

This research was supported by National Institutes of Health,  
Institute of General Medical Sciences, in part by Career Development  
Grant No. GM 0038-16 and in part by Grant No. GM 12868-06.

THE ANALYSIS OF CATEGORICAL DATA  
WITH SUPPLEMENTED MARGINS INCLUDING  
APPLICATIONS TO MIXED MODELS

by

Donald William Reinfurt  
Department of Biostatistics  
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 697

July 1970

## ABSTRACT

REINFURT, DONALD WILLIAM. The Analysis of Categorical Data with Supplemented Margins Including Applications to Mixed Models. (Under the direction of GARY GROVE KOCH.)

The general linear model approach for the analysis of categorical data is extended to the case of supplemental data. In the usual case, each individual is classified according to each of the  $d$  dimensions of the corresponding contingency table. With supplementation, there is additional information on  $d' < d$  dimensions for a random sample selected from the same homogeneous population. This situation might arise if there was special interest in only certain of the responses and/or possibly for reasons of economy and is distinctly different from the empty cell problem.

Starting with the simplest case, maximum likelihood and iterated weighted least squares estimates are obtained for the basic parameters for  $2 \times 2$  tables with one margin supplemented. The maximum likelihood estimates are unbiased and have the same asymptotic covariance matrix as do the iterated weighted least squares estimates. Using conditional expectations and variances, improved estimates of the precision of the maximum likelihood estimates are obtained. For the case with both margins supplemented, iterative solutions are indicated. Extensions to  $r \times c$  tables and higher-dimensional tables are outlined.

A general two-stage test procedure is presented which requires certain modifications of the existing computer program. Several detailed examples are given including mixed categorical data models with supplemented margins. The situations in which these special "no factor, multi-response" models arise are identified as are the

appropriate tests of hypotheses, namely, tests of marginal symmetry and, for quantitative variables, tests of equality of mean scores. It is noted that supplementation should be of most use in exactly these mixed categorical data model situations.

## BIOGRAPHY

The author was born August 30, 1938, in Wilkes-Barre, Pennsylvania, and, being the son of a Methodist minister, was raised in the eastern part of the United States. He graduated from Unadilla Central High School in Unadilla, New York, in 1956.

He received the Bachelor of Science degree with a major in mathematics from the State University of New York at Albany in 1960, and the Master of Arts degree in mathematics from the State University of New York at Buffalo in 1963 where he was a teaching fellow. Since then, the author has combined his studies in the Department of Experimental Statistics at North Carolina State University at Raleigh with work at the North Carolina Highway Safety Research Center in Chapel Hill.

The author is married to the former Mary Karen Hillix of St. Joseph, Missouri, and has a daughter, Kristin Elizabeth, 18 months old.

## ACKNOWLEDGMENTS

The author extends his sincere appreciation to the many persons who have helped in so many ways in the preparation of this study. He is especially indebted to his advisor, Dr. Gary G. Koch, for his invaluable advice, guidance, and constant encouragement. In addition, he would like to thank the other members of his advisory committee, Drs. John T. Gentry, Daniel G. Horvitz, Henry L. Lucas, and Harry Smith, Jr., for their constructive criticism and suggestions.

Special gratitude is expressed to his employer, Dr. B. J. Campbell, Director of the North Carolina Highway Safety Research Center, for making it possible to complete this research at an earlier date than would otherwise have been possible.

Generous financial support for this study has been provided by the National Institutes of Health, Institute of General Medical Services.

Finally, the author expresses his thanks to his wife, Karen, and his daughter, Kristin, for their continued patience and sacrifice during the course of this research.

## TABLE OF CONTENTS

	Page
LIST OF TABLES . . . . .	vi
1. INTRODUCTION AND REVIEW OF LITERATURE . . . . .	1
1.1 Introduction . . . . .	1
1.2 General Categorical Data Models . . . . .	2
1.2.1 Historical Remarks . . . . .	2
1.2.2 Some Recent Developments in Model Formulation. . . . .	4
1.2.3 The Categorical Data Framework . . . . .	6
1.2.4 Hypotheses of Interest . . . . .	9
1.3 Methods of Inference . . . . .	11
1.3.1 Hypothesis Formulation . . . . .	11
1.3.2 Estimation Procedures . . . . .	13
1.3.3 Test Statistics . . . . .	17
1.4 The General Linear Model . . . . .	20
1.4.1 Notation and Assumptions . . . . .	20
1.4.2 Methods of Inference . . . . .	22
2. MIXED CATEGORICAL DATA MODELS . . . . .	28
2.1 Introduction . . . . .	28
2.2 Mixed Models in General . . . . .	30
2.3 Mixed Categorical Data Models of Order 2 . . . . .	32
2.4 Mixed Categorical Data Models of Higher Orders . . . . .	37
2.5 Split Plot Contingency Tables . . . . .	41
3. MORE GENERAL CATEGORICAL DATA MODELS FOR THE CASE OF SUPPLEMENTED MARGINS . . . . .	47
3.1 Introduction . . . . .	47
3.2 More General Linear Models in the Case of Normal Random Variables . . . . .	50
3.3 Notation for the More General Categorical Data Model . . . . .	53
3.4 Estimation . . . . .	56
3.5 Tests of Hypotheses . . . . .	59
3.6 Examples Using the More General Model . . . . .	60
3.6.1 Perinatal Health Survey Example . . . . .	60
3.6.2 Hand Example . . . . .	65
3.6.3 Drug Example . . . . .	73

## TABLE OF CONTENTS (continued)

	Page
4. ESTIMATION THEORY AND HYPOTHESIS TESTING FOR TWO-WAY CONTINGENCY TABLES WITH SUPPLEMENTED MARGINS . . . . .	81
4.1 Introduction . . . . .	81
4.2 2 x 2 Case with One Margin Supplemented . . . . .	81
4.2.1 Notation . . . . .	81
4.2.2 Methods of Estimation . . . . .	82
4.2.3 Properties of the Estimators . . . . .	88
4.3 2 x 2 Case with Both Margins Supplemented . . . . .	94
4.3.1 Methods of Estimation . . . . .	94
4.3.2 Properties of the Estimators . . . . .	100
4.4 Extension of the Theory of Supplemented Margins . . . . .	103
4.4.1 The Case of One Margin Supplemented . . . . .	103
4.4.2 The Case of Several Margins Supplemented . . . . .	107
4.5 Tests of Hypotheses . . . . .	108
5. SUMMARY AND SUGGESTIONS FOR FURTHER RESEARCH . . . . .	113
5.1 Summary . . . . .	113
5.2 Suggestions for Further Research . . . . .	117
6. LIST OF REFERENCES . . . . .	119
7. APPENDIX . . . . .	123

## LIST OF TABLES

	Page
1.4.1.1 Expected cell probabilities (cell frequencies) for the standard contingency table . . . . .	20
2.2.1 Summary of assumptions and appropriate test procedures for various types of mixed models . . . . .	32
2.3.1 Vision records for women employees of the Royal Ordnance factories . . . . .	33
2.3.2 Frequency distribution of the difference between right and left eye grades . . . . .	36
2.4.1 Responses of patients to drugs A, B, and C . . . . .	37
2.5.1 Responses of males to culturally masculine and anatomically feminine objects with weak intensity . . . . .	42
3.3.1 Frequency distribution for experimental units with complete information . . . . .	55
3.3.2 Frequency distribution for experimental units with information on A and B only . . . . .	55
3.3.3 Frequency distribution for experimental units with information on A only . . . . .	55
3.6.1.1 Serum bilirubin readings and 5-minute Apgar scores for a sample of premature infants . . . . .	62
3.6.2.1 Finger data indicating classification with respect to the presence or absence of sweat or sensation . . . . .	66
3.6.2.2 Frequency distribution of complete and supplemental data by agreement (A) and disagreement (D) for each finger . . . . .	67
3.6.3.1 Frequency distribution of responses to drugs A and B . . . . .	74
4.2.1.1 Expected cell probabilities (cell frequencies) for the 2 x 2 case with one margin supplemented . . . . .	82
4.5.1 Observed frequencies for a 2 x 2 contingency table with both margins supplemented . . . . .	110
7.1 Iterative solution for the MLE's ( $p_{ij}$ ) and WLSE's ( $\tilde{\pi}_{ij}$ ) for Table 4.5.1 . . . . .	125

## 1. INTRODUCTION AND REVIEW OF LITERATURE

### 1.1 Introduction

In the usual categorical data situation, there is complete information available for each entry in the corresponding contingency table, i.e., each unit is classified according to each dimension of the table. However, in many situations, the experimenter is not equally interested in all of the response variates or perhaps he is able to obtain additional information on a subset of the dimensions of classification relatively easily and economically. The question then arises as to how this additional or supplemental information might be used in improving the estimates of the basic probabilities of the table, thereby improving the corresponding tests of hypotheses.

The primary goal of this research is to extend the general linear model approach for the analysis of categorical data (cf. Grizzle et al., 1969) to the above-mentioned case of contingency tables with supplemented margins. The following questions are relevant to this investigation: What are the properties of maximum likelihood estimates (MLE's) and iterated weighted least squares estimates (WLSE's) with and without supplementation? What method of estimation should be used in a given situation? How can the results on estimation be incorporated into the general Grizzle, Starmer, Koch (GSK) framework for hypothesis testing? What are the properties of these tests and for which tests is supplementation most useful?

There are two additional goals of this research. The first is the investigation of a particular "no factor, multi-response" situation in which supplementation would appear to be most useful since the main

hypotheses of interest are ones of marginal symmetry and, if the response variables are quantitative, equality of mean scores. How can the GSK approach be applied to analyze the most general of these "mixed categorical data model" situations with supplementation?

The second is a comprehensive review of the vast and scattered literature on categorical data methods. There is special emphasis on the linear model approach since it is the method employed in the remainder of this research.

## 1.2 General Categorical Data Models

### 1.2.1 Historical Remarks

Categorical data and its analysis have been of interest to statisticians ever since the earliest origins of the subject. For example, in the area of vital statistics, Graunt, in the early 1600's, presented data on causes of death (e.g., consumption, diseases of infancy, tooth diseases) which appeared in the form of frequency tabulations according to the sex and marital status of the deceased. Since 1790, census studies have collected and analyzed numerous cross-tabulations on a wide variety of demographic variables. In the area of genetics, Mendel, in the mid-1800's, collected data on garden pea varieties (one-way tables) leading to the now-famous Mendelian heritability ratios (e.g., with complete dominance for each of two characteristics, genotype ratios of offspring of 9:3:3:1). In the field of epidemiology, Greenwood and Yule, in the early 1900's, had considerable data on typhoid attacks on inoculated and non-inoculated individuals. The list of other examples is virtually unlimited.

In the analysis of categorical data, the binomial and Poisson probability distributions and later the multinomial distribution have played an essential role. One of the most important discoveries (Laplace, de Moivre) in early statistical theory was that, as the sample size becomes "large", data generated by any of these distributions is approximately normally or multinormally distributed. This property has been used repeatedly in developing appropriate methods for analyzing categorical data from the original Pearson chi-square ( $\chi^2$ ) statistic to the more complex Wald statistics for testing the hypothesis of "no interaction" in multi-dimensional contingency tables.

Aside from this historical background, categorical data appearing in the form of frequency tabulations in one-way, two-way, and multi-way tables is a familiar data array to virtually all statisticians in today's society. In many fields, much of the data arises from questionnaires and thus is often nominal or, at best, ordinal. Such fields include the social sciences, the medical and health sciences, and occasionally the physical sciences.

Unfortunately, one consequence of categorical data being of interest to persons in widely divergent areas of interest has been the development of a number of very different approaches to analyzing such data. In the familiar situations involving continuous random variables, the general linear model includes such apparent special models as multiple regression models, fixed effects, random effects, and mixed effects analysis of variance models, models with covariables, etc. It uses a general least squares approach as the method of inference. On the other hand, with contingency tables, the underlying models have not

always been obvious and hence hypotheses have often been formulated in a fragmented way. As a result, it has appeared that many types of complex categorical data situations required their own special methods of inference. The abundant and colorful literature for contingency tables unfortunately documents this in considerable detail. Thus, the practitioner has often had to search the literature for a procedure appropriate to his problem; when he has failed to find such a technique, he has either been required to use an inappropriate analysis for his data or leave his data unanalyzed.

Another difficulty in the analysis of categorical data has been the existence of numerous methods of estimation and corresponding test criteria upon which the statistical inferences are based. As a result, even when the researcher has decided upon the hypotheses of interest, he still is confronted with a wide choice of particular methods to apply. In most cases, this choice is not crucial since the various methods have been shown to be asymptotically equivalent. However, since various authors use different methods, to the uninitiated the variety of choices is often confusing.

#### 1.2.2 Some Recent Developments in Model Formulation

Pearson (1947) was among the first to note that, even in the simple  $2 \times 2$  contingency table, the same configuration could have arisen from different sampling schemes. Thus, it appeared necessary to carefully specify the underlying probability model since different models could lead to different statistical procedures with obviously different interpretations.

Underscoring the work of Professor S. N. Roy and his students, Kastenbaum, Mitra, Diamond, Bhapkar and Sathe, has been the careful consideration of the sampling scheme from which the data arose. The particular factor-response structure (i.e., sampling scheme) of the model has dictated the hypotheses of relevance just as in univariate and multivariate analysis of variance for continuous variables.

The combined contributions of these researchers form the basis of the general approach found in Bhapkar and Koch (1968a, 1968b) and especially in Grizzle et al. (1969). They recommend multi-factor, multi-response models as the key to formulating relevant hypotheses. Their test statistics are derived through weighted least squares procedures and correspond identically to the minimum modified chi-square statistics due to Neyman (1949) or equivalently the generalized quadratic form criteria due to Wald (1943).

The main alternative school of thought has been developed by Goodman, Good, Bishop, Fienberg, Kullback, Kupperman and Ku, among others. The major papers from this group include Goodman (1968, 1970), Bishop (1969), Bishop and Fienberg (1969), Kullback et al. (1962) and Ku and Kullback (1968). Goodman, Good, Bishop and Fienberg have been motivated by the resemblance of multi-dimensional contingency tables to certain complex factorial designs. They then apply maximum likelihood methods to the assumed multiplicative underlying models although their methods can be modified to apply to other models as well. Much of their effort is directed toward special problems such as incomplete tables. On the other hand, Kullback, Kupperman, and Ku use the method of minimum discrimination information estimation (which closely

resembles maximum likelihood estimation) and the minimum discrimination information statistic as the test criterion "for measuring the divergence of the alternative hypothesis, as evidenced by the sample values, from the null hypothesis."

The major portion of this research will utilize the multi-factor, multi-response framework with statistical inference based on the minimum modified chi-square method of Neyman. However, in certain cases, maximum likelihood results will be given for the sake of completeness.

### 1.2.3 The Categorical Data Framework

In the analysis of any multi-dimensional contingency table, there are three fundamental aspects to be considered. These include

- (i) the specification of the underlying model (e.g., no factor, two response model);
- (ii) the formulation (in terms of this model) of the hypotheses to be tested (e.g., independence of responses) and the calculation of the corresponding test statistics;
- (iii) the interpretation (with respect to the model) of the results (e.g., there is an association or relationship between the two responses).

In order to specify the underlying model, it must be realized that each subject (or experimental unit) may give rise to two types of data.

These are

- (i) a description of the experimental conditions which the subject undergoes (or of the sub-population of units to which the experimental unit belongs) -- henceforth referred to as "factors" or "populations";
- (ii) a description of what subsequently happens to each subject-- henceforth referred to as "responses".

In other words, the data in any multi-dimensional contingency table are

the frequencies with which subjects belonging to the same sub-population or combination of factor categories yielded the same combination of responses. The dimension,  $d$ , of a contingency table refers to the total number of factors and responses (e.g.,  $d = 2$  in the previous example since there were no factors and two responses) while the levels represent the sub-categories within the factors and responses (e.g., handedness has three levels: left, right, and ambidextrous).

More specifically, consider  $r$  independent random samples taken from  $r$  multinomial populations where  $n_{i_0}$  represents the size of the sample from the  $i$ -th population and  $n_{ij}$  the observed number falling in the  $j$ -th category of the  $i$ -th sample where  $i = 1, 2, \dots, r$  and  $j = 1, 2, \dots, s$ . It should be noted that  $i$  and/or  $j$  may be multiple subscripts as in the case of multi-factor and/or multi-response models. For example,

$$i = \underline{i}' = (i_1, i_2, \dots, i_{d_f}) \text{ with } i_\alpha = 1, 2, \dots, r_\alpha, \alpha = 1, 2, \dots, d_f \text{ and}$$

$$j = \underline{j}' = (j_1, j_2, \dots, j_{d_r}) \text{ with } j_\beta = 1, 2, \dots, s_\beta, \beta = 1, 2, \dots, d_r. \text{ In}$$

this example, the model is that of a  $d_f$ -factor,  $d_r$ -response contingency table with the factors at levels  $i_\alpha$ ,  $\alpha = 1, 2, \dots, d_f$ , and the responses at levels  $j_\beta$ ,  $\beta = 1, 2, \dots, d_r$ . Note also that all response-level combinations are assumed to occur with positive probability; however, situations where this assumption does not hold may be handled by methods analogous to those given by Goodman. On the other hand, not all factor-level combinations are required to appear. In particular, if the sample design is an incomplete block design, not all factor-level combinations can appear.

Now, if  $\pi_{ij}$  represents the probability of an observation from the  $i$ -th population (or factor level) falling in the  $j$ -th response category (i.e., if the  $\pi_{ij}$  represent individual cell probabilities), and the sampling is either from a very large population or with replacement, then the probability distribution of the observed frequencies,  $n_{ij}$ , is given by the product-multinomial distribution

$$\phi = \prod_{i=1}^r n_{i0}! \prod_{j=1}^s \frac{(\pi_{ij})^{n_{ij}}}{n_{ij}!} \quad (1.2.3.1)$$

where  $\sum_{j=1}^s \pi_{ij} = 1$  and  $\sum_{j=1}^s n_{ij} = n_{i0}$  (fixed) for all  $i$  and where it is

assumed that  $\pi_{ij} > 0$  for all  $i, j$ . This basic model (1.2.3.1) will be assumed throughout and the hypotheses of interest will be expressible as functions of the unknown probabilities in the model, say  $f(\underline{\pi})$ . It should be noted that, if sampling is without replacement from a relatively small population and hence the sampling fraction is large (say, exceeding 10 percent), methods of analysis based on the theory of sampling from finite populations would be appropriate. The details of the analysis for this situation in terms of finite population correction factors is discussed in Johnson and Koch (1970).<sup>1</sup>

---

<sup>1</sup>Johnson, W. D. and Koch, G. G. Some aspects of the statistical analysis of qualitative data from sample surveys. Part I: Proportions, mean scores, and other linear functions. Submitted to Health Services Research, 1970. Department of Biostatistics, University of North Carolina at Chapel Hill.

