, or otherwise restrict the extent to which CLT arguments can be applied to the distribution of \mathbf{F} . With respect to the hypotheses H_{SJ} and H_{CJ} , CLT arguments cannot be applied with confidence to the estimators $p_{ij} = (n_{ij}/n_i)$ of the π_{ij} unless most of the $n_{ij} \geq 5$.

This condition, however, is usually not satisfied for situations involving moderately large samples, $n_i \geq 100$, except when both L and d are small, i.e., $d \leq 3, L \leq 3$. Thus, in such cases, the potential presence of many zero frequencies implies that these hypotheses cannot be validly tested by the general methodological approach given in this paper.

On the other hand, if attention can be restricted to hypotheses like H_{SM} and H_{CM} or H_{SAM} and H_{CAM} , which involve the *first order marginal probabilities* $\{\phi_{igk}\}$, then the sample size requirements are considerably less severe. In particular, CLT can be applied with reasonable validity to estimates of the $\{\phi_{igk}\}$ obtained either by \mathbf{A}_i matrix operations like (A.14) or by direct computations like (A.27) provided the overall within sub-population sample sizes $n_i \geq 25$ and most of the first order marginal frequencies

$$n_{igk} = \sum_{j \text{ with } i_g = k} \dots \sum n_{i_1, i_2, \dots, i_d} \geq 5. \quad (\text{A.33})$$

Since $\sum_{k=1}^L n_{igk} = n_i$, the conditions (A.33) tend to hold in most situations where the sample sizes $n_i \geq 25$ and where the number of response categories is small, i.e., $L \leq 3$.

However, if L is not small, then cases where (A.33) tends not to hold can be handled by pooling certain response categories together so that (A.33) is satisfied by the pooled frequencies. Of course, attempts should be made to base such pooling on conceptual similarities among the response categories as well as to prevent such pooling from masking any obvious differences among either sub-populations or among conditions. Otherwise, the logical relationship between the interpretation of the hypotheses H_{SM} (or H_{CM}) for the pooled categories vs. the original categories is analogous to the logical relationship between H_{SAM} (or H_{CAM}) and H_{SM} (or H_{CM}). Alternatively, many of the situations where L is large involve ordinally scaled response categories, and hence, the quantities of interest are estimates of *mean score functions* $\{\eta_{ig}\}$. CLT can be applied with reasonable validity to these statistics provided the overall within sub-population sample sizes $n_i \geq 25$, and the corresponding observed response distributions for each measurement condition are not degenerate in the sense of being concentrated in a single response category. This latter requirement is satisfied for most purposes if there exist response categories $k(i, g)$ for each sub-population x condition combination such that,

$$\sum_{l=1}^{k(i, g)} n_{ig l} \geq 5, \quad \sum_{l=(k(i, g)+1)}^L n_{ig l} \geq 5. \quad (\text{A.34})$$

The previous remarks deal with the extent to which the presence of zero frequencies affects the types of functions to which CLT arguments can be applied in a valid manner. Moreover, when such functions are chosen judiciously, the resulting estimated covariance matrix $\mathbf{V}\mathbf{F}$ usually will be nonsingular. However, if $\mathbf{V}\mathbf{F}$ is singular, then a number of technical difficulties arise in applying the Appendix 1 methodology because matrix inversion of $\mathbf{V}\mathbf{F}$ can no longer be performed. For these situations, the vector \mathbf{F} can be analyzed by the Appendix 1 methodology provided that the estimated covariance matrix $\mathbf{V}\mathbf{F}$ is replaced by a suitably similar, nonsingular, pseudo estimated covariance matrix $\mathbf{V}\mathbf{F}^*$.

If both L and d are small, e.g., $d \leq 3, L \leq 3$, such a nonsingular matrix $\mathbf{V}\mathbf{F}^*$ can be obtained by applying the same operations used to form $\mathbf{V}\mathbf{F}$ to the pseudo full ($s \times r$) contingency table in which some of the "0" frequencies are replaced by an appropriate small number like "(1/2)." The motivation behind this is the findings of Berkson [1955] who investigated its effects on estimation involving the logit, i.e., natural logarithm of the ratio of the proportions, associated with $r = 2$ responses. For the case where $r = L^d$ is large, the properties of this rule are largely unknown; and thus the recommendation here is *not* to replace

all "0" frequencies by $(1/2)$, but to modify only those which are necessary to the construction of a non-singular estimated covariance matrix $\mathbf{V_F^*}$ for \mathbf{F} . Other aspects of the analysis of \mathbf{F} then proceed as described in Appendix 1, except that $\mathbf{V_F^*}$ is used as the estimated covariance matrix instead of $\mathbf{V_F}$.

The previously described method for constructing $\mathbf{V_F^*}$ is not practical when $d \geq 4$, $L \geq 4$ because the full contingency table is too large to manipulate efficiently. Moreover, in these cases, attention is primarily directed at the hypotheses pertaining to the *first order marginal probabilities* $\{\phi_{igk}\}$ for which unbiased estimators $\{\bar{y}_{igk}\}$ are determined via (A.26) and (A.27) and the corresponding estimated covariance matrix $\mathbf{V_{\bar{y}}}$ is determined via (A.30). Unfortunately, if $\mathbf{V_{\bar{y}}}$ is singular, considerable effort may be required to construct a nonsingular, pseudo estimated covariance matrix $\mathbf{V_{\bar{y}}^*}$ for $\bar{\mathbf{y}}$. For this purpose, the basic strategy is to augment the raw data for the respective subjects with pseudo data vectors with smaller weights like $(1/2)$ until the required nonsingular matrix $\mathbf{V_{\bar{y}}^*}$ is determined. For further discussion, see Koch, Freeman, Freeman, and Lehnen [1974].

In summary, questions pertaining to zero cells and singularities are best resolved by statistical judgments which take into account the special features of each experimental (or survey) situation where such methods are applied. For many situations, this can be accomplished by restricting attention to hypotheses involving the *first order marginal probabilities* $\{\phi_{igk}\}$ (or *mean scores* $\{\eta_{ig}\}$). In other cases, the use of modular MLEs as discussed in Koch and Tolley [1975] and Koch, Imrey, Freeman and Tolley [1976] may be required to "smooth" away the singularities in $\mathbf{V_F}$ through the assumption of certain a priori constraints on the likelihood (A.1). However, if singularities are still present, then the analysis of \mathbf{F} can often be effectively undertaken through the use, with appropriate caution, of a non-singular, pseudo estimated covariance matrix $\mathbf{V_F^*}$.