#### 1.2.4 Hypotheses of Interest

In order to formulate appropriate hypotheses to be tested, the underlying factor-response structure of the table should be identified. Following Bhapkar and Koch (1968a), the four types of multi-dimensional contingency tables of interest (including examples from the literature) are the following:

- Model I. No factor, multi-response tables  
 Bhapkar (1966): no factor, two responses (left and right eye grades);  $d = 2$
- Model II. Uni-factor, multi-response tables  
 Mosteller (1968): one factor (country), two responses (status of father's occupation, status of son's occupation);  $d = 3$
- Model III. Multi-factor, uni-response tables  
 Bhapkar and Koch (1968a): two factors (driver group, accident type), one response (accident severity);  $d = 3$
- Model IV. Multi-factor, multi-response tables  
 Bhapkar and Koch (1968b): two factors (subject group, anatomical meaning), two responses (classifications at exposure rates of 1/5 second and 1/1000 second);  $d = 4$

Obviously, for one-way tables, only Model I can occur; for two-way tables, only Models I and II can occur and Model II actually represents uni-factor, uni-response tables; for three-way tables, only Models I, II, and III can occur. Otherwise ( $d > 3$ ), all four models can arise in various situations.

To identify the hypotheses appropriate to the different models, consider, for simplicity, the case of three-dimensional tables. In the case of "no factor, three response" tables (Model I), the primary interest lies in the relationships among the three responses which are analogous to problems of independence and correlation in normal

multivariate analysis of variance. Questions of interest include total independence of the three responses, independence of any one response from the other two, pairwise independence, and partial association between two responses within given levels of the third. In addition, as will be explained later, hypotheses of marginal symmetry and equality of mean scores for structured responses (where scores, e.g., ridits or percentile scores, are assigned to the response levels) are of considerable interest.

If the experiment is of the "one factor, two response" type (Model II), interest lies in the association between the responses as well as the effect of the factor on the responses. Here the questions posed are similar to those encountered in one-way normal multivariate analysis of variance. Appropriate hypotheses for this case include independence of the responses within each factor level, homogeneity for each marginal response (i.e., does the factor level affect the marginal distribution of the response?), and joint homogeneity (i.e., does the factor level affect the joint distribution of the responses?).

Finally, if the experiment is of the "two factor, one response" type (Model III), the questions of interest are similar to those arising in univariate normal analysis of variance, i.e., how do the factors (cf. treatments or independent variables) combine to determine the response (cf. "yield" or dependent variable)? Here the hypotheses of interest include total homogeneity (i.e., do both factors affect the distribution of the response?), partial homogeneity (i.e., does one factor affect the distribution of the response?), and "no interaction" between factors in the way they affect the response (i.e., do the

factors determine the response in a purely additive or multiplicative way or are the relationships between the factors and the response more complicated?).

For completeness, note that, for Model IV situations ("multi-factor, multi-response" experiments and hence  $d > 3$ ), interest centers in both the relationships among the responses and in the way the factors combine to affect the responses.

Thus, the structure of the table (which is not always obvious or unambiguous) dictates the types of hypotheses that are appropriate to the given experimental situation. The test statistics used and the interpretation of the results then follow from the specified model and the corresponding hypotheses of interest.

### 1.3 Methods of Inference

#### 1.3.1 Hypothesis Formulation

A general formulation for many hypotheses of interest in the analysis of categorical data (e.g., in two-way tables, independence in the Model I situation, homogeneity in the case of Model II) is given by the following set of constraints on the cell probabilities:

$$H_{01}: \frac{f}{tx1}(\underline{\pi}) = \underline{0} \quad (1.3.1.1)$$

where

$$(\underline{f}(\underline{\pi}))' = (f_1(\underline{\pi}), f_2(\underline{\pi}), \dots, f_t(\underline{\pi})) \quad t \leq r(s-1)$$

with

$$\begin{aligned}\underline{\pi}' &= (\pi_{11}, \pi_{12}, \dots, \pi_{1s}; \pi_{21}, \pi_{22}, \dots, \pi_{2s}; \dots; \pi_{r1}, \pi_{r2}, \dots, \pi_{rs}) \\ &= (\underline{\pi}'_1, \underline{\pi}'_2, \dots, \underline{\pi}'_r)\end{aligned}$$

and the functions,  $f_k(\underline{\pi})$ , are functionally independent and independent of the constraint,  $\sum_{j=1}^s \pi_{ij} = 1$ ,  $i = 1, 2, \dots, r$ . In addition, the

$f_k(\underline{\pi})$  must have continuous first and second partial derivatives with

respect to the  $\pi_{ij}$ 's and  $H_{\text{txrs}}^o(\underline{\pi}) = \left( \frac{\partial f_k(\underline{\pi})}{\partial \pi_{ij}} \right)_{\underline{\pi}=\underline{\pi}^o}$  must be of rank  $t$

(i.e., full rank) for any  $\underline{\pi}$  in the neighborhood of  $\underline{\pi}$ . For the case of independence in the two-way case mentioned previously, the constraints,  $f_k(\underline{\pi}) = 0$ , take the form

$$\ln \left\{ \frac{\pi_{j_1 j_2} \pi_{j'_1 j'_2}}{\pi_{j_1 j'_1} \pi_{j_2 j'_2}} \right\} = 0 \quad j_1 \neq j'_1, j_2 \neq j'_2$$

where  $\ln$  is the natural or Napierian logarithm, and where, for convenience; in the case of the single multinomial (i.e., the situation where two or more responses are given for each member of a given population and hence  $i \equiv 1$  since  $r = 1$ ), the  $i$ -subscript is suppressed. For the case of homogeneity, the  $f_k(\underline{\pi})$  take the form

$$\pi_{ij} - \pi_{i'j} = 0 \quad \text{for all } i \neq i' \text{ and each } j.$$

Alternatively, functions of the cell probabilities can be expressed in the following linear model:

$$H_{o2}: \quad \underline{f}_{\text{txl}}(\underline{\pi}) = \sum_{\text{txu}} \underline{\beta}_{\text{uxl}} \quad (1.3.1.2)$$

where  $\underline{f}(\pi)$  is as before,  $X$  is a known design matrix of rank  $u \leq t$  and  $\underline{\beta}$  is a vector of unknown parameters. It should be noted that  $X$  differs from the usual design matrix when more than one function is constructed within each population. Tests for fit of the model are appropriate under (1.3.1.2). Given that the model fits the data, hypotheses concerning various constraints on the model parameters can be formulated as follows:

$$H_{03}: \begin{matrix} C & \underline{\beta} & = & \underline{0} \\ c_{xu} & u_{x1} & & c_{x1} \end{matrix} \quad (1.3.1.3)$$

where  $C$  is a matrix of known constants of rank  $c \leq u$ . For example, in the case of split-plot categorical models as described in Koch and Reinfurt (1970),<sup>2</sup> the significance of whole-plot effects, split-plot effects, and interaction between whole-plot and split-plot effects may be tested using appropriate  $C$  matrices.

### 1.3.2 Estimation Procedures

If the hypothesis of interest is formulated in terms of constraints on the cell probabilities as in (1.3.1.1), then it is necessary to obtain estimates of the cell probabilities subject to those constraints. These constrained estimates are then incorporated into one of several asymptotically equivalent test criteria and the test is performed.

In the framework of (1.3.1.1), there are essentially two different classes of estimation procedures. These are

---

<sup>2</sup>Koch, G. G. and Reinfurt, D. W. The analysis of categorical data from mixed models. Submitted to Biometrics, 1970. Department of Biostatistics, University of North Carolina at Chapel Hill.

- (1) Maximum likelihood and minimum discrimination information;
- (2) Minimum chi-square (minimum  $\chi^2$ ) and modified minimum chi-square (minimum  $\chi_1^2$ ) in the sense of Neyman.

Both classes of estimation procedures satisfy a certain optimality property, namely, they all yield BAN (best asymptotically normal) estimates. Thus, these estimates,  $\hat{\pi}_{ij}$ , are

- (i) Consistent and hence asymptotically unbiased (i.e., they converge in probability to the  $\pi_{ij}$ );
- (ii) Asymptotically normal as  $N = \sum_{i=1}^r n_{i0} \rightarrow \infty$  with  $\frac{n_{i0}}{N} = \xi_i$  where  $\xi_i$  is a constant;
- (iii) Efficient (i.e., the variance of any other consistent, asymptotically normal estimate is at least as large as the variance of  $\hat{\pi}_{ij}$ );
- (iv) Sufficiently regular (i.e.,  $\frac{\partial \hat{\pi}_{ij}}{\partial \pi_{ij}}$  exist and are continuous in  $\pi_{ij}$  for all  $i, j$ ).

Historically, maximum likelihood estimates (MLE's) were probably the first estimates derived for (1.3.1.1) and have certainly been the most popular to date. To obtain MLE's of the  $\pi_{ij}$ , (1.2.3.1) must be maximized with respect to the  $\pi_{ij}$ 's and subject to (1.3.1.1) leading to a system of simultaneous equations which are often non-linear in the  $\pi_{ij}$  and consequently difficult to solve. Although various iterative schemes have been proposed (e.g., Roy and Kastenbaum, 1956; Kastenbaum and Lamphiear, 1959), the variety of possible hypotheses has precluded the formulation of a general solution of the resulting systems of

equations. However, due to the results of Birch (1963), fairly general maximum likelihood procedures have recently appeared. These include the various iterative proportional fitting schemes presented in Mosteller (1968), Goodman (1968), Bishop and Fienberg (1969), and Bishop (1969), and are based on the realization that "the MLE's of the expected cell frequencies are uniquely determined by the marginal totals being equal to the MLE's of their expectations."

Alternatively, Kullback and his associates have recommended the use of minimum discrimination information estimates. These BAN estimates are obtained by minimizing, with respect to the  $\pi_{ij}$ 's,

$$I(\underline{\pi}; \underline{\tilde{p}}) = \sum_{i=1}^r \sum_{j=1}^s \pi_{ij} \ln\left(\frac{\pi_{ij}}{\tilde{p}_{ij}}\right) \quad (1.3.2.1)$$

subject to (1.3.1.1) where the  $\tilde{p}_{ij}$  are fixed by hypothesis, observed, or estimated. Similar to maximum likelihood estimation, this procedure also usually requires iterative procedures.

Minimum  $\chi^2$  estimates are obtained by minimizing

$$\sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{i0} \pi_{ij})^2}{n_{i0} \pi_{ij}} \quad (1.3.2.2)$$

with respect to the  $\pi_{ij}$ 's and subject to (1.3.1.1). Again, this technique usually involves the solution of a system of complicated, non-linear simultaneous equations, and, since minimum  $\chi^2$  estimates are not known to possess any desirable properties that MLE's lack, these estimates have seldomly been used.

If the denominator of (1.3.2.2) is replaced by the observed cell frequency,  $n_{ij}$ , and the resulting sum

$$\sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{i0} \pi_{ij})^2}{n_{ij}} \quad (1.3.2.3)$$

minimized with respect to the  $\pi_{ij}$ 's and subject to (1.3.1.1), the resulting minimum  $\chi_1^2$  estimates are obtained by solving only linear equations provided the  $f_k(\underline{\pi})$  are linear in the  $\pi_{ij}$ 's. Neyman (1949) proved that, if the  $f_k(\underline{\pi})$  are not linear in the  $\pi_{ij}$ 's, the  $f_k(\underline{\pi})$  may be replaced by their Taylor series expansion about the point,  $\underline{\pi} = \underline{p}$ , where  $p_{ij} = n_{ij}/n_{i0}$ ,  $i = 1, \dots, r$ ;  $j = 1, \dots, s$ , namely

$$f_k^*(\underline{\pi}) = f_k(\underline{p}) + \sum_{i=1}^r \sum_{j=1}^s \left( \frac{\partial f_k(\underline{\pi})}{\partial \pi_{ij}} \right)_{\underline{\pi}=\underline{p}} \cdot (\pi_{ij} - p_{ij}) \quad (1.3.2.4)$$

$$k = 1, 2, \dots, t$$

thereby reducing the problem to the linear case. In either case, the resulting minimum  $\chi_1^2$  estimates are BAN estimates and do not require iterative procedures for their solution.

In the framework of (1.3.1.2) where a linear model is fitted to the data, Grizzle et al. (1969) estimate  $\underline{f}(\underline{\pi})$  by replacing  $\underline{\pi}$  by its unrestricted MLE,  $\underline{p}$ , and then estimate the parameters ( $\underline{\beta}$ ) of the model by ordinary weighted least squares procedures applied to

$$\frac{\underline{f}}{tx1}(\underline{p}) = X \frac{\underline{\beta}}{txu \ ux1} \quad (1.3.2.5)$$

This procedure yields the familiar weighted least squares estimate (WLSE),  $\underline{b}$ , of  $\underline{\beta}$ , namely

$$\underline{b}_{ux1} = (X'S^{-1}X)^{-1} X'S^{-1} \underline{f}(\underline{p}) \quad (1.3.2.6)$$

where

$$S_{txt} = HV(\underline{p})H' = \text{sample estimate of var}(\underline{f}(\underline{p}))$$

with

$$V(\underline{p})_{rsxrs} = \text{sample estimate of var}(\underline{p})$$

$$H_{txrs} = \left( \frac{\partial f_k(\underline{\pi})}{\partial \pi_{ij}} \right)_{\underline{\pi}=\underline{p}}$$

### 1.3.3 Test Statistics

A number of asymptotically equivalent test statistics have been proposed for testing the hypotheses specified by (1.3.1.1) and (1.3.1.2). Asymptotic equivalence requires that, regardless of whether the null hypothesis holds, the probability of any two tests being

contradictory tends to zero as  $N = \sum_{i=1}^r n_{i0} \rightarrow \infty$  with  $\frac{n_{i0}}{N} = \xi_i$  where  $\xi_i$

is a constant. These test statistics for (1.3.1.1) fall into three essentially different classes. They are

(1) Pearson's chi-square statistic

$$X_P^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{i0} \hat{\pi}_{ij})^2}{n_{i0} \hat{\pi}_{ij}} \quad (1.3.3.1)$$

(2) (a) Neyman-Pearson's likelihood ratio chi-square statistic

$$X_L^2 = \sum_{i=1}^r \sum_{j=1}^s \left[ -2 \ln \left( \frac{n_{i0} \hat{\pi}_{ij}}{n_{ij}} \right) \right] n_{ij} . \quad (1.3.3.2)$$

(b) Minimum discrimination information statistic

$$X_I^2 = \sum_{i=1}^r \sum_{j=1}^s \left[ 2N \hat{\pi}_{ij} \ln \left( \frac{\hat{\pi}_{ij}}{p_{ij}} \right) \right] . \quad (1.3.3.3)$$

(3) (a) Neyman's chi-square statistic

$$X_N^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n_{i0} \hat{\pi}_{ij})^2}{n_{ij}} . \quad (1.3.3.4)$$

(b) Wald's statistic

$$X_W^2 = (\underline{f}(\underline{p}))' (H\underline{V}(\underline{p})H')^{-1} (\underline{f}(\underline{p})) \quad (1.3.3.5)$$

where  $\underline{p}$ ,  $\underline{f}(\underline{p})$ ,  $\underline{V}(\underline{p})$ ,  $H$ ,  $\tilde{\underline{p}}$ , and  $N$  are defined in Section 1.3.2, the  $\hat{\pi}_{ij}$  are any BAN estimates of the  $\pi_{ij}$ , and the  $\hat{\pi}_{ij}$  are the minimum discrimination information estimates obtained by minimizing (1.3.2.1).

For testing the fit of the linear model (1.3.1.2), the usual analysis of variance error sum of squares is used, i.e.,

$$X_F^2 = SS(\underline{f}(\underline{\pi}) = X\underline{\beta}) = (\underline{f}(\underline{p}))' S^{-1} \underline{f}(\underline{p}) - (X\underline{b})' S^{-1} (X\underline{b}) \quad (1.3.3.6)$$

where  $S$  and  $\underline{b}$  are defined in Section 1.3.2. Then, given that the model adequately fits the data, tests involving contrasts of the model

parameters  $(\underline{C} \quad \underline{\beta} = \underline{0})$  are produced by usual analysis of variance

sums of squares, i.e.,

$$X_C^2 = SS(\underline{C}\underline{\beta} = \underline{0}) = (\underline{C}\underline{b})' [C(X'S^{-1}X)^{-1}C']^{-1} (\underline{C}\underline{b}) \quad (1.3.3.7)$$

where  $C$  is a matrix of constants of rank  $c \leq u$ .

Under  $H_{01}$ , Neyman (1949) has shown that  $X_P^2$ ,  $X_{LR}^2$  and  $X_N^2$ , using any BAN estimates of the  $\pi_{ij}$ , are asymptotically  $\chi^2$  variates with  $t$  degrees of freedom (D.F.) and hence are asymptotically equivalent provided that

(i)  $\frac{n_{i0}}{N}$ ,  $i = 1, 2, \dots, r$ , remain constant as  $N \rightarrow \infty$ ,

(ii)  $f_k(\underline{\pi}) = 0$ ,  $k = 1, 2, \dots, t$ , has at least one solution such that  $\pi_{ij} > 0$  for all  $i, j$ .

Bhappkar (1966) has shown that  $X_W^2$  (Wald's statistic) for testing linear hypotheses in categorical data is algebraically identical to  $X_N^2$  whenever  $X_N^2$  is defined, i.e., whenever all the  $n_{ij}$  are positive. Similarly, for testing non-linear hypotheses,  $X_W^2$  is identical to  $X_N^2$  using Neyman's linearization technique on the hypothesis constraints,  $\underline{f}(\underline{\pi}) = \underline{0}$ .

Kullback (1959) has shown that  $X_I^2$  (the minimum discrimination information statistic), under  $H_{01}$ , is also asymptotically  $\chi^2$  with D.F. =  $t$  so that all of these tests used for  $H_{01}$  are asymptotically equivalent when using the appropriate estimates of the individual cell probabilities.

In the context of the linear model as in (1.3.1.2), the tests are derived by conventional methods of weighted least squares and hence are the same as Neyman's chi-square tests when translated into constraints.

Under  $H_{02}$ ,  $X_F^2$  for testing the fit of the model is asymptotically  $\chi^2$

with D.F. =  $t-u$ , and, given the model,  $X_C^2$  for testing contrasts of the model parameters is asymptotically  $\chi^2$  with D.F. =  $c$  under the null hypothesis,  $C\beta = \underline{0}$ .

#### 1.4 The General Linear Model

##### 1.4.1 Notation and Assumptions

For a large portion of this research, unless stated otherwise, the underlying probability model is that defined by (1.2.3.1) with the notation (see Grizzle et al., 1969) referring to the expected cell probabilities and hypothetical data shown in Table 1.4.1.1.

Table 1.4.1.1 Expected cell probabilities (cell frequencies) for the standard contingency table

Populations (factors)	Response Categories				Totals
	1	2	...	s	
1	$\pi_{11}$ ( $n_{11}$ )	$\pi_{12}$ ( $n_{12}$ )	...	$\pi_{1s}$ ( $n_{1s}$ )	1 ( $n_{1o}$ )
2	$\pi_{21}$ ( $n_{21}$ )	$\pi_{22}$ ( $n_{22}$ )	...	$\pi_{2s}$ ( $n_{2s}$ )	1 ( $n_{2o}$ )
.	.	.	...	.	.
.	.	.	...	.	.
.	.	.	...	.	.
r	$\pi_{r1}$ ( $n_{r1}$ )	$\pi_{r2}$ ( $n_{r2}$ )	...	$\pi_{rs}$ ( $n_{rs}$ )	1 ( $n_{ro}$ )

Let

$$\begin{aligned} \underline{\pi}'_{1 \times r s} &= (\pi_{11}, \pi_{12}, \dots, \pi_{1s}; \pi_{21}, \pi_{22}, \dots, \pi_{2s}; \dots; \pi_{r1}, \pi_{r2}, \dots, \pi_{rs}) \\ &= (\underline{\pi}'_1, \underline{\pi}'_2, \dots, \underline{\pi}'_r) \end{aligned} \quad (1.4.1.1)$$

$$\begin{aligned} \underline{p}'_{1 \times r s} &= \underline{\pi}'_{\underline{\pi}=\underline{p}} = \text{unrestricted MLE of } \underline{\pi}', \text{ where } p_{ij} = \frac{n_{ij}}{n_{i0}} \quad (1.4.1.2) \\ & \quad i = 1, 2, \dots, r; j = 1, 2, \dots, s \end{aligned}$$

$$\left( \underline{f}_{1 \times t}(\underline{\pi}) \right)' = (f_1(\underline{\pi}), f_2(\underline{\pi}), \dots, f_t(\underline{\pi})) \quad (1.4.1.3)$$

where

$$\begin{aligned} f_k(\underline{\pi}) &= \text{any function of } \underline{\pi} \text{ having first and second partial} \\ & \quad \text{derivatives with respect to the } \pi_{ij} \text{ where} \\ & \quad k = 1, 2, \dots, t \leq r(s-1) \end{aligned}$$

$$\underline{f}'_{1 \times t} = (\underline{f}(\underline{p}))' = (\underline{f}(\underline{\pi}))'_{\underline{\pi}=\underline{p}}$$

$$\begin{aligned} V(\underline{\pi}_i) = \text{var}(\underline{p}_i) &= \frac{1}{n_{i0}} \begin{bmatrix} \pi_{i1}(1-\pi_{i1}) & -\pi_{i1}\pi_{i2} & \dots & -\pi_{i1}\pi_{is} \\ & \pi_{i2}(1-\pi_{i2}) & \dots & -\pi_{i2}\pi_{is} \\ & & \ddots & \\ & & & \pi_{is}(1-\pi_{is}) \end{bmatrix} \\ & \quad \text{(symmetric)} \end{aligned} \quad (1.4.1.4)$$

$$\hat{V}_{i \text{ sxs}} = V(\underline{p}_i) = V(\underline{\pi}_i)_{\underline{\pi}_i = \underline{p}_i} = \text{sample estimate of } V(\underline{\pi}_i) \quad (1.4.1.5)$$

$$\hat{V}_{rs \times rs} = V(\underline{p}) = \text{block diagonal matrix with diagonal blocks} \quad (1.4.1.6)$$

of the form  $\frac{1}{n_{io}} (D_{\underline{p}_i} - \underline{p}_i \underline{p}_i')$

where  $D_{\underline{p}_i}$  = diagonal  $(p_{i1}, p_{i2}, \dots, p_{is})$

$$H_{txrs} = H(\underline{p}) = \left( \frac{\partial f_k(\underline{\pi})}{\partial \pi_{ij}} \right)_{\underline{\pi}=\underline{p}} \quad (1.4.1.7)$$

$$S_{txt} = \hat{H}VH' = \text{sample estimate of var}(\underline{f}) \quad (1.4.1.8)$$

It is assumed that the  $f_k(\underline{\pi})$  are functionally independent and

independent of the constraint,  $\sum_{j=1}^s \pi_{ij} = 1$ , for all  $i$ , and hence  $H$  and

$\hat{H}VH'$  are of full rank. If some of the  $n_{ij} = 0$ , they will be replaced by  $1/s$  so that  $S$  will be of full rank.

#### 1.4.2 Methods of Inference

Assume that  $\underline{f}(\underline{\pi}) = X\underline{\beta}$  where the  $f_k$ 's are possibly non-linear and the rank of  $X$  is  $u$ . For the no factor, multi-response case (i.e., one population problems), the relevant hypotheses are those of various types of independence and are tested using

$$X_I^2 = SS(\underline{f}(\underline{\pi}) = \underline{0}) = \underline{f}' S^{-1} \underline{f} \quad (1.4.2.1)$$

which, under  $H_0: \underline{f}(\underline{\pi}) = \underline{0}$ , is asymptotically  $\chi^2$  with D.F. =  $t$ . If  $\underline{f}(\underline{\pi})$  is linear, i.e.,  $\underline{f}(\underline{\pi}) = A\underline{\pi}$  where

$$A = \begin{bmatrix} a_{(1)11} \cdots a_{(1)1s}; a_{(1)21} \cdots a_{(1)2s}; \cdots; a_{(1)r1} \cdots a_{(1)rs} \\ a_{(2)11} \cdots a_{(2)1s}; a_{(2)21} \cdots a_{(2)2s}; \cdots; a_{(2)r1} \cdots a_{(2)rs} \\ \vdots \\ a_{(t)11} \cdots a_{(t)1s}; a_{(t)21} \cdots a_{(t)2s}; \cdots; a_{(t)r1} \cdots a_{(t)rs} \end{bmatrix}$$

$$= \begin{bmatrix} a'_{(1)} \\ a'_{(2)} \\ \vdots \\ a'_{(t)} \end{bmatrix}$$

is a matrix of arbitrary constants of rank  $t$ , then (1.4.2.1) is more explicitly given by

$$X_1^2 = SS(\underline{f}(\underline{\pi}) = A\underline{\pi} = \underline{0}) = (A\underline{p})' (\hat{A}VA')^{-1} (A\underline{p}) \quad (1.4.2.2)$$

A second class of functions often arising in categorical data problems is the class consisting of the logarithmic functions which can be expressed by  $\underline{f}(\underline{\pi}) = K \underline{\ln}(A\underline{\pi})$  where  $\underline{\ln}(A\underline{\pi})$  denotes the vector of natural (or Naperian) logarithms of the elements of  $A\underline{\pi}$ , and  $K = (k_{\alpha\gamma})_{v \times t}$  is a matrix of arbitrary constants of rank  $v \leq t$ . The appropriate test statistic for this case is given by

$$\begin{aligned}
 X_L^2 &= SS(\underline{f}(\underline{\pi}) = K \underline{\ln}(A\underline{\pi}) = \underline{0}) & (1.4.2.3) \\
 &= (K \underline{\ln}(A\underline{p}))' (KD^{-1}A\hat{V}A'D^{-1}K')^{-1} (K \underline{\ln}(A\underline{p}))
 \end{aligned}$$

where  $D = \text{diagonal } (\underline{a}'_{(1)} \cdot \underline{p}, \underline{a}'_{(2)} \cdot \underline{p}, \dots, \underline{a}'_{(t)} \cdot \underline{p})$ . Under  $H_0$ ,  $X_L^2$  is asymptotically  $\chi^2$  with D.F. =  $v$ . It should be noted that caution must be taken in constructing  $A$  so that no element of the vector,  $A\underline{p}$ , is zero. Most frequently,  $A = I = (\text{identity matrix})$  corresponding to tests for various types of interaction. Since empty cells are to be replaced by  $1/s$ , each element of  $I\underline{p}$  ( $= \underline{p}$ ) will be positive so that no difficulty is encountered with this common application of the logarithmic functions.

For simplicity, consider, as an example of the preceding, the test of independence for the two-way case cited previously and suppose  $j_1 = 1, 2; j_2 = 1, 2$ . Then

$$H_0: \underline{f}(\underline{\pi}) = K \underline{\ln}(A\underline{\pi}) = \underline{0}$$

is given by

$$\begin{aligned}
 \underline{f}(\underline{\pi}) &= \underline{\ln} \left( \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} \right) \\
 &= \underline{\ln} \pi_{11} - \underline{\ln} \pi_{12} - \underline{\ln} \pi_{21} + \underline{\ln} \pi_{22} = 0
 \end{aligned}$$

so that

$$A = I_4 = \text{identity matrix}$$

$$K = (1, -1, -1, 1)$$

$$\hat{V} = V(\underline{p}) = \frac{1}{N} \begin{bmatrix} p_{11}(1-p_{11}) & -p_{11}p_{12} & -p_{11}p_{21} & -p_{11}p_{22} \\ & p_{12}(1-p_{12}) & -p_{12}p_{21} & -p_{12}p_{22} \\ & & p_{21}(1-p_{21}) & -p_{21}p_{22} \\ \text{(symmetric)} & & & p_{22}(1-p_{22}) \end{bmatrix}$$

$$\text{where } N = \sum_{j_1=1}^2 \sum_{j_2=1}^2 n_{j_1 j_2}$$

$$D = \text{diagonal } (p_{11}, p_{12}, p_{21}, p_{22})$$

and hence

$$KD^{-1}A = \left( \frac{1}{p_{11}}, -\frac{1}{p_{12}}, -\frac{1}{p_{21}}, \frac{1}{p_{22}} \right)$$

$$KD^{-1}A\hat{V} = \frac{1}{N} (1, -1, -1, 1)$$

$$KD^{-1}A\hat{V}A'D^{-1}K' = \frac{1}{N} \left( \frac{1}{p_{11}} + \frac{1}{p_{12}} + \frac{1}{p_{21}} + \frac{1}{p_{22}} \right)$$

and the test statistic with D.F. = 1 is given by the familiar

$$X_L^2 = SS(\underline{\hat{f}}(\underline{\pi}) = K \underline{\ell n}(A\underline{\pi}) = \underline{0})$$

$$= \frac{N(\ell n p_{11} - \ell n p_{12} - \ell n p_{21} + \ell n p_{22})^2}{\frac{1}{p_{11}} + \frac{1}{p_{12}} + \frac{1}{p_{21}} + \frac{1}{p_{22}}}$$

For the uni- or multi-factor, uni- or multi-response case (i.e., several population problems),  $X$  is known and of rank  $u \leq t$ . A test of fit of the model  $\underline{f}(\pi) = X\underline{\beta}$ , is given by the residual sum of squares

$$X_F^2 = SS(\underline{f}(\pi) = X\underline{\beta}) = \underline{f}'S^{-1}\underline{f} - (\underline{Xb})'S^{-1}(\underline{Xb}) \quad (1.4.2.4)$$

where  $\underline{b}$  is the vector that minimizes  $(\underline{f} - \underline{Xb})'S^{-1}(\underline{f} - \underline{Xb})$  and is given by

$$\underline{b} = (X'S^{-1}X)^{-1}X'S^{-1}\underline{f}. \quad (1.4.2.5)$$

If the model fits the data,  $X_F^2$  is asymptotically  $\chi^2$  with D.F. =  $t - u$ . Recalling that for the linear case,  $\underline{f} = A\underline{p}$ ,  $S = A\hat{V}A'$ , and for the logarithmic case,  $\underline{f} = K \ln(A\underline{p})$ ,  $S = KD^{-1}A\hat{V}A'D^{-1}K'$ , explicit expressions for the test statistic given in (1.4.2.4) can easily be obtained for these two classes of functions.

Given that the model fits, various hypotheses concerning constraints on the model parameters,  $\underline{\beta}$ , may be of interest as, for example, significance of differential hospital effects and differential surgical procedure effects relating to severity of the dumping syndrome for the example in Grizzle et al. (1969). A test of the hypothesis,  $H_0: C\underline{\beta} = \underline{0}$ , is produced by

$$X_c^2 = SS(C\underline{\beta} = \underline{0}) = (\underline{Cb})'[C(X'S^{-1}X)^{-1}C']^{-1}(\underline{Cb}) \quad (1.4.2.6)$$

where  $C$  is a  $(cxu)$  matrix of arbitrary constants of rank  $c \leq u$ . Under  $H_0$ ,  $X_c^2$  is asymptotically  $\chi^2$  with D.F. =  $c$ . Again, explicit expressions for (1.4.2.6) can easily be obtained for the linear and logarithmic classes of functions.

The two factor, one response example in Grizzle et al. (1969) involving severity of the dumping syndrome provides a good illustration of the linear model treatment of the several population problems and thus will not be repeated here.

## 2. MIXED CATEGORICAL DATA MODELS

### 2.1 Introduction

In considering two-way contingency tables, two types of models have been identified (see Bhapkar and Koch, 1968a): the "no factor, two response" situation in which neither margin is fixed and the "one factor, one response" situation in which exactly one margin is fixed. For the first case, the hypothesis of primary interest is one of independence or no association between the responses; in the second case, the hypothesis of primary interest is one of homogeneity over the factor levels.

However, there exists a third type of model which will be called a "mixed categorical data model of order 2" for the case of two-way tables. The experimental situation corresponding to such a model involves exposing each of  $n$  randomly chosen subjects from some homogeneous population to both levels of a binary factor (e.g., control vs treated) and classifying each of the two responses into one of  $r$  categories. The resulting data, which will be assumed to follow a multinomial distribution, can then be represented by an  $r \times r$  contingency table. Thus, this is a categorical data version of the classical matched pairs design. As such, it has been studied in a somewhat different context recently by Miettinen (1968, 1969, 1970), among others. In addition, the methods of Mantel and Haenszel (1959) and Mantel (1963) are also of interest for this experimental situation. Although these authors deal with appropriate statistical procedures for the analysis of certain types of mixed categorical data models of order 2, they do not indicate the way in which such contingency table data can be analyzed as a special case of a unified general approach.

One such methodology which will be the basis of inference in this research is that of Grizzle et al. (1969). As a result, test statistics are derived through weighted least squares analysis of certain appropriately formulated linear models as indicated in Section 1.4.2. Alternative methods are also appropriate in certain contexts, e.g., those of Lewis (1968) and Goodman (1970) which are based on maximum likelihood and that of Ku and Kullback (1968) based on minimum discrimination information.

To apply the methods of Grizzle, Starmer, Koch (GSK), it is first necessary to recognize the underlying factor-response structure for the data. Since only the sample size  $n$  is fixed, this is a special case of the "no factor, two response" model. However, the hypothesis of interest is one of homogeneity over the factor levels; i.e., there is no difference between the effects of control and treated. This hypothesis can be formulated specifically as an hypothesis of marginal symmetry in the two-way table.

One of the principal purposes of this research is to demonstrate that tests of marginal symmetry are not only applicable to mixed categorical data models, but are the only tests directed at the question of primary interest associated with such experimental situations. Responses from the same subject (or units in the same matched pair) to different factor levels are obviously associated since they have the same initial source in common. Thus, the traditional hypothesis of independence which is usually applied to "no factor, two response" tables is not of much interest here.

Finally, it is important to recognize that if the  $r$  categories can be described in terms of some quantitative scale, the question of principal interest is one of equality of mean scores over the two marginals. This hypothesis bears a definite resemblance to the relationship tested by the classical matched pairs  $t$ -test used with continuous data.

## 2.2 Mixed Models in General

For mixed model experiments in the continuous case (cf. Scheffé, 1959), each of  $n$  randomly selected subjects responds to each of  $d$  treatments exactly once. The outcome of the experiment constitutes an observation matrix composed of  $n$  response vectors, each of length  $d$ . The hypothesis of interest in mixed models is the equality of treatment effects (or homogeneity over the factor levels).

The following basic assumptions are made (see Koch and Sen, 1968):

- A.1 The  $n$  response vectors are mutually independent.
- A.2 Each response vector has a continuous  $d$ -variate c.d.f. (cumulative distribution function).
- A.3 For any subject, the joint distribution of any linearly independent set of contrasts among the  $d$  observations is diagonally symmetric.

For the normal (or parametric) case, an additional assumption is required, namely,

- A.4 The  $n$  error vectors have a common  $d$ -variate normal distribution,  $N_d(\underline{0}, \Sigma)$ .

If it can be assumed, in addition, that

- A.5 The subject effects are additive, i.e., the joint distribution of the  $i$ -th subject effect and  $\bar{i}$ -th error vector is the same for all subjects,

the appropriate test procedure is based on Hotelling's  $T^2$  (cf. Case III of Koch and Sen). If furthermore, it can be assumed that

A.6 There is compound symmetry of the error vectors (e.g.,

$$\Sigma = \rho\sigma^2J + (1-\rho)\sigma^2I$$

where J is a (d x d) matrix of 1's and I is the (d x d) identity matrix),

the usual analysis of variance F is appropriate (Case IV of Koch and Sen). Case I (which requires A.1 - A.4 only) and Case II (which requires A.1 - A.4 and A.6) are not of interest in the parametric case. Non-parametric analogues of these cases are given in Koch and Sen (1968).

The mixed models primarily considered in this research may be viewed as belonging to the categorical data version of Case III for which the Hotelling  $T^2$ -statistic is applicable in the continuous case with multivariate normal responses. As it is assumed in Case III that there are no subject effects, it is appropriate to assume that a single multinomial distribution characterizes the entire contingency table.

If assumptions A.5 and A.6 are not tenable in the categorical data framework (Case II), it is useful to adopt a univariate point of view as in the matched case-control situations considered by Mantel and Haenszel (1959) and Mantel (1963). The principal difference with Case III is that here the data are represented by n two-way contingency tables with d rows and r columns where the response of each subject to each treatment is classified making the total frequency nd rather than n.

Cases I and IV are of rather limited interest in the categorical data framework. Table 2.2.1 summarizes the three most important cases along with the appropriate test procedures.

Table 2.2.1 Summary of assumptions and appropriate test procedures for various types of mixed models

Case	Assumptions (A.4 for Parametric Tests)	Continuous Data		Categorical Data Tests
		Parametric Tests	Non-Parametric Tests	
II <sup>a</sup>	A.1 - A.3 A.6		Generalized Friedman's $\chi^2$ test	Mantel-Haenszel test (case-control) Tests of Mantel (matched sets)
III	A.1 - A.3, A.5	Test based on Hotelling $T^2$	d-variate signed rank Wilcoxon test	Linear model tests of mixed categor- ical data models
IV	A.1 - A.3, A.5, A.6	Analysis of variance F-test	Sen's aligned block rank test	Classical $\chi^2$ test of homogeneity

<sup>a</sup>Case I is omitted since it is of little interest here.

The extension to generalized mixed models will be considered at a later time. With this extension, analyses of "matched set" experiments (e.g., the study reported in Miettinen (1969) where each propositus was matched with four controls) will readily be performed using the basic GSK approach. The present discussion is limited to more standard situations.

### 2.3 Mixed Categorical Data Models of Order 2

First let us consider the following data (see Table 2.3.1) from case records of the eye-testing of 7477 employees in Royal Ordnance factories which have been used as an illustrative example for tests of marginal symmetry by Stuart (1955), Bhapkar (1966), Grizzle et al.

Table 2.3.1 Vision records for women employees of the Royal Ordnance factories

Right Eye Grade	Left Eye Grade				Total
	Highest (1)	Second (2)	Third (3)	Lowest (4)	
Highest (1)	1520	266	124	66	1976
Second (2)	234	1512	432	78	2256
Third (3)	117	362	1772	205	2456
Lowest (4)	36	82	179	492	789
Total	1907	2222	2507	841	7477

(1969), and Ireland et al. (1969). These data may be viewed as arising from a mixed model in which each of  $n = 7477$  subjects is classified according to both levels of a factor (eye position) with the corresponding cell frequencies being assumed to follow a multinomial distribution. As such, the hypothesis of principal interest is that the distributions of grades of vision are the same for both eyes, i.e., marginal symmetry. The values of the test statistics (each approximately distributed as  $\chi^2$  with D.F. = 3 under the null hypothesis obtained by the preceding authors are approximately equal (namely, 11.96 (Stuart), 11.97 (Bhaskar), 11.98 (Grizzle et al.), and 12.00 (Ireland et al.)). Thus, the marginal distributions of eye grades are significantly different.

Looking at the problem of marginal symmetry in a more formal way, let  $\pi_{jj'}$  denote the probability that a subject is classified according to the  $j$ -th grade category on the right eye and the  $j'$ -th grade

category on the left eye. Then the hypothesis of marginal symmetry may be written as

$$\pi_{j0} = \pi_{0j'} \quad j = j' = 1, 2, 3, 4 \quad (2.3.1)$$

where  $\pi_{j0} = \sum_{j'=1}^4 \pi_{jj'}$  and  $\pi_{0j'} = \sum_{j=1}^4 \pi_{jj'}$ . Following the GSK

approach, form the unrestricted maximum likelihood estimates

$p_{jj'} = (n_{jj'}/n)$  of  $\pi_{jj'}$ , where  $n_{jj'}$  represents the number of subjects in the  $jj'$ -th cell. From these, the desired estimates of  $\pi_{j0}$  and  $\pi_{0j'}$  are obtained by taking the appropriate marginal sums of  $p_{jj'}$ , namely,

$p_{j0} = \sum_{j'=1}^4 p_{jj'}$  and  $p_{0j'} = \sum_{j=1}^4 p_{jj'}$ , respectively. If the following

linear model is fitted to the  $p_{j0}$  and  $p_{0j'}$

$$E \begin{bmatrix} p_{10} \\ p_{20} \\ p_{30} \\ p_{01} \\ p_{02} \\ p_{03} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \quad (2.3.2)$$

by weighted least squares, the residual sum of squares ( $X_{MS}^2$ ) with D.F. = 3 is the GSK test statistic for (2.3.1) which is the same as that of Bhapkar (1966) (except for round-off error).

Alternatively, if scores  $a_1, a_2, a_3, a_4$  are assigned to the response categories, the hypothesis of equality of mean scores over the two marginals may be written as

$$H_0: \alpha_{RE} \equiv \sum_{j=1}^4 a_j \pi_{j0} = \sum_{j'=1}^4 a_{j'} \pi_{0j'} \equiv \alpha_{LE} . \quad (2.3.3)$$

One test statistic for this hypothesis is the residual sum of squares,  $X_M^2$ , with D.F. = 1, which is obtained by forming  $\hat{\alpha}_{RE} = \sum_{j=1}^4 a_j p_{j0}$

and  $\hat{\alpha}_{LE} = \sum_{j'=1}^4 a_{j'} p_{0j'}$  and fitting the linear model

$$E \begin{bmatrix} \hat{\alpha}_{RE} \\ \hat{\alpha}_{LE} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \mu$$

by weighted least squares. For the data in Table 1, the scores  $a_1 = 1, a_2 = 2, a_3 = 3, a_4 = 4$  have been assigned to the grade categories for both eyes. The residual sum of squares is then given by

$$\begin{aligned} X_M^2 &= \frac{n(\hat{\alpha}_{RE} - \hat{\alpha}_{LE})^2}{\sum_{j=1}^4 \sum_{j'=1}^4 [(a_j - a_{j'}) - (\hat{\alpha}_{RE} - \hat{\alpha}_{LE})]^2 p_{jj'}} & (2.3.4) \\ &= \frac{7477 \left( \frac{17012}{7477} - \frac{17236}{7477} \right)^2}{\left[ 1 \left( \frac{1678}{7477} \right) + 4 \left( \frac{401}{7477} \right) + 9 \left( \frac{102}{7477} \right) \right] - \left( \frac{17012}{7477} - \frac{17236}{7477} \right)^2} \\ &= 11.97 . \end{aligned}$$

The similarity of the test given in (2.3.4) to the matched pairs t-test may be seen clearly by writing down the frequency distribution of the difference ( $d_{jj'} = a_j - a_{j'}$ ) between the right eye grade and the left eye grade (see Table 2.3.2). The average difference,  $\bar{d}$ , is  $(-244/7477) \approx -.03$  and its estimated standard error,  $s_{\bar{d}}$ , is 0.0087.

Hence

$$t^2 = (\bar{d} / s_{\bar{d}})^2 = (-3.46)^2 = 11.97 \quad (2.3.5)$$

which is identical to the value obtained for  $X_M^2$ .

Table 2.3.2 Frequency distribution of the difference between right and left eye grades

Difference ( $d_{jj'}$ )	Frequency ( $f_{jj'}$ )
-3	66
-2	202
-1	903
0	5296
1	775
2	199
3	36

#### 2.4 Mixed Categorical Data Models of Higher Orders

In general a "mixed categorical data model of order  $d$ " involves exposing each of  $n$  randomly chosen subjects from some homogeneous population to each level of a factor with  $d$  levels and classifying each response into one of  $r$  categories. The resulting data, which will be assumed to follow a multinomial distribution, are then represented by an  $r \times r \times \dots \times r$  contingency table of  $d$  dimensions. This, then, is the categorical data version of the classical mixed model with  $n$  subjects and  $d$  treatments (see Section 2.2).

For the case of higher order mixed categorical data models with binary response levels ( $r=2$ ), Bhapkar (1965) has formulated a test in terms of  $d$ -th order marginal symmetry. This procedure is illustrated in Grizzle et al. (1969) for data indicating the responses of 46 subjects to each of the drugs A, B, and C as either being favorable or not favorable (see Table 2.4.1). Thus, this example is a mixed categorical data model of order  $d = 3$  with drug type being the factor with levels A, B, and C. The hypothesis of interest is equality of the

Table 2.4.1 Responses of patients to drugs A, B, and C

Response to A		Favorable			Not Favorable			Total
Response to B		Fav.	Not Fav.	Total	Fav.	Not Fav.	Total	
Response to C	Fav.	6	2	8	2	6	8	16
	Not Fav.	16	4	20	4	6	10	30
	Total	22	6	28	6	12	18	46

probability of a favorable response for each of the three drugs; i.e., equality of the corresponding marginal distributions or marginal symmetry of the three one-way margins. For the data in Table 2.4.1, the GSK approach yields  $X_{MS}^2 = 6.58$  with D.F. = 2 indicating a significant difference ( $\alpha = .05$ ) between drugs.

In order to extend the theory given in the previous section to the general case, let  $\pi_{j_1 j_2 \dots j_d}$  denote the probability of cell  $(j_1, j_2, \dots, j_d)$  where  $j_\xi = 1, 2, \dots, r$  with  $\xi = 1, 2, \dots, d$  in a  $d$ -dimensional table. Now, the hypothesis of total symmetry is only of limited practical interest since it is the first order marginal distributions (and the corresponding tests of marginal symmetry) that provide the clearest indication of the relative effects of the various treatments or factor levels. Thus, if  $\phi_{\xi j}$  represents the marginal probability of level  $j$  for the  $\xi$ -th dimension and is defined as

$$\phi_{\xi j} = \sum_{\substack{v=1 \\ v \neq \xi}}^d \sum_{j_v=1}^r \pi_{j_1 j_2 \dots j_d} \quad (2.4.1)$$

the hypothesis of  $d$ -th order marginal symmetry for the one-way margins may be written

$$H_0: \phi_{1j} = \phi_{2j} = \dots = \phi_{dj} \quad (2.4.2)$$

$$j = 1, 2, \dots, r .$$

Let  $n_{j_1 j_2 \dots j_d}$  denote the observed frequency of cell  $(j_1, j_2, \dots, j_d)$ .

An appropriate A-matrix is then constructed which generates the



$$E \begin{bmatrix} p_{11} \\ p_{12} \\ \dots \\ p_{1,r-1} \\ p_{21} \\ p_{22} \\ \dots \\ p_{2,r-1} \\ \dots \\ \dots \\ \dots \\ p_{d1} \\ p_{d2} \\ \dots \\ p_{d,r-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & 1 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & 1 \\ & & \dots & \\ & & \dots & \\ & & \dots & \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_{r-1} \end{bmatrix} \quad (2.4.3)$$

The residual sum of squares,  $X_{MS}^2$ , provides the test statistic for the hypothesis of marginal symmetry in (2.4.2). When (2.4.2) holds,  $X_{MS}^2$  is approximately distributed as  $\chi^2$  with D.F. =  $(d-1)(r-1)$ .

Alternatively, if scores  $a_1, a_2, \dots, a_r$  are assigned to the response categories, the hypothesis of equality of mean scores over the  $d$  marginals may be written as

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_d \quad (2.4.4)$$

where  $\alpha_\xi = \sum_{j=1}^r a_j p_{\xi j}$ ,  $\xi = 1, 2, \dots, d$ . The hypothesis (2.4.4) may be

tested by obtaining the estimates  $\hat{\alpha}_\xi = \sum_{j=1}^r a_j p_{\xi j}$ ,  $\xi = 1, 2, \dots, d$ ,

and fitting the linear model

$$E \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \dots \\ \hat{\alpha}_d \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix} \mu \quad (2.4.5)$$

by weighted least squares. As before the test statistic is the residual sum of squares,  $X_M^2$ , but with D.F. = (d-1).

The tests given here for (2.4.2) and (2.4.4) appear similar to the classical ones of homogeneity in "one factor, one response" tables. However, here the underlying table is d-dimensional and the  $p_{\xi j}$  are correlated with each other while in the "one factor, one response" tables the underlying table is two-dimensional and the  $p_{\xi j}$  are independent of  $p_{\xi' j}$  for all  $\xi' \neq \xi$ . Hence, the two procedures are quite different with respect to the underlying estimates of variance and covariance.

Finally, it should be mentioned that other types of hypotheses of marginal symmetry (e.g., equality of all two-way margins) may be easily tested by this approach. All that is required is choosing an appropriate A-matrix, and then fitting a linear model specific to the hypothesis of interest. However, such tests are not as readily interpreted experimentally in the context of mixed models as are the tests involving one-way margins.

## 2.5 Split Plot Contingency Tables

A somewhat more complex example than those considered previously is provided by the following data (see Table 2.5.1) of Lessler (1962) which have been described in Bhapkar and Koch (1968b). In this case,

Table 2.5.1 Responses of males to culturally masculine and anatomically feminine objects with weak intensity

Response at 1/5 Sec.		M			F			Total	
Response at 1/100 Sec.		M	F	Total	M	F	Total		
A	Response at 1/1000 Sec.	M	171	6	177	7	7	14	191
		F	18	12	30	7	56	63	93
Total			189	18	207	14	63	77	284
C	Response at 1/1000 Sec.	M	184	10	194	7	20	27	221
		F	38	14	52	7	114	121	173
Total			222	24	246	14	134	148	394

there are two groups of male subjects involved in a study of the nature of sexual symbolism in objects: Group A members were not told the purpose of the experiment while Group C subjects were. Each subject was asked to classify as masculine or feminine (M or F) certain objects which had been previously characterized as being culturally masculine and anatomically feminine but with weak intensity at three different exposure rates: 1/5 second, 1/100 second, and 1/1000 second. Hence, this is a "one factor, three response" situation with Group being the factor and the M or F classification at different exposure rates being the responses. However, since it is of interest to compare the responses at different exposure rates, it is apparent that the data from either one of the Groups A or C by themselves is a mixed categorical data model of order 3. Thus, the data for the combined groups may be referred to as a "split plot categorical data" model. Hence, the hypotheses of interest are equality of group (whole plot) effects,

equality of exposure rate (split plot) effects, and "no interaction" between exposure rates and groups. These hypotheses may be readily formulated using the GSK approach. First, define the relative frequency vectors for the two groups as follows:

$$\begin{aligned} \underline{p}_A' &= (171, 18, 6, 12, 7, 7, 7, 56) (1/284) \\ &1 \times 8 \\ \underline{p}_C' &= (184, 38, 10, 14, 7, 7, 20, 114) (1/394) \\ &1 \times 8 \end{aligned} \quad (2.5.1)$$

$$\underline{p}' = (\underline{p}_A', \underline{p}_C') \quad 1 \times 16$$

In order to obtain estimates of the probability of masculine classification at the exposure rates 1/5 second, 1/100 second, 1/1000 second, respectively, in Group A and in Group C, define

$$\underline{A} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \quad (2.5.2)$$

Then fit the following linear model by weighted least squares

$$E(\underline{f}) = \begin{bmatrix} 1 & 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \mu_A \\ \lambda_{A_1} \\ \lambda_{A_2} \\ \mu_C \\ \lambda_{C_1} \\ \lambda_{C_2} \end{bmatrix} \quad (2.5.3)$$

where  $\underline{f} = A\underline{p} = \begin{bmatrix} 0.729 \\ 0.715 \\ 0.673 \\ 0.624 \\ 0.599 \\ 0.561 \end{bmatrix}$

and  $\mu_A$  is a mean effect in Group A,  $\lambda_{A_1}$  is the linear effect of exposure in Group A,  $\lambda_{A_2}$  is the non-linear effect of exposure in Group A, and  $\mu_C$ ,  $\lambda_{C_1}$ ,  $\lambda_{C_2}$  are similarly defined quantities in Group C. The hypothesis of no (group x exposure rate) interaction may be formulated as

$$H_0: \lambda_{A_1} = \lambda_{C_1}, \quad \lambda_{A_2} = \lambda_{C_2} \quad (2.5.4)$$

which may be tested by using the following hypothesis matrix

$$C_{2 \times 6} = \begin{bmatrix} 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \quad (2.5.5)$$

in accordance with the principles of weighted regression analysis. The resulting  $X^2_I = 0.20$  with D.F. = 2 indicates that the data are consistent with this hypothesis of no interaction. Hence, the model may be revised as follows:

$$E(\underline{f}) = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & 0 & 2 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 0 & 2 \\ 1 & -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \gamma \\ \lambda_1 \\ \lambda_2 \end{bmatrix} \quad (2.5.6)$$

where  $\mu$  is an overall mean,  $\gamma$  is the differential group effect,  $\lambda_1$  is the linear effect of exposure, and  $\lambda_2$  is the non-linear effect of exposure. The hypothesis of no group main effects may be formulated as

$$H_0: \gamma = 0 \quad (2.5.7)$$

and tested by using the hypothesis matrix

$$C_{1 \times 4} = [0 \ 1 \ 0 \ 0] \quad (2.5.8)$$

while the hypothesis of no exposure rate main effects may be formulated as

$$H_0: \lambda_1 = \lambda_2 = 0 \quad (2.5.9)$$

and tested by using the hypothesis matrix

$$C_{2 \times 4} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.5.10)$$

The resulting test statistics are  $X_G^2 = 11.57$  with D.F. = 1 and  $X_E^2 = 14.55$  with D.F. = 2, respectively. Hence, it is apparent that there are both significant differences between the two groups and among the three exposure rates. In fact, the differences among the three exposure rates are linear as can be seen by testing separately for the linear and for the non-linear effect of exposure.

Thus, we see that split plot contingency tables can be analyzed by the linear model approach as readily as the more standard types of multi-factor, multi-response experiments treated in Grizzle et al. (1969).

### 3. MORE GENERAL CATEGORICAL DATA MODELS FOR THE CASE OF SUPPLEMENTED MARGINS

#### 3.1 Introduction

As has been indicated previously, the approach of Grizzle et al. (1969) is based upon applying weighted least squares analysis to certain appropriately formulated linear models similar to those arising in univariate and/or multivariate analysis. As such, the following assumptions are made:

- (i) These linear models are completely specified by a single design matrix.
- (ii) The data is complete in the sense that every experimental unit is classified according to each of the  $d$  dimensions of the table.
- (iii) All cells are assumed to have positive probability of occurrence.

In many practical problems, some of these assumptions must be relaxed. Goodman (1968) and Bishop and Fienberg (1969), among others, have considered the implications of having a priori certain empty cells. Their approach utilizes multiplicative models with maximum likelihood estimation. The relaxation of (i) is partially accounted for in the GSK approach by allowing various functions,  $f(\pi)$ , of the cell probabilities. The relaxation of (ii) is the primary concern of this research.

In the usual categorical data situation, there is complete information available for each entry in the corresponding contingency table. For instance, in the two factor (subject group, anatomical meaning), two response (classification at exposure rates of 1/5 second and 1/1000 second) example in Bhapkar and Koch (1969b), each entry has

been classified according to all four dimensions. However, in many situations, the experimenter is not equally interested in the response variates or perhaps additional information on one particular dimension could be obtained relatively easily and economically. Alternatively, perhaps, as described in Kleinbaum (1970) for the continuous case, there is missing data for some individuals in the sense that classifications on only  $d' < d$  dimensions have been made. Retaining and utilizing this partial information is distinctly different from the "apparently" similar problem of empty cells (see Assumption (iii)).

In this research, the additional information is obtained by design rather than by chance and as such is referred to as "supplemental information" and the corresponding margins of the contingency table as "supplemented margins". This supplemental information is used to improve the precision of the estimates of the marginal probabilities over the supplemented dimension(s), and it is also incorporated into the estimation of the individual cell probabilities in order to improve certain tests of hypotheses.

Examples where such supplemental information might arise include questionnaire surveys similar to the recently-completed national census. In such surveys, every respondent would be expected to answer a basic set of questions. In addition, a certain proportion would be asked to respond to a second set of questions judged to be less important or more difficult to answer than the questions in the basic set. Supplemental information would derive from the answers to questions in the basic set from those respondents not requested to consider the second set of questions. Similarly, in automobile crash investigations,

the investigating officer might routinely obtain certain biographical information (e.g., race, sex, age) for all drivers but, due to the difficult conditions of the interview, obtain, only for a specified subset, additional information such as sobriety of the driver, purpose of the trip, etc. In this case, the supplemental information would consist of the answers to those biographical questions for those drivers who were not examined in detail. A number of additional examples are given later in this chapter.

It should be noted here that the use of supplemented margins in the analysis of contingency tables resembles the technique of "double sampling" as described in Cochran (1963). In "double sampling," a preliminary sample (cf. supplemental information) is used to estimate the distribution of an auxiliary variable (cf. marginal distribution of particular response or combination of responses) and a second stratified sample is used to estimate the characteristic of interest (cf. cell probability).

Previous work in this area appears to be limited to but one author. Blumenthal (1968) treats essentially the same problem as considered in this research although he restricts attention to one-way tables where each of the  $I$  main categories has  $J_i$ ,  $i = 1, 2, \dots, I$ , sub-categories of classification. The major difference in the two approaches lies in Blumenthal's assumption that a sample belonging in cell  $(i, j)$  has probability  $\alpha_{ij}$  of being only "partially classified" as opposed to having a secondary sample for which supplementary information on a subset of the study variables is obtained. Imposing certain simplifying assumptions for this one-way situation, Blumenthal obtains

MLE's of the individual cell probabilities as well as the bias and variance of these estimates. He concludes by suggesting the extension to double dichotomies which is treated in the present research (see Chapter 4) from the point of view of "supplemented margins" rather than "partially categorized" data.

It should be noted that supplementation is considered only for margins that are not fixed (i.e., response margins rather than factor margins). Supplementing a factor margin is of no interest since that margin being considered fixed a priori yields complete information about its composition.

In order to see more clearly the implications of using supplemental data in the GSK approach, the recent work of Kleinbaum (1970) for the case of normal random variables will be considered.

### 3.2 More General Linear Models in the Case of Normal Random Variables

In the case of normal multivariate theory where the standard multivariate (SM) linear model is given by

$$E(y) = E(\underline{y}_1, \underline{y}_2, \dots, \underline{y}_p) = X(\underline{\beta}_1, \underline{\beta}_2, \dots, \underline{\beta}_p) = \begin{matrix} X & \beta \\ N \times u & u \times p \end{matrix} \quad (3.2.1)$$

with corresponding variance matrix,  $\Sigma$ , two basic assumptions are made.

These assumptions are that

- (i) Each of the  $p$  response variates is measured for each of the  $N$  experimental units (i.e., each of the  $N$  response vectors is complete).
- (ii) The design matrix,  $X$ , is the same for each of the  $p$  responses (i.e., the same blocking system is used on each response variable).

Examples where certain of these assumptions do not hold are numerous. For example, in psychological testing with batteries of (say) 20 different tests, each subject might take a combination of only 5 tests since taking many tests might tend to bias the later results. Hence condition (i) is violated. (Analogous situations occur in the categorical data framework.) Assumption (ii) is violated in many econometric problems involving sets of different but correlated regressions as in certain time series analyses. Likewise, in certain agricultural experiments, efficiency might dictate the use of various blocking systems on different response variates leading to a number of design matrices for the  $p$  responses.

Various authors, including most recently Afifi and Elashoff (1966, 1967, 1969a, 1969b), have treated the case where one or the other of the above assumptions is violated. The General Incomplete Multivariate (GIM) model pertains to the situation where (i) does not hold. A special case of the GIM model is the Hierarchical Incomplete Multivariate (HIM) model where the missing data has a simple monotonic structure. The Multi-Design Multivariate (MDM) model allows the various response variates to have different design matrices and hence allows for the relaxation of (ii).

Kleinbaum (1970) has allowed for the simultaneous relaxation of these basic assumptions in his development of the More General Linear Multivariate (MGLM) model. The SM linear model as well as the GIM, HIM, and MDM models arise as special cases of the MGLM model. However, since the Growth Curve Multivariate (GCM) model is not a special case of the MGLM model, the adjective "More" rather than "Most" has been used.

The most useful form of the MGLM model is the "vector version" given by

$$E(\underline{z}) = E \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_p \end{bmatrix} = \begin{bmatrix} W_1 & 0 & \dots & 0 \\ & W_2 & & \\ & & \ddots & \\ & & & W_p \\ 0 & & & \dots & 0 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_p \end{bmatrix} = W \underline{\xi} \quad (3.2.2)$$

$N_T \times 1$ 

 $N_T \times M_T$ 
 $M_T \times 1$

with  $\text{var}(\underline{z}) = \Omega$  where  $\underline{z}_\ell$ ,  $\ell = 1, 2, \dots, p$ , is a  $(N_\ell \times 1)$  vector comprising all observations on the  $\ell$ -th response variate in the entire experiment where  $0 < N_\ell \leq N$ ,  $W_\ell$  the corresponding  $(N_\ell \times M_\ell)$  design matrix for  $\underline{z}_\ell$ , and  $\xi_\ell$  the  $(M_\ell \times 1)$  vector of parameters corresponding to  $\underline{z}_\ell$  where  $N_T = \sum_{k=1}^p N_k$  and  $M_T = \sum_{k=1}^p M_k$ . It is assumed that  $W$  is of full rank,  $M_T \leq N_T$ , and hence  $W_\ell$  is of rank  $M_\ell$ . For this assumption to be met, it may be necessary to reparameterize the model. Note that, if (i) holds,  $N_\ell = N$ ; if (ii) holds,  $W_\ell = W_1$  for all  $\ell$ .

The test statistic proposed by Kleinbaum for testing linear hypotheses for the MGLM model is a Wald statistic with BAN estimators of linear functions of the "treatment" parameters and consistent estimators of the variance parameters. More specifically, the hypotheses to be tested are specified by the following set of constraints on the model parameters:

$$H_0: \begin{matrix} C \\ c \times M_T \end{matrix} \begin{matrix} \xi \\ M_T \times 1 \end{matrix} = \begin{matrix} 0 \\ c \times 1 \end{matrix} \quad (3.2.3)$$

where  $C$  is a matrix of arbitrary constants of full rank,  $c \leq M_T$ . In a number of applications,  $C$  has the form,  $C = \text{diagonal } (C_1, C_2, \dots, C_p)$  where  $C_\ell$  is a  $(c_\ell \times M_\ell)$  matrix of arbitrary constants of rank  $c_\ell \leq M_\ell$ . MLE's of  $C\underline{\xi}$  (say,  $C\underline{\hat{\xi}}$ ) are obtained using the likelihood function for the MGLM model. The consistent estimators of the variances,  $(W'\hat{\Omega}^{-1}W)^{-1}$ , of the estimated parameters are theoretically obtained by inverting the derived Fisher's information matrix. In actuality, Kleinbaum obtains the information matrices for the GIM and MDM models which are considerably less complicated, and, from them, infers the solution for the general case. Inserting these estimates into the appropriate quadratic form gives the Wald statistic for testing (3.2.3), namely,

$$X_{W(K)}^2 = (C\underline{\hat{\xi}})' [C(W'\hat{\Omega}^{-1}W)^{-1}C']^{-1} (C\underline{\hat{\xi}}). \quad (3.2.4)$$

Under the null hypothesis,  $X_{W(K)}^2$  is asymptotically  $\chi^2$  with D.F. =  $c$ . Kleinbaum has shown, in addition, that the Wald statistic obtained for the SM linear model as a special case of the MGLM model is equivalent to Hotelling's trace criterion (cf. the previous discussion of mixed categorical data models).

(Several detailed examples of the application of the MGLM model can be found in Kleinbaum (1970) and thus will not be repeated here.)

### 3.3 Notation for the More General Categorical Data Model

In order to extend the GSK approach to the analysis of contingency tables with supplemented margins, several assumptions must be made. First, it is assumed for this generalization that there is no interaction between subjects and the presence of supplemental data (i.e.,

for the responses classified, the joint marginal distributions for subjects with complete data and for those with supplemental data are the same). Also, it is assumed that the supplemental data arises by design (e.g., fewer than  $d$  measurements are made on certain of the experimental units due to cost considerations).

Conditional on the number of individuals with various types of complete or supplemental data being fixed by design a priori, the basic probability model is the product of several multinomials with marginal and/or cell probabilities as fundamental parameters.

The details involved in modifying the GSK computer program to handle the general case of "incomplete" data will be carried out at a later time. However, the "vector version" of the extended model will be illustrated by considering a  $2 \times 2 \times 2$  contingency table with all possible degrees of supplementation and therefore a sizable total number of experimental units of varying degrees of categorization. There are three basic types of experimental units:

- (1) The  $n_{ooo}$  units with complete information on the three responses A, B, C (see Table 3.3.1).
- (2) The  $n_{oo*}$  units with information on A and B but not on C (see Table 3.3.2); similar tables apply to situations with "missing" data on A and correspondingly on B.
- (3) The  $n_{o**}$  units with information on A only (i.e., "missing" on B, C, BC, AB, and AC) (see Table 3.3.3); similar tables apply to situations with information on B only and on C only.

Table 3.3.1 Frequency distribution for experimental units with complete information

		B				
		1	2	1	2	
		C		C		
		1	2	1	2	
A	1	$n_{111}$	$n_{112}$	$n_{121}$	$n_{122}$	$n_{100}$
	2	$n_{211}$	$n_{212}$	$n_{221}$	$n_{222}$	$n_{200}$
		↓	↘	↘	↓	$n_{000}$
		$n_{001}$			$n_{002}$	

Table 3.3.2 Frequency distribution for experimental units with information on A and B only

		B		
		1	2	
A	1	$n_{11*}$	$n_{12*}$	$n_{10*}$
	2	$n_{21*}$	$n_{22*}$	$n_{20*}$
		$n_{01*}$	$n_{02*}$	$n_{00*}$

Table 3.3.3 Frequency distribution for experimental units with information on A only

A	1	$n_{1**}$
	2	$n_{2**}$
		$n_{0**}$

Also, let

$$\frac{\pi'}{1 \times 8} = (\pi_{111}, \pi_{112}, \pi_{121}, \pi_{122}, \pi_{211}, \pi_{212}, \pi_{221}, \pi_{222}) \quad (3.3.1)$$

$$\frac{n'_G}{1 \times 26} = (\underline{n}', \underline{n}'_{(C)}, \underline{n}'_{(B)}, \underline{n}'_{(A)}, \underline{n}'_{(BC)}, \underline{n}'_{(AC)}, \underline{n}'_{(AB)}) \quad (3.3.2)$$

with

$$\frac{n'}{1 \times 8} = (n_{111}, n_{112}, n_{121}, n_{122}, n_{211}, n_{212}, n_{221}, n_{222})$$

$$\frac{n'}{1 \times 4}_{(C)} = (n_{11*}, n_{12*}, n_{21*}, n_{22*})$$

and similarly for  $\frac{n'}{1 \times 4}_{(A)}$ ,  $\frac{n'}{1 \times 4}_{(B)}$

$$\frac{n'}{1 \times 2}_{(BC)} = (n_{1**}, n_{2**})$$

and similarly for  $\frac{n'}{1 \times 2}_{(AB)}$ ,  $\frac{n'}{1 \times 2}_{(AC)}$ .

### 3.4 Estimation

As was pointed out in Chapter 1, maximum likelihood estimation for complete tables usually involves the solution of a system of simultaneous equations which are non-linear and consequently difficult to solve. The situation is considerably more complicated in the more general case with supplemental data. As a result, maximum likelihood estimation of the cell probabilities will be used only in the simpler cases (e.g., two-way tables as in Chapter 4) where certain explicit solutions can be obtained. For most complex tables, iterated weighted least squares estimation will be used to obtain estimates of the cell

probabilities (or functions thereof). Here, if the iterated weighted least squares estimates (WLSE's) are unique, then these WLSE's will be MLE's.

The theory for maximum likelihood estimation and weighted least squares estimation for supplemented 2 x 2 tables is presented in detail in Chapter 4. For the 2 x 2 x 2 case cited in the previous section, the table is sufficiently complex that weighted least squares estimation would be used to derive the required estimates of the cell probabilities,  $\pi_{ijk}$ ,  $i, j, k = 1, 2$ , and their asymptotic covariance matrix,  $\hat{V}$ , using the model

$$E(\underline{n}_G) = X_G \underline{\pi} \quad (3.4.1)$$

where  $\underline{n}_G$  and  $\underline{\pi}$  are as defined previously and

$$X_G =$$

26x8

$n_{000}$	0	0	0	0	0	0	0
0	$n_{000}$	0	0	0	0	0	0
0	0	$n_{000}$	0	0	0	0	0
0	0	0	$n_{000}$	0	0	0	0
0	0	0	0	$n_{000}$	0	0	0
0	0	0	0	0	$n_{000}$	0	0
0	0	0	0	0	0	$n_{000}$	0
0	0	0	0	0	0	0	$n_{000}$
$n_{00*}$	$n_{00*}$	0	0	0	0	0	0
0	0	$n_{00*}$	$n_{00*}$	0	0	0	0
0	0	0	0	$n_{00*}$	$n_{00*}$	0	0
0	0	0	0	0	0	$n_{00*}$	$n_{00*}$
$n_{0*0}$	0	$n_{0*0}$	0	0	0	0	0
0	$n_{0*0}$	0	$n_{0*0}$	0	0	0	0
0	0	0	0	$n_{0*0}$	0	$n_{0*0}$	0
0	0	0	0	0	$n_{0*0}$	0	$n_{0*0}$
$n_{*00}$	0	0	0	$n_{*00}$	0	0	0
0	$n_{*00}$	0	0	0	$n_{*00}$	0	0
0	0	$n_{*00}$	0	0	0	$n_{*00}$	0
0	0	0	$n_{*00}$	0	0	0	$n_{*00}$
$n_{0**}$	$n_{0**}$	$n_{0**}$	$n_{0**}$	0	0	0	0
0	0	0	0	$n_{0**}$	$n_{0**}$	$n_{0**}$	$n_{0**}$
$n_{*0*}$	$n_{*0*}$	0	0	$n_{*0*}$	$n_{*0*}$	0	0
0	0	$n_{*0*}$	$n_{*0*}$	0	0	$n_{*0*}$	$n_{*0*}$
$n_{**0}$	0	$n_{**0}$	0	$n_{**0}$	0	$n_{**0}$	0
0	$n_{**0}$	0	$n_{**0}$	0	$n_{**0}$	0	$n_{**0}$

(3.4.2)

### 3.5 Tests of Hypotheses

Testing various hypotheses for the more general case of supplemental data using a modification of the Grizzle, Starmer, Koch (GSK) program is essentially a two-stage procedure. The first stage replaces the first portion of the GSK program while the second stages are, for the most part, the same.

More specifically, the first stage consists of the following:

- (1) Input the observed frequencies (e.g.,  $n_G$  for the  $2 \times 2 \times 2$  case).
- (ii) Obtain estimates of the cell probabilities corresponding to the unrestricted MLE's for the case with complete data:
  - (1) If the table is fairly simple, obtain the MLE's (as in (4.2.2.7)). This is the closest analogue to the unrestricted MLE's for the complete data version and is most consistent with tests using Wald statistics.
  - (2) For most cases, compute the WLSE's (iterated if required) using a formulation similar to (3.4.1) for the particular table of interest.
  - (3) The most general (and heuristic) approach is to use WLSE's of  $A\pi$  (thereby estimating functions of the cell probabilities). This is particularly appropriate in the mixed model situation.

Note that, since in the complete case, tests (e.g., those involving Wald statistics,  $\chi^2_N$ , etc.) using MLE's or WLSE's are asymptotically equivalent, it seems reasonable that the same should apply to the present case of supplemental data.

- (iii) Obtain an estimate,  $\hat{V}$  (or  $\hat{A}\hat{V}\hat{A}'$  for (3)), of the asymptotic covariance matrix using the estimates derived in (ii) (see Section 4.5 for certain details).

The second stage consists of the following:

- (i) As in the existing computer program, input the following matrices as required:

A = contrast matrix

X = design matrix

C = hypothesis matrix

K (for logarithmic functions of the cell probabilities)

- (ii) Use the adjusted (for supplemental data) estimates of the cell probabilities as obtained in (ii) of the first stage in place of the estimates given in (1.4.1.2) along with the estimate of the asymptotic covariance matrix in (iii) of the first stage in place of the usual estimate given in (1.4.1.6). Proceed with the testing of various hypotheses exactly as in the existing GSK framework.

### 3.6 Examples Using the More General Model

#### 3.6.1 Perinatal Health Survey Example

To illustrate the more general categorical data model and its applicability, consider the following medical example involving 456 premature live births. To be considered premature in this situation, the birthweight could not exceed 2000 grams (or 4.4 lbs.) nor could the period of gestation exceed 34 weeks. One of the main questions of

interest was that of an association between the 5-minute Apgar index and the serum bilirubin level for premature infants. The 5-minute Apgar index is a composite of heart rate, respiratory effort, reflex irritability, muscle tone and color observed five minutes after birth. Since each component receives a score of 0, 1, or 2, the index range is 0-10 with the lower values representing the healthier infants. A serum bilirubin count exceeding 1.0 mg per 100 ml indicates a malfunctioning kidney in the premature infant which is usually accompanied by considerable jaundice. If allowed to persist, this condition can lead to damage of the central nervous system. Thus, the recognition of an association between the Apgar index and the serum bilirubin count in premature infants could help in the detection of potentially serious conditions by, for example, the immediate recognition of a jaundiced infant, in turn, suggesting heart or respiratory trouble.

As indicated by Table 3.6.1.1, this is a "no factor, two response" situation with supplementation on both margins. The large number (153) of infants with no serum bilirubin reading might be accounted for by advanced jaundice for these premature infants making the reading unnecessary. If this is indeed the case, the large proportion with low Apgar scores provides evidence contrary to the hypothesis of a positive association between the two responses (which is only slightly supported by the 279 individuals with both measurements available). Alternatively, perhaps those infants with low Apgar scores and normal complexions were not examined for serum bilirubin levels. Quite possibly (from external evidence), they would have had low bilirubin levels and thus supported the association hypothesis. At any rate, it

would seem that methods which utilize the considerable information in the one-way margins are most appropriate for this example.

Table 3.6.1.1 Serum bilirubin readings and 5-minute Apgar scores for a sample of premature infants

5-Minute Apgar Score	Serum Bilirubin Reading		Sub- total	Supplementation on Apgar Score	Total
	0-1.0	1.1 and above			
0-6	35	57	92	117	209
7-10	75	112	187	36	223
Sub-total	110	169	279	153	432
Supplementation on Bilirubin Reading	11	13	24		
Total	121	182	303		456

As has been stated, the test of interest is that of independence of the 5-minute Apgar score and serum bilirubin reading for premature infants, i.e.,

$$H_0: f(\underline{\pi}) = \ln\left(\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}\right) = 0 \quad (3.6.1.1)$$

where

$$\pi_{ij} = \Pr[\text{Apgar score } i, \text{ serum bilirubin reading } j; i, j = 1, 2] .$$

In the two-stage test procedure, the appropriate method of estimation

in the first stage is either (1) or (2) as given in Section 3.5. In the former, iterated MLE's,  $p_{ij}$ ,  $i, j = 1, 2$ , are obtained using the relationship (4.3.1.3) with the computer program given in Chapter 7. The resulting MLE's are given by

$$\begin{aligned} p_{11} &= 0.1877 & p_{12} &= 0.2960 \\ p_{21} &= 0.2092 & p_{22} &= 0.3071 \end{aligned} \quad (3.6.1.2)$$

But these are exactly the iterated WLSE's from (2). The required estimate of the asymptotic covariance matrix for  $\underline{p}$  is obtained by substituting the MLE's in (3.6.1.2) into (4.3.2.1) which is the inverse of the information matrix. Thus, the required estimate is given by

$$\hat{V} = \begin{bmatrix} 0.483 & -0.259 & -0.108 & -0.117 \\ & 0.612 & -0.126 & -0.227 \\ & & 0.522 & -0.288 \\ \text{(symmetric)} & & & 0.632 \end{bmatrix} 10^{-3} \quad (3.6.1.3)$$

For the second stage (cf. Section 1.4.2 for the corresponding test of independence for ordinary  $2 \times 2$  tables using the GSK approach), the following matrices are required:

$$\begin{aligned} A &= I_4 = \text{identity matrix} \\ K &= (1, -1, -1, 1) \\ \hat{V} &\text{ as given in (3.6.1.3)} \\ D &= \text{diagonal } (p_{11}, p_{12}, p_{21}, p_{22}) \end{aligned} \quad (3.6.1.4)$$

The program then provides a test of independence which is the test of fit of the model,  $f(\underline{\pi}) = 0$ . The test statistic is given by

$$\begin{aligned} X_I^2 &= (f(\underline{p}))' [KD^{-1} \hat{A} \hat{V} A' D^{-1} K']^{-1} (f(\underline{p})) \\ &= \frac{(f(\underline{p}))^2}{(KD^{-1} \hat{V} D^{-1} K')} \end{aligned} \quad (3.6.1.5)$$

where

$$f(\underline{p}) = \ln p_{11} - \ln p_{12} - \ln p_{21} + \ln p_{22}$$

$$\begin{aligned} KD^{-1} \hat{V} D^{-1} K' &= \frac{1}{p_{11}} \left( \frac{v_{11}}{p_{11}} - \frac{v_{12}}{p_{12}} - \frac{v_{13}}{p_{21}} + \frac{v_{14}}{p_{22}} \right) - \frac{1}{p_{12}} \left( \frac{v_{12}}{p_{11}} - \frac{v_{22}}{p_{12}} - \frac{v_{23}}{p_{21}} + \frac{v_{24}}{p_{22}} \right) \\ &\quad - \frac{1}{p_{21}} \left( \frac{v_{13}}{p_{11}} - \frac{v_{23}}{p_{12}} - \frac{v_{33}}{p_{21}} + \frac{v_{34}}{p_{22}} \right) + \frac{1}{p_{22}} \left( \frac{v_{14}}{p_{11}} - \frac{v_{24}}{p_{12}} - \frac{v_{34}}{p_{21}} + \frac{v_{44}}{p_{22}} \right) \end{aligned}$$

with  $\hat{V} = (v_{k\ell})$   $k, \ell = 1, 2, 3, 4$ . Under  $H_0$ ,  $X_I^2$  is asymptotically a  $\chi^2$ -variable with D.F. = 1.

For this example,

$$\begin{aligned} X_I^2 &= \frac{(-1.661 + 1.204 + 1.561 - 1.171)^2}{(.0191) - (-.0121) - (-.0171) + (.0117)} \\ &= 0.0748 . \end{aligned} \quad (3.6.1.6)$$

Thus, using the simplified double dichotomy to illustrate the procedure, it would appear that the serum bilirubin reading and the Apgar score are independent.

### 3.6.2 Hand Example

In physical therapy, it has been hypothesized that sweat and sensation in the hand are linked (i.e., sweat is present if, and only if, sensation is also present). In order to investigate this conjecture, 35 patients at the Hand Rehabilitation Center at the University of North Carolina at Chapel Hill were studied. These patients all had severed nerves in the hand which was examined. Because the ring finger is influenced by both the median and ulnar nerves while the other fingers by only one, data was obtained for this finger for only the 12 patients where both nerves were affected by the particular injury (with the exception of patient 35 whose ring finger had been amputated; see Table 3.6.2.1 presenting the results reported in Lachenbruch and Perry (1970)<sup>3</sup>).

Thus, this is a mixed categorical data model of order  $d = 5$  (cf. eye grade example in Chapter 2) with binary ( $r = 2$ ) response categories (agree or disagree) for each finger. In addition, there is supplementary (or incomplete) information for 23 patients representing, in this particular example, the bulk of the available information.

Clearly, outcomes "1" and "4" agree with the hypothesis of association of sweat and sensation while "2" and "3" disagree with this hypothesis. Thus, the results of the investigation can be summarized in terms of "agree with the hypothesis" or "disagree with the hypothesis" as shown in Table 3.6.2.2.

---

<sup>3</sup>Lachenbruch, P. A. and Perry, J. Testing equality of proportion of success of several correlated binomial variates. Submitted to Biometrics, 1970. Department of Biostatistics, University of North Carolina at Chapel Hill.

Table 3.6.2.1 Finger data indicating classification with respect to the presence or absence of sweat or sensation

Patient	Nerve Affected	Finger				
		Thumb	Index	Middle	Fifth	Ring
1	U	1	1	1	3	-
2	C	2	4	4	4	4
3	C	1	3	4	4	4
4	C	3	1	2	2	3
5	C	1	1	1	1	1
6	M	4	4	4	1	-
7	U	1	1	1	4	-
8	M	4	4	4	1	-
9	C	1	1	1	4	1
10	U	1	1	1	4	-
11	U	1	1	1	4	-
12	C	2	2	2	2	2
13	C	2	2	2	2	2
14	U	1	1	1	4	-
15	M	3	4	4	1	-
16	C	4	4	4	4	4
17	M	3	4	4	1	-
18	U	1	1	1	4	-
19	M	1	2	2	1	-
20	M	4	4	4	1	-
21	M	1	1	4	1	-
22	M	2	4	1	1	-
23	M	1	2	2	1	-
24	M	1	2	1	1	-
25	C	1	2	2	2	2
26	C	4	4	4	4	4
27	C	2	4	4	4	4
28	U	1	1	1	4	-
29	M	4	4	4	1	-
30	M	2	1	2	1	-
31	U	1	1	1	4	-
32	M	4	4	2	1	-
33	C	2	2	2	2	2
34	U	1	1	1	2	-
35	C	2	2	2	2	-

where 1 = sweat and sensation present  
 2 = sweat present, sensation absent  
 3 = sweat absent, sensation present  
 4 = sweat and sensation absent

U = ulnar (thumb, index, middle, ring)

M = median (fifth, ring)

C = combination of both U and M (ring finger)

Table 3.6.2.2 Frequency distribution of complete and supplemental data by agreement (A) and disagreement (D) for each finger

Finger	Complete (12 patients)		Supplemental (23 patients)		Total (35 patients)	
	A	D	A	D	A	D
Thumb	6	6	18	5	24	11
Index	7	5	19	4	26	9
Middle	7	5	18	5	25	10
Fifth	7	5	20	3	27	8
Ring	7	5	-	-	7	5
Total	34	26	75	17	109	43

The main questions of interest are (a) whether the marginal probabilities of disagreement are the same for the five fingers and, if so, (b) whether this common value is near zero (i.e., whether there is an association of sweat and sensation). Therefore, the hypotheses of interest are

$$H_{oa}: \pi_{20000} = \pi_{02000} = \pi_{00200} = \pi_{00020} = \pi_{00002} \quad (3.6.2.1)$$

where  $\pi_{20000} = \text{Pr} [\text{disagreement for the thumb}]$

(similarly for the index, middle, fifth, and ring fingers, respectively)

$$H_{ob}: \pi_D = 0 \quad \text{given that } H_{oa} \text{ holds} \quad (3.6.2.2)$$

(where  $\pi_D = \pi_{20000} = \dots = \pi_{00002}$ ).

Since the results for one finger cannot be assumed independent of those for the other fingers, the classical Pearson  $\chi^2$ -test of homogeneity for the proportions is not appropriate for (3.6.2.1). However, the extended linear model for supplemental categorical data is appropriate for this situation and uses all of the available data.

For the estimation involved at the first stage, method (3) of Section 3.5 is most appropriate since interest lies in the marginal probabilities of disagreement rather than individual cell probabilities. Thus, WLSE's of the five marginal probabilities of disagreement (which are functions of the individual cell probabilities) are obtained using the model

$$E(A_1 \begin{matrix} \hat{\pi}_G \\ 9 \times 48 \end{matrix}) = X_1 \begin{matrix} \underline{\beta} \\ 48 \times 1 \end{matrix} \quad (3.6.2.3)$$

where

$$\begin{matrix} \hat{\pi}'_G \\ 1 \times 48 \end{matrix} = (\hat{\pi}', \hat{\pi}'_{(R)}) \quad (3.6.2.4)$$

with

$$\begin{matrix} \hat{\pi}'_G \\ 1 \times 32 \end{matrix} = (\hat{\pi}_{11111}, \hat{\pi}_{11112}, \hat{\pi}_{11121}, \dots, \hat{\pi}_{12222}, \hat{\pi}_{21111}, \hat{\pi}_{21112}, \dots, \hat{\pi}_{22222})$$

where  $\pi_{ijk\ell m} = \text{Pr}$  [level  $i$  of thumb,  $j$  of index,  $k$  of middle,  $\ell$  of fifth, and  $m$  of ring finger,  $i, j, k, \ell, m = 1, 2$ ]

i.e.,  $\hat{\pi}'_G = (4, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 2, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 3) (1/12)$

$$\hat{\underline{\pi}}'_{(R)} = (\hat{\pi}_{1111o}, \hat{\pi}_{1112o}, \hat{\pi}_{1121o}, \dots, \hat{\pi}_{1222o}, \hat{\pi}_{2111o}, \hat{\pi}_{2112o}, \dots, \hat{\pi}_{2222o})_{1 \times 16}$$

where

$$\pi_{ijk\ell o} = \sum_{m=1}^2 \pi_{ijk\ell m} \quad i, j, k, \ell = 1, 2$$

i.e.,  $\hat{\underline{\pi}}'_{(R)} = (12, 2, 1, 0, 1, 0, 2, 0, 3, 0, 1, 0, 0, 0, 0, 1) (1/23)$

from the supplemental information;

$$\underline{\beta}'_{1 \times 5} = (\pi_{2o00o}, \pi_{o200o}, \pi_{oo2oo}, \pi_{ooo2o}, \pi_{oooo2}) \quad (3.6.2.5)$$

where

$$\pi_{2o00o} = \text{Pr ["disagreement with the hypothesis" for the thumb]}$$

$$= \sum_{j=1}^2 \sum_{k=1}^2 \sum_{\ell=1}^2 \sum_{m=1}^2 \pi_{2j k \ell m}$$

and similarly for the remaining fingers;

$$A_1 = \begin{bmatrix} A_{(.)} & 0 \\ 0 & A_{(R)} \end{bmatrix} \quad (3.6.2.6)$$

where



$$\hat{V}_{\pi_G} = \begin{bmatrix} \hat{V}_{(\cdot)\pi} & \underline{0} \\ \underline{0} & \hat{V}_{(R)\pi} \end{bmatrix} \quad (3.6.2.8)$$

48x48

with

$$\hat{V}_{(\cdot)\pi} = \frac{1}{n_{00000}} (D_{\pi} - \hat{\pi} \hat{\pi}') \quad 32 \times 32$$

where  $n_{00000} = 12$ ;  $\hat{\pi}$  is given in (3.6.2.4);

$$\hat{V}_{(R)\pi} = \frac{1}{n_{00000}^*} (D_{\pi(R)} - \hat{\pi}_{(R)} \hat{\pi}_{(R)}') \quad 16 \times 16$$

where  $n_{00000}^* = 23$ ;  $\hat{\pi}_{(R)}$  is given in (3.6.2.4).

A test with D.F. = 4 of (3.6.2.1) is obtained (as in GSK) using the WLSE's,  $\underline{b}$ , of  $\underline{\beta}$ , from the first stage with the contrast matrix

$$C_a = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \quad 4 \times 5$$

since  $H_{0a}$  can equivalently be represented by  $C_a \underline{\beta} = \underline{0}$ . More specifically, the test statistic is given by

$$X_a^2 = (C_a \underline{b})' [C_a (X_1' S^{-1} X_1)^{-1} C_a']^{-1} (C_a \underline{b}) \quad (3.6.2.9)$$



where  $A_{(.)}$  and  $A_{(R)}$  are given in (3.6.2.6). However, for these data, the estimate of the covariance matrix for  $A_{2\underline{G}}$ , namely  $A_{2\underline{G}}\hat{V}_{\underline{G}}A'_{2\underline{G}}$ , is singular. This is due to the middle, ring, and fifth fingers being perfectly correlated for those 12 individuals with complete information. Although the data could be manipulated to avoid this difficulty, the preceding should adequately describe a test procedure which appropriately utilizes all of the available information.

It should be noted in addition that having both nerves (median and ulnar) severed appears to have a quite different effect on sweat and sensation than having only one nerve severed (see Table 3.6.2.2). For the 12 patients with both nerves severed, there is strong evidence against the hypothesis that sweat and sensation in the hand are linked. A notably different distribution of proportion of "disagreements with the hypothesis" occurs for those 23 patients with only one severed nerve. This raises the question about the appropriateness of the uniform hypothesis for all five fingers, especially in light of the different physiology of the ring finger. Nevertheless, the example is still useful in showing the potential of the more general categorical data model.

### 3.6.3 Drug Example

This is an example of a mixed categorical data model of order 2 with supplemental information on both margins. Suppose a certain pharmaceutical company wished to compare the effects of drugs A and B. Of the 80 patients available for study, 50 patients (Group I) received both drugs under essentially the same conditions including a suitable

time delay between applications (with the order of presentation of A and B random); 16 patients (Group II) received drug A only; and 14 patients (Group III) received drug B only. In each case, the patient's response to the particular drug was observed and recorded as "none", "slight", or "strong". The results are given in Table 3.6.3.1. The main question of interest is whether the distribution of strength of response is the same for both drugs.

Table 3.6.3.1 Frequency distribution of responses to drugs A and B

[Group I]	Response to B			Sub- total	Response to A only [Group II]	Total
Response to A	None (1)	Slight (2)	Strong (3)			
None (1)	4	6	0	10	4	14
Slight (2)	4	8	9	21	7	28
Strong (3)	0	5	14	19	5	24
Sub-total	8	19	23	50	16	66
Response to B only [Group III]	3	4	7	14		
Total	11	23	30	64		80

For the estimation involved at the first stage, methods (2) and (3) of Section 3.5 are appropriate. For the former,

$$\frac{\pi'}{1 \times 9} = (\pi_{11}, \pi_{12}, \pi_{13}, \pi_{21}, \pi_{22}, \pi_{23}, \pi_{31}, \pi_{32}, \pi_{33}) \quad (3.6.3.1)$$

$$\frac{n'}{1 \times 15} = (n', n'_{(B)}, n'_{(A)})$$

where

$$\underline{n}'_{1 \times 9} = (4, 6, 0, 4, 8, 9, 0, 5, 14)$$

$$\underline{n}'_{1 \times 3}(B) = (4, 7, 5)$$

$$\underline{n}'_{1 \times 3}(A) = (3, 4, 7)$$

$$\underline{X}_{15 \times 9} = \begin{bmatrix} 50 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 50 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 50 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 50 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 50 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 50 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 50 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 50 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 50 \\ 16 & 16 & 16 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 16 & 16 & 16 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 16 & 16 & 16 \\ 14 & 0 & 0 & 14 & 0 & 0 & 14 & 0 & 0 \\ 0 & 14 & 0 & 0 & 14 & 0 & 0 & 14 & 0 \\ 0 & 0 & 14 & 0 & 0 & 14 & 0 & 0 & 14 \end{bmatrix} \quad (3.6.3.2)$$

with estimated covariance matrix for  $\underline{n}_G$

$$\hat{V}_{15 \times 15} = \begin{bmatrix} \hat{V}_{( )} & Q & Q \\ Q & \hat{V}_{(B)} & Q \\ Q & Q & \hat{V}_{(A)} \end{bmatrix} \quad (3.6.3.3)$$

where

$$\hat{V}_{( )} = n_{oo} (D_{\hat{\pi}} - \hat{\pi} \hat{\pi}')_{9 \times 9}$$

with  $n_{oo} = 50;$

$$\begin{aligned} D_{\hat{\pi}} &= \text{diagonal } (\hat{\pi}_{11}, \hat{\pi}_{12}, \dots, \hat{\pi}_{33}) \\ &= \text{diagonal } \left( \frac{4}{50}, \frac{6}{50}, \dots, \frac{14}{50} \right) \end{aligned}$$

$$\hat{V}_{(B)} = n_{o*} (D_{\hat{\pi}_o} - \hat{\pi}_o \hat{\pi}'_o)_{3 \times 3}$$

with  $n_{o*} = 16;$

$$\begin{aligned} D_{\hat{\pi}_o} &= \text{diagonal } (\hat{\pi}_{1o}, \hat{\pi}_{2o}, \hat{\pi}_{3o}) \\ &= \text{diagonal } \left( \frac{4}{16}, \frac{7}{16}, \frac{5}{16} \right) \end{aligned}$$

$$\hat{V}_{(A)} = n_{*o} (D_{\hat{\pi}_o} - \hat{\pi}_o \hat{\pi}'_o)_{3 \times 3}$$

with  $n_{*o} = 14;$

$$\begin{aligned} D_{\hat{\pi}_0} &= \text{diagonal } (\hat{\pi}_{01}, \hat{\pi}_{02}, \hat{\pi}_{03}) \\ &= \text{diagonal } \left( \frac{3}{14}, \frac{4}{14}, \frac{7}{14} \right) . \end{aligned}$$

Thus, the first stage estimates of  $\underline{\pi}$  could be obtained as in Example 1.

However, two adjustments must be made, namely,

- (i) Reparameterization;
- (ii) Modification for zero cells.

The first can be accomplished by deleting the ninth, twelfth, and fifteenth rows and last column of  $X$  with the corresponding adjustments made on  $\underline{n}$ ,  $\underline{n}_G$  and  $\hat{V}$ . For the second adjustment, replace each zero cell frequency by a small positive quantity (e.g.,  $1/s = 1/9$  where  $s =$  number of responses for Group I). Note, that, if there were a zero marginal frequency, it could be replaced by the corresponding  $1/3$  in this case. With these adjustments made, replace the usual first stage estimates from the GSK program by the resulting first stage WLSE's ( $\tilde{\underline{\pi}}$  and  $\tilde{V}$ ), formulate the hypotheses of interest, e.g., marginal symmetry as specified by

$$H_0: \pi_{i0} = \pi_{0i} \quad i = 1, 2 \quad (3.6.3.4)$$

and proceed as in the second stage of GSK (as indicated for the previous examples).

Alternatively, for this example, it appears reasonable to assign scores to the three levels of the two responses and to test for equality of mean scores over the two marginals, namely

$$H_0: \alpha_A = \alpha_B \quad (3.6.3.5)$$

where  $\alpha_A$  and  $\alpha_B$  are the mean scores for the responses to drugs A and B, respectively. Therefore, assigning scores  $a_1 = 1$ ,  $a_2 = 2$ , and  $a_3 = 3$  to the response categories for both drugs A and B and using initial relative frequencies  $(\hat{\pi}_G)$  in place of the observation vector  $(n_G)$ , the first stage estimation involves the process referred to in (3) of Section 3.5. Specifically, WLSE's,  $\tilde{\alpha}_A$  and  $\tilde{\alpha}_B$ , are obtained using the model

$$E \left( \begin{matrix} A_1 \\ \hat{\pi}_G \end{matrix} \right) = \begin{matrix} X_1 \\ \alpha \end{matrix} \quad (3.6.3.6)$$

$\begin{matrix} 4 \times 15 & 15 \times 1 & & 4 \times 2 & 2 \times 1 \end{matrix}$

where

$$\hat{\pi}_G = (\hat{\pi}', \hat{\pi}'_{\cdot 0}, \hat{\pi}'_{\cdot 0.}) \quad \text{as in (3.6.3.1)} \quad (3.6.3.7)$$

with

$$\hat{\pi}' = \frac{1}{50} (4, 6, 0, 4, 8, 9, 0, 5, 14)$$

$$\hat{\pi}'_{\cdot 0} = \frac{1}{16} (4, 7, 5)$$

$$\hat{\pi}'_{\cdot 0.} = \frac{1}{14} (3, 4, 7)$$

$$\alpha' = (\alpha_A, \alpha_B)$$

$$A_1 = \begin{bmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 1 & 2 & 3 & 1 & 2 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 \end{bmatrix} \quad (3.6.3.8)$$

$$X_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (3.6.3.9)$$

The inverse ( $\hat{V}$ ) of the weight matrix is given by  $A_1 \hat{V}_{\underline{\pi}_G} A_1'$  where

$$\hat{V}_{\underline{\pi}_G} = \begin{bmatrix} \hat{V}_{( ) \underline{\pi}} & \underline{0} \\ \underline{0} & \hat{V}_{(B) \underline{\pi}} \\ \underline{0} & \underline{0} & \hat{V}_{(A) \underline{\pi}} \end{bmatrix} \quad (3.6.3.10)$$

$15 \times 15$

with

$$\hat{V}_{( ) \underline{\pi}} = \frac{1}{n_{00}^2} \hat{V}_{( )} \quad \text{as given in (3.6.3.3)}$$

$9 \times 9$

$$\hat{V}_{(B) \underline{\pi}} = \frac{1}{n_{0*}^2} \hat{V}_{(B)}$$

$3 \times 3$

$$\hat{V}_{(A) \underline{\pi}} = \frac{1}{n_{*0}^2} \hat{V}_{(A)}$$

$3 \times 3$

Finally, a test with D.F. = 1 for (3.6.3.5) is obtained (as in GSK) using  $\tilde{\alpha}_A$  and  $\tilde{\alpha}_B$  from the first stage with the contrast matrix,  $C = (1, -1)$ . Alternatively, if the linear model

$$E(\tilde{\alpha}) = X_2 \mu \quad \text{where} \quad X_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (3.6.3.11)$$



#### 4. ESTIMATION THEORY AND HYPOTHESIS TESTING FOR TWO-WAY CONTINGENCY TABLES WITH SUPPLEMENTED MARGINS

##### 4.1 Introduction

In this chapter, the estimation theory for two-way tables with supplemented margins is investigated in detail since it is possible to obtain certain explicit results for maximum likelihood estimation in this relatively simple case and to compare these results with other competing methods of estimation. Various properties of the resulting estimates are presented and the more general categorical data model approach to hypothesis testing is illustrated for a  $2 \times 2$  table with both margins supplemented.

##### 4.2 2 x 2 Case with One Margin Supplemented

###### 4.2.1 Notation

Consider the following extension of the usual  $2 \times 2$  contingency table (see Table 4.2.1.1), where  $\pi_{ij}$ ,  $\pi_{i0}$ ,  $\pi_{0j}$ ,  $n_{ij}$ ,  $n_{i0}$ ,  $n_{0j}$ , and  $n_{00}$ ,  $i, j = 1, 2$ , are as defined previously. Then,  $n_{0*}$  additional observations are available for response A with  $n_{i*}$  responses at level  $i$ ,  $i = 1, 2$ . For simplicity it is assumed throughout this research that

$$\begin{aligned} \pi_{i*} &= \text{Pr} [\text{level } i \text{ of response A for the supplemented dimension}] \\ &= \pi_{i1} + \pi_{i2} = \pi_{i0} \quad i = 1, 2 \end{aligned}$$

although a number of alternative assumptions on the supplemental prob-

abilities are possible (e.g.,  $\pi_{i*} = \sum_{j=1}^2 a_{ij} \pi_{ij} \neq \pi_{i0}$  where

Table 4.2.1.1 Expected cell probabilities (cell frequencies) for the 2 x 2 case with one margin supplemented

Response A Levels	Response B Levels		Sub-total	Supplemented Margin	Total
	1	2			
1	$\pi_{11}$ ( $n_{11}$ )	$\pi_{12}$ ( $n_{12}$ )	$\pi_{10}$ ( $n_{10}$ )	$\pi_{1*}$ ( $n_{1*}$ )	( $N_{10}$ )
2	$\pi_{21}$ ( $n_{21}$ )	$\pi_{22}$ ( $n_{22}$ )	$\pi_{20}$ ( $n_{20}$ )	$\pi_{2*}$ ( $n_{2*}$ )	( $N_{20}$ )
Total	$\pi_{01}$ ( $n_{01}$ )	$\pi_{02}$ ( $n_{02}$ )	1 ( $n_{00}$ )	1 ( $n_{0*}$ )	(N)

$\sum_{j=1}^2 a_{ij} = 1$ ). It will also be assumed for simplicity that

$$0 < \pi_{ij} < 1 \quad i, j = 1, 2$$

so that  $\Pr [n_{i0} = 0] = (1 - \pi_{i0})^{n_{00}} \rightarrow 0$  as  $n_{00} \rightarrow \infty$ . Thus, a total sample of  $N$  individuals is taken with both responses observed for  $n_{00}$  subjects and with information on response A only for  $n_{0*} = N - n_{00}$  individuals.

#### 4.2.2 Methods of Estimation

Intuitive estimates of the individual cell probabilities using the supplemental information are given by the following:

$$\hat{\pi}_{ij} = \Pr [\text{level } i \text{ of response A, level } j \text{ of response B}]$$

$$= \Pr [\text{row } i] \Pr [\text{column } j | \text{row } i]$$

$$= \frac{N_{i0}}{N} \cdot \frac{n_{i0}}{n_{00}} \quad i, j = 1, 2$$

$$= \left( \frac{n_{ij}}{n_{00}} \right) \left[ \frac{1 + \frac{n_{i*}}{n_{i0}}}{1 + \frac{n_{0*}}{n_{00}}} \right]$$

$$= \left( \begin{array}{c} \text{unsupplemented} \\ \text{unrestricted MLE} \end{array} \right) \times \left[ \frac{\begin{array}{c} \text{ratio of supplemented to} \\ 1 + \text{unsupplemented in } i^{\text{th}} \text{ row} \\ \text{ratio of total supplemented} \\ 1 + \text{to total unsupplemented} \end{array}}{\quad} \right]$$

where  $\hat{\pi}_{ij} = 0$  with probability approaching zero as  $n_{00}$  becomes large. But  $\hat{\pi}_{ij}$  is exactly the unrestricted MLE of  $\pi_{ij}$  (say  $\hat{\pi}_{ij} = p_{ij}$ , as before). This can be seen as follows: By virtue of the independence of the two samples (with sample sizes  $n_{00}$  and  $n_{0*}$ ), the joint likelihood function is given by

$$\begin{aligned} \ell &= n_{00}! \prod_{i=1}^2 \prod_{j=1}^2 \frac{(\pi_{ij})^{n_{ij}}}{n_{ij}!} \cdot n_{0*}! \prod_{i=1}^2 \frac{(\pi_{i0})^{n_{i*}}}{n_{i*}!} \\ &= c_1 \prod_{i=1}^2 \prod_{j=1}^2 (\pi_{ij})^{n_{ij}} \prod_{i=1}^2 (\pi_{i1} + \pi_{i2})^{n_{i*}} \end{aligned}$$

where

$$c_1 = \frac{n_{00}! n_{0*}!}{\prod_{i=1}^2 \prod_{j=1}^2 n_{ij}! \prod_{i=1}^2 n_{i*}!}$$

Utilizing a Lagrangian multiplier for the restriction  $\sum_{i=1}^2 \sum_{j=1}^2 \pi_{ij} = 1$ ,

the log likelihood function (with the restriction) is given by

$$f = c_2 + \sum_{i=1}^2 \sum_{j=1}^2 n_{ij} \ln \pi_{ij} + \sum_{i=1}^2 n_{i*} \ln \pi_{i0} - \lambda \left( \sum_{i=1}^2 \sum_{j=1}^2 \pi_{ij} - 1 \right) \quad (4.2.2.1)$$

where  $c_2 = \ln c_1$

with corresponding normal equations

$$\frac{\partial f}{\partial \pi_{ij}} \Big|_{\pi=p} = \frac{n_{ij}}{p_{ij}} + \frac{n_{i*}}{p_{i0}} - \lambda \stackrel{\text{set}}{=} 0 \quad (4.2.2.2)$$

From (4.2.2.2), we have that

$$n_{ij} + n_{i*} \frac{p_{ij}}{p_{i0}} - \lambda p_{ij} = 0$$

and, upon summing over  $j$  and then  $i$ , it follows that  $\lambda = N$ . Now

$$\frac{n_{ij}}{p_{ij}} + \frac{n_{i*}}{p_{i0}} = N \quad i, j = 1, 2 \quad (4.2.2.3)$$

implies that

$$p_{ij} = \frac{n_{ij}}{N} + \frac{n_{i*}}{N} \frac{p_{ij}}{p_{io}} \quad (4.2.2.4)$$

Hence

$$\begin{aligned} p_{io} &= \sum_{j=1}^2 \left( \frac{n_{ij}}{N} + \frac{n_{i*}}{N} \frac{p_{ij}}{p_{io}} \right) \quad \text{from (4.2.2.4)} \\ &= \frac{N_{io}}{N} \quad (4.2.2.5) \end{aligned}$$

But (4.2.2.3) also implies that

$$p_{ij} = \frac{n_{ij}}{N - \frac{n_{i*}}{p_{io}}} \quad (4.2.2.6)$$

so that an explicit solution of the normal equations is given by

$$\begin{aligned} p_{ij} &= \hat{\pi}_{ij} = \frac{n_{ij}}{N - \frac{n_{i*}}{p_{io}}} \quad \text{from (4.2.2.6)} \\ &= \frac{n_{ij}}{N - n_{i*} / \left( \frac{N_{io}}{N} \right)} \quad \text{from (4.2.2.5)} \\ &= \frac{N_{io}}{N} \cdot \frac{n_{ij}}{n_{io}} \quad i, j = 1, 2 \quad (4.2.2.7) \end{aligned}$$

which is exactly the intuitive estimate given previously.

Alternatively, let

$$\underline{n}_G' = (n_{11}, n_{12}, n_{21}, n_{1*})$$

1x4

$$\underline{\pi}'_s = (\pi_{11}, \pi_{12}, \pi_{21})$$

1x3

$$X = \begin{bmatrix} n_{oo} & 0 & 0 \\ 0 & n_{oo} & 0 \\ 0 & 0 & n_{oo} \\ n_{o*} & n_{o*} & 0 \end{bmatrix}$$

4x3

Weighted least squares estimates,  $\tilde{\pi}_{ij}$ , are obtained using the model

$$E(\underline{n}_G) = X\underline{\pi}_s \quad (4.2.2.8)$$

$$\text{Since } V = \text{var}(\underline{n}_G) = \begin{bmatrix} n_{oo}(D_{\underline{\pi}_s} - \underline{\pi}_s \underline{\pi}'_s) & \underline{0} \\ \underline{0}' & n_{o*} \pi_{1o} (1 - \pi_{1o}) \end{bmatrix}$$

4x4

where  $D_{\underline{\pi}_s} = \text{diagonal} (\pi_{11}, \pi_{12}, \pi_{21})$

$\underline{0}$  is a (3 x 1) vector of 0's,

the weighted least squares estimates,  $\tilde{\pi}_s$ , are given by

$$\tilde{\pi}_s = (X' \tilde{V}^{-1} X)^{-1} X' \tilde{V}^{-1} \underline{n}_G \quad (4.2.2.9)$$

where  $\tilde{V} = V_{\frac{\tilde{\pi}_s - \tilde{\pi}_s}{\tilde{\pi}_s}}$

$$\tilde{V}^{-1} = \begin{bmatrix} \frac{1}{n_{00}} (D_{\tilde{\pi}_s}^{-1} + \frac{1}{\tilde{\pi}_{22}} \cdot J) & 0 \\ 0' & \frac{1}{n_{0*} \tilde{\pi}_{10} (1 - \tilde{\pi}_{10})} \end{bmatrix}$$

with J a (3 x 3) matrix of 1's. Performing the indicated matrix multiplication and simplifying yields

$$\tilde{\pi}_{ij} = \frac{1}{N} (n_{ij} + n_{i*} \frac{\tilde{\pi}_{ij}}{\tilde{\pi}_{i0}} + \frac{n_{0*}}{n_{00}} \cdot \frac{n_{ij} \tilde{\pi}_{ij'} - n_{ij'} \tilde{\pi}_{ij}}{\tilde{\pi}_{i0}}) \quad j \neq j' \quad (4.2.2.10)$$

which is in a format convenient for iterating. E.g., in the right hand side of (4.2.2.10), let

$$\tilde{\pi}_{1j}^{(0)} = \frac{n_{1j}}{n_{00}} \quad j = 1, 2$$

= unrestricted MLE without supplemented margins

obtaining  $\tilde{\pi}_{1j}^{(1)}$ ,  $j = 1, 2$ . Replace  $\tilde{\pi}_{1j}^{(0)}$  by  $\tilde{\pi}_{1j}^{(1)}$  in the right hand side of (4.2.2.10) and continue the iteration until

$$\max (|\tilde{\pi}_{11}^{(k)} - \tilde{\pi}_{11}^{(k-1)}|, |\tilde{\pi}_{12}^{(k)} - \tilde{\pi}_{12}^{(k-1)}|) < \epsilon \quad \text{for some prescribed } \epsilon.$$

From numerical work on the computer, this process appears to converge quite rapidly.

It should be noted that both  $\tilde{\pi}_{ij}$  and  $p_{ij} (= \hat{\pi}_{ij} = \hat{\pi}_{ij})$  reduce to the ordinary unrestricted MLE when  $n_{o*} = 0$  (i.e., when there is no supplemental information). Also,  $p_{ij}$  is a stationary solution of (4.2.2.10) (i.e., the MLE's in (4.2.2.7) satisfy the relations (4.2.2.10) for the weighted least squares estimates). Thus, if the iterated WLSE's are unique,  $p_{ij} = \tilde{\pi}_{ij}$ ,  $i, j = 1, 2$ .

#### 4.2.3 Properties of the Estimators

In order to find approximations to the means, variances and covariances of the MLE's,  $p_{ij} = \frac{N_{io}}{N} \cdot \frac{n_{ij}}{n_{io}}$ ,  $i, j = 1, 2$ , it is convenient to use the Taylor series expansion of  $p_{ij}$  about the point

$$\begin{aligned} E(\underline{q}') &= E(q_{11}, q_{12}, q_{21}, q_{1*}) \\ &= E\left(\frac{n_{11}}{n_{oo}}, \frac{n_{12}}{n_{oo}}, \frac{n_{21}}{n_{oo}}, \frac{n_{1*}}{n_{o*}}\right) \\ &= (\pi_{11}, \pi_{12}, \pi_{21}, \pi_{1o}) \end{aligned}$$

Now

$$\begin{aligned} p_{ij} &= f(\underline{q}) = \frac{n_{io} + n_{i*}}{N} \cdot \frac{n_{ij}}{n_{io}} \\ &= \frac{n_{oo}}{N} q_{ij} \left(1 + \frac{n_{o*}}{n_{oo}} \frac{q_{i*}}{q_{io}}\right) \quad i, j = 1, 2 \end{aligned} \quad (4.2.3.1)$$

where  $q_{io} = q_{i1} + q_{i2}$

so that the Taylor series expansion (through the linear terms) is given by

$$p_{ij} = \hat{\pi}_{ij} = f(\underline{q}) \quad (4.2.3.2)$$

$$\begin{aligned} & \doteq \pi_{ij} + \left(1 - \frac{n_{o*}}{N} \frac{\pi_{ij}}{\pi_{io}}\right) (q_{ij} - \pi_{ij}) - \frac{n_{o*}}{N} \frac{\pi_{ij}}{\pi_{io}} (q_{ij'} - \pi_{ij'}) \\ & \quad + \frac{n_{o*}}{N} \frac{\pi_{ij}}{\pi_{io}} (q_{i*} - \pi_{io}) \quad i, j = 1, 2 \quad j \neq j' . \end{aligned}$$

The MLE expressed by its Taylor series expansion (even including quadratic terms) is unbiased. Using (4.2.3.2), it can be shown that

$$\begin{aligned} \text{var}(p_{ij}) & \doteq \left(\frac{n_{o*}}{N}\right)^2 \left[ \left(\frac{\pi_{ij'}}{\pi_{io}} + \frac{n_{oo}}{n_{o*}}\right)^2 \frac{\pi_{ij}(1-\pi_{ij})}{n_{oo}} + \left(\frac{-\pi_{ij}}{\pi_{io}}\right)^2 \frac{\pi_{ij'}(1-\pi_{ij'})}{n_{oo}} \right. \\ & \quad \left. + \left(\frac{\pi_{ij}}{\pi_{io}}\right)^2 \frac{\pi_{io}(1-\pi_{io})}{n_{o*}} + 2\left(\frac{\pi_{ij'}}{\pi_{io}} + \frac{n_{oo}}{n_{o*}}\right) \left(\frac{-\pi_{ij}}{\pi_{io}}\right) \left(\frac{-\pi_{ij'}\pi_{ij'}}{n_{oo}}\right) \right] \\ & = \frac{\pi_{ij}(1-\pi_{ij})}{N} + \frac{n_{o*}}{Nn_{oo}} \frac{\pi_{ij}\pi_{ij'}}{\pi_{io}} \quad i, j = 1, 2 \\ & \quad \quad \quad j \neq j' . \end{aligned} \quad (4.2.3.3)$$

This is exactly the asymptotic variance obtained by inverting Fisher's information matrix,  $I$ , where

$$I = \left( -E \left( \frac{\partial^2 \ln L(\underline{n}; \underline{\pi}_s)}{\partial \pi_{ij} \partial \pi_{i'j'}} \right) \right)_{3 \times 3}$$

where  $L(\underline{n}; \underline{\pi}_s)$  denotes  $f$  as defined in (4.2.2.1) but without the restriction. It is also the asymptotic variance of the WLSE,  $\tilde{\pi}_{ij}$ ,

namely,  $(X'V^{-1}X)^{-1}$ . In addition, the MLE's satisfy the relations given in (4.2.2.10) for the WLSE's. Thus, it would appear that the MLE's and iterated WLSE's are identical for the case of  $2 \times 2$  tables with one margin supplemented. This is indeed the case if the iterated WLSE's are unique.

The asymptotic covariances of the MLE's are also elements of the inverse of the information matrix which is given by the following:

$$I^{-1} = \frac{\pi_{22}}{n_{00}} \begin{bmatrix} 1 + \frac{\pi_{22}}{\pi_{11}} + \alpha_r & 1 + \alpha_r & 1 \\ & 1 + \frac{\pi_{22}}{\pi_{12}} + \alpha_r & 1 \\ \text{(symmetric)} & & 1 + \frac{\pi_{22}}{\pi_{21}} \end{bmatrix}^{-1} \quad (4.2.3.4)$$

$$= \frac{1}{N} \begin{bmatrix} \pi_{11}(1-\pi_{11}) + \gamma_{r_1} & -\pi_{11}\pi_{12} - \gamma_{r_1} & -\pi_{11}\pi_{21} \\ & \pi_{12}(1-\pi_{12}) + \gamma_{r_1} & -\pi_{12}\pi_{21} \\ \text{(symmetric)} & & \pi_{21}(1-\pi_{21}) + \gamma_{r_2} \end{bmatrix}$$

where

$$\alpha_r = \frac{n_{0*}}{n_{00}} \frac{\pi_{20}}{\pi_{10}\pi_{20}}$$

$$\gamma_{r_i} = \frac{n_{0*}}{n_{00}} \frac{\pi_{i1}\pi_{i2}}{\pi_{i0}} \quad i = 1, 2$$

which reduces to the usual covariance matrix when  $n_{0*} = 0$  (i.e., when there is no supplemental information). Also depending on the sample sizes selected (i.e.,  $n_{00}$  and  $n_{0*}$ ) and the values of  $\pi_{ij}$ ,  $\pi_{ij'}$ , ( $j \neq j'$ ),  $p_{ij}$  may be more efficient than the corresponding MLE,  $p_{ij}(u)$ , for the usual situation since

$$\text{var}(p_{ij}) = \frac{n_{00}}{N} \text{var}(p_{ij}(u)) + \frac{n_{0*}}{Nn_{00}} \frac{\pi_{ij}\pi_{ij'}}{\pi_{i0}} \quad (j \neq j') .$$

Note also that the multiparameter extension for the Cramér-Rao lower bound could be used to determine if the MLE's are jointly optimal. In addition, since the MLE's are BAN estimators, they are consistent (i.e.,  $\Pr [ |p_{ij} - \pi_{ij}| > \epsilon \text{ for } i, j = 1, 2 ] \xrightarrow[N \rightarrow \infty]{} 0$ ) and asymptotically efficient.

Alternatively, exact means as well as improved estimates of the variances of the  $p_{ij}$ ,  $i, j = 1, 2$ , can be derived using conditional means and variances. More explicitly, the following relationships will be used:

$$E(p_{ij}) = E[E(p_{ij} | n_{i0})] \quad (4.2.3.5)$$

$$\text{var}(p_{ij}) = E[\text{var}(p_{ij} | n_{i0})] + \text{var}[E(p_{ij} | n_{i0})] \quad i, j = 1, 2 .$$

Now, since  $(n_{i1}, n_{i2})$  is Multinomial  $(n_{i0}; \pi_{i1}, \pi_{i2})$  and  $n_{i0}$  is Binomial  $(n_{00}; \pi_{i0})$ , it follows that  $(n_{ij} | n_{i0})$  is Binomial  $(n_{i0}; \pi_{ij}/\pi_{i0})$ . Thus

$$\begin{aligned}
E(p_{ij}) &= E[E(p_{ij} | n_{io})] \\
&= E\left[E\left(\frac{n_{io} + n_{i^*}}{N} \cdot \frac{n_{ij}}{n_{io}} \mid n_{io}\right)\right] \\
&= E\left[\frac{n_{io} + n_{o^*} \pi_{io}}{N} \cdot \frac{n_{io} \pi_{ij} / \pi_{io}}{n_{io}}\right] \quad (4.2.3.6) \\
&= E\left[\frac{n_{io} + n_{o^*} \pi_{io}}{N} \frac{\pi_{ij}}{\pi_{io}}\right] \\
&= \frac{n_{oo} \pi_{io} + n_{o^*} \pi_{io}}{N} \cdot \frac{\pi_{ij}}{\pi_{io}} \\
&= \pi_{ij} \quad i, j = 1, 2
\end{aligned}$$

i.e.,  $p_{ij}$  is an unbiased estimator for  $\pi_{ij}$ ,  $i, j = 1, 2$ .

One component of the variance is readily obtained, namely

$$\begin{aligned}
\text{var}[E(p_{ij} | n_{io})] &= \text{var}\left[\frac{n_{io} + n_{o^*} \pi_{io}}{N} \frac{\pi_{ij}}{\pi_{io}}\right] \quad \text{from (4.2.3.6)} \\
&= \left(\frac{\pi_{ij}}{N \pi_{io}}\right)^2 n_{oo} \pi_{io} (1 - \pi_{io}) \quad (4.2.3.7) \\
&= \frac{n_{oo}}{N^2} \pi_{ij}^2 \frac{\pi_{i'o}}{\pi_{io}} \quad i, j = 1, 2 \quad i \neq i' .
\end{aligned}$$

The other component of  $\text{var}(p_{ij})$  is considerably more complex and involves a minor approximation, namely the expectation of the reciprocal of the binomial random variable,  $n_{io}$ . This approximation involves taking the expectation of the Taylor series expansion (through the

quadratic term) of  $f(n_{io}) = 1/n_{io}$  about the point  $E(n_{io}) = n_{oo}\pi_{io}$ .

The resulting approximation is given by the following:

$$\begin{aligned} E\left(\frac{1}{n_{io}}\right) &\doteq E\left[\frac{1}{n_{oo}\pi_{io}} - \frac{n_{io} - n_{oo}\pi_{io}}{(n_{oo}\pi_{io})^2} + \frac{(n_{io} - n_{oo}\pi_{io})^2}{(n_{oo}\pi_{io})^3}\right] \\ &= \frac{1}{n_{oo}\pi_{io}} + \frac{1 - \pi_{io}}{(n_{oo}\pi_{io})^2} \quad i = 1, 2 \quad (4.2.3.8) \end{aligned}$$

Thus, omitting the considerable algebra involved,

$$\begin{aligned} \text{var}(p_{ij} | n_{io}) &= \text{var}\left[\frac{n_{io} + n_{i^*}}{N} \frac{n_{ij}}{n_{io}} \mid n_{io}\right] \\ &= \frac{1}{(Nn_{io})^2} \text{var}[(n_{io} + n_{i^*})n_{ij} \mid n_{io}] \quad (4.2.3.9) \\ &= \frac{\pi_{ij}}{N^2 n_{io}^2 \pi_{io}} [n_{io}^2 \pi_{ij}' + n_{io} (2n_{o^*} \pi_{ij}' \pi_{io} + n_{o^*} \pi_{ij} \pi_{io} (1 - \pi_{io})) \\ &\quad + (n_{o^*} \pi_{ij}' \pi_{io} (n_{o^*} \pi_{io} + \pi_{i^* o}))] \quad i, j = 1, 2 \\ &\quad \quad \quad i \neq i', j \neq j' \end{aligned}$$

so that, using the approximation mentioned previously,

$$\begin{aligned} E[\text{var}(p_{ij} | n_{io})] &\doteq \left[ \frac{n_{oo}}{N^2} \frac{\pi_{ij} \pi_{ij}'}{\pi_{io}} + \frac{n_{o^*}}{N^2} \frac{\pi_{ij} (2\pi_{ij}' + \pi_{ij} (1 - \pi_{io}))}{\pi_{io}} \right. \\ &\quad \left. + \frac{n_{o^*}}{N^2} \frac{\pi_{ij} \pi_{ij}'}{\pi_{io}} (n_{o^*} \pi_{io} + \pi_{i^* o}) \left( \frac{1}{n_{oo}\pi_{io}} + \frac{1 - \pi_{io}}{(n_{oo}\pi_{io})^2} \right) \right]. \quad (4.2.3.10) \end{aligned}$$

Finally, combining (4.2.3.7) and (4.2.3.10) and simplifying, the

improved estimate of the variance is given by

$$\begin{aligned} \text{var}(p_{ij}) &= \frac{\pi_{ij}(1 - \pi_{ij})}{N} + \frac{n_{o*}}{Nn_{oo}} \frac{\pi_{ij}\pi_{ij'}}{\pi_{io}} & (4.2.3.11) \\ &+ \frac{n_{o*}}{N^2 n_{oo}^2} \frac{\pi_{ij}\pi_{ij'}\pi_{i'o}(N\pi_{io} + \pi_{i'o})}{\pi_{io}^3} \quad \begin{array}{l} i, j = 1, 2 \\ i \neq i' \quad j \neq j' \end{array} \end{aligned}$$

where the first two terms of (4.2.3.11) give the variance approximation previously derived (see (4.2.3.3)). (The covariance terms, although more complicated, could be derived in an analogous manner.)

### 4.3 2 x 2 Case with Both Margins Supplemented

#### 4.3.1 Methods of Estimation

The notation used for this case will be the obvious extension of the notation used previously (see Table 4.2.1.1). Thus,  $\pi_{*j}$ ,  $j = 1, 2$ , represent the probabilities over the second supplemented margin where it is assumed that  $\pi_{*j} = \pi_{1j} + \pi_{2j} = \pi_{oj}$ ; a sample of  $n_{*o}$  responses are observed for the second dimension of the table with  $n_{*j}$  responses at level  $j$ ,  $j = 1, 2$ ;  $N_{oj} = n_{oj} + n_{*j}$ ,  $j = 1, 2$ ; and  $N_r = \sum_{i=1}^2 N_{io}$ ,

$$N_c = \sum_{j=1}^2 N_{oj} \text{ so that}$$

$$N = n_{oo} + n_{o*} + n_{*o} = N_r + n_{*o} = N_c + n_{o*} .$$

A number of intuitive estimates for the individual cell probabilities were examined. These included the following:

$$(i) \quad \hat{\pi}_{ij} = w p_{ij}^{(r)} + (1-w) p_{ij}^{(c)}$$

where

$$p_{ij}^{(r)} = \frac{N_{io}}{N_r} \frac{n_{ij}}{n_{io}}$$

$$p_{ij}^{(c)} = \frac{N_{oj}}{N_c} \cdot \frac{n_{ij}}{n_{oj}}$$

with the following natural choices for w:

$$(1) \quad w = \frac{n_{o*}}{n_{o*} + n_{*o}}$$

$$(2) \quad w = \frac{n_{oo} + n_{o*}}{(n_{oo} + n_{o*}) + (n_{oo} + n_{*o})} = \frac{N_r}{N_r + N_c}$$

$$(ii) \quad \hat{\pi}_{ij} = \sqrt{p_{ij}^{(r)} p_{ij}^{(c)}}$$

$$(iii) \quad \hat{\pi}_{ij} = \frac{n_{ij}}{n_{oo}} \begin{pmatrix} 1 + \frac{n_{i*}}{n_{io}} \\ \frac{n_{o*}}{1 + \frac{n_{o*}}{n_{oo}}} \end{pmatrix} \begin{pmatrix} 1 + \frac{n_{*j}}{n_{oj}} \\ \frac{n_{*o}}{1 + \frac{n_{*o}}{n_{oo}}} \end{pmatrix} \quad i, j = 1, 2 .$$

However, none of these estimates provides an explicit solution for the MLE although numerical investigation suggests that the estimate in

(i) with  $w = \frac{n_{o*}}{n_{o*} + n_{*o}}$  gives a good approximation to the MLE.

Differentiating the log likelihood function (including the

restriction,  $\sum_{i=1}^2 \sum_{j=1}^2 \pi_{ij} = 1$ ) with respect to  $\pi_{ij}$ ,  $i, j = 1, 2$ , yields

the following normal equations:

$$\frac{n_{ij}}{p_{ij}} + \frac{n_{i*}}{p_{io}} + \frac{n_{*j}}{p_{oj}} - \lambda = 0 \quad i, j = 1, 2 . \quad (4.3.1.1)$$

But (4.3.1.1) implies that

$$n_{ij} + n_{i*} \frac{p_{ij}}{p_{io}} + n_{*j} \frac{p_{ij}}{p_{oj}} = \lambda p_{ij} \quad i, j = 1, 2 . \quad (4.3.1.2)$$

Summing (4.3.1.2) over  $i$  and  $j$  (and changing the order of summation where appropriate), it follows that  $\lambda = N$ . Thus,

$$p_{ij} = \frac{1}{N} \left( n_{ij} + n_{i*} \frac{p_{ij}}{p_{io}} + n_{*j} \frac{p_{ij}}{p_{oj}} \right) \quad i, j = 1, 2 \quad (4.3.1.3)$$

or, equivalently,

$$p_{ij} = \frac{n_{ij}}{N - \frac{n_{i*}}{p_{io}} - \frac{n_{*j}}{p_{oj}}} \quad i, j = 1, 2 \quad (4.3.1.4)$$

with either formulation requiring an iterated solution (see Chapter 7).

The method of iterative proportional fitting (IPF) would appear to be a reasonable alternative to iterating either (4.3.1.3) or (4.3.1.4). This intuitively-appealing method has been discussed and utilized by, among others, Mosteller (1968), Goodman (1968) and Bishop (1969) and

is based on certain results on maximum likelihood estimation derived by Birch (1963). Basically, the IPF scheme manipulates the rows and columns, respectively, of the given table in order to satisfy conditions on the margins while maintaining the original association within the table (as, for example, retaining the same cross-product ratio,  $\Delta = \ln[(\pi_{11}\pi_{22})/(\pi_{12}\pi_{21})]$ , in a 2 x 2 table). That this method yields MLE's of the cell probabilities is based on the realization by Birch that "the MLE's of the cell probabilities are determined uniquely by the marginal totals being equal to the MLE's of their expectations".

To apply IPF to the present problem, the question arises concerning the correct marginals to fit. It would appear that  $n_{1*}$  is relevant only to estimation of  $\pi_{10}$  and correspondingly for  $n_{*1}$  and that neither is relevant to the association since, for each, only one response was measured. But this implies that the MLE for  $\pi_{10}$  should be the estimate obtained for the case with one margin supplemented, namely,

$$p_{10} = \hat{\pi}_{10} = (N_{10}/N) = (n_{10} + n_{1*})/(n_{00} + n_{0*}) \text{ and similarly for } \pi_{01}.$$

However, we have from numerical investigation (i.e., the exact iteration for the MLE's of the cell probabilities and, through the invariance of MLE's, the MLE's of the marginal probabilities) that the  $p_{10}$  and  $p_{01}$  (and resulting  $\hat{\Delta}$ ) selected for the IPF are not the MLE's of the marginal probabilities (and  $\Delta$ )! Thus, since the proposed IPF scheme is not based on MLE's, it will not give MLE's of the cell probabilities (although this procedure does work fairly well for  $\Delta \doteq 0$ ). It appears that when both margins are supplemented, the estimation of the margins becomes complexly entangled with the association, and the individual effects are no longer separable.

Alternatively, let

$$\underline{n}'_G = (n_{11}, n_{12}, n_{21}, n_{1*}, n_{*1})$$

$1 \times 5$

$$\underline{\pi}'_s = (\pi_{11}, \pi_{12}, \pi_{21})$$

$1 \times 3$

$$X = \begin{bmatrix} n_{00} & 0 & 0 \\ 0 & n_{00} & 0 \\ 0 & 0 & n_{00} \\ n_{0*} & n_{0*} & 0 \\ n_{*0} & 0 & n_{*0} \end{bmatrix}$$

$5 \times 3$

Then, weighted least squares estimates,  $\tilde{\pi}_{ij}$ , obtained using the model

$$E(\underline{n}_G) = X\underline{\pi}_s \quad (4.3.1.5)$$

are given by

$$\tilde{\underline{\pi}}_s = (X' \tilde{V}^{-1} X)^{-1} X' \tilde{V}^{-1} \underline{n}_G$$

where

$$\tilde{V}^{-1} = \begin{bmatrix} \frac{1}{n_{00}} (D^{-1} + \frac{1}{n_{00}} J) & \underline{0} & \underline{0} \\ \underline{0}' & \frac{1}{n_{0*} \tilde{\pi}_{10} (1 - \tilde{\pi}_{10})} & 0 \\ \underline{0}' & 0 & \frac{1}{n_{*0} \tilde{\pi}_{01} (1 - \tilde{\pi}_{01})} \end{bmatrix}$$

$5 \times 5$

(4.3.1.6)

Performing the indicated matrix multiplication and simplifying yields

$$\begin{aligned}
 \tilde{\pi}_{ij} = \frac{1}{\tilde{N}} & \left[ n_{ij} + n_{i*} \frac{\tilde{\pi}_{ij}}{\tilde{\pi}_{i0}} + n_{*j} \frac{\tilde{\pi}_{ij}}{\tilde{\pi}_{0j}} + \frac{n_{0*}}{n_{00}} \frac{n_{ij} \tilde{\pi}_{ij'} - n_{ij'} \tilde{\pi}_{ij}}{\tilde{\pi}_{i0}} \right. \\
 & + \frac{n_{*0}}{n_{00}} \frac{n_{ij} \tilde{\pi}_{i'j} - n_{i'j} \tilde{\pi}_{ij}}{\tilde{\pi}_{0j}} + \tilde{\gamma} \left\{ \frac{n_{ij}}{\tilde{\pi}_{ij}} + \frac{\left( \frac{n_{00}}{n_{*0}} n_{*j} - n_{i'j} \right)}{\tilde{\pi}_{i'j}} \right. \\
 & \left. \left. + \frac{\left( \frac{n_{00}}{n_{0*}} n_{i*} - n_{ij'} \right)}{\tilde{\pi}_{ij'}} + \frac{\left( \frac{n_{00}}{n_{*0}} n_{*j} - \frac{n_{00}}{n_{0*}} n_{i'*} + n_{i'j'} \right)}{\tilde{\pi}_{i'j'}} \right\} \right] \\
 & \quad i, j = 1, 2 \quad i \neq i' \quad j \neq j'
 \end{aligned} \tag{4.3.1.7}$$

where

$$\tilde{N}_{\tilde{\gamma}} = N + n_{00} \tilde{\gamma} \sum_{i=1}^2 \sum_{j=1}^2 \left( \frac{1}{\tilde{\pi}_{ij}} \right)$$

with

$$\tilde{\gamma} = \frac{n_{0*} n_{*0}}{n_{00}} \cdot \frac{\prod_{i=1}^2 \prod_{j=1}^2 (\tilde{\pi}_{ij})}{\prod_{i=1}^2 (\tilde{\pi}_{i0}) \prod_{j=1}^2 (\tilde{\pi}_{0j})}$$

(4.3.1.7) also requires iterative procedures (see Chapter 7).

As before,  $p_{ij}$  and  $\tilde{\pi}_{ij}$  reduce to the ordinary MLE when  $n_{0*} = n_{*0} = 0$  and they reduce to the previous estimates when  $n_{*0} = 0 \neq n_{0*}$  (i.e., when one margin is supplemented). Also,  $p_{ij} = \text{MLE}$  is a stationary

solution of (4.3.1.7) (i.e., the  $p_{ij}$  satisfying (4.3.1.4) also satisfy the relations (4.3.1.7) for weighted least squares estimates). To show this, (4.3.1.3) must be used along with the relationships

$$\frac{n_{ij}}{\tilde{\pi}_{ij}} - \frac{n_{ij'}}{\tilde{\pi}_{ij'}} = \frac{n_{*j'}}{\tilde{\pi}_{0j'}} - \frac{n_{*j}}{\tilde{\pi}_{0j}} = \frac{n_{i'j}}{\tilde{\pi}_{i'j}} - \frac{n_{i'j'}}{\tilde{\pi}_{i'j'}} \quad (4.3.1.8)$$

$$\frac{n_{ij}}{\tilde{\pi}_{ij}} - \frac{n_{i'j'}}{\tilde{\pi}_{i'j'}} = \frac{n_{i'*}}{\tilde{\pi}_{i'o}} - \frac{n_{i*}}{\tilde{\pi}_{io}} = \frac{n_{ij'}}{\tilde{\pi}_{ij'}} - \frac{n_{i'j'}}{\tilde{\pi}_{i'j'}}$$

$$i, j = 1, 2 \quad i \neq i' \quad j \neq j' .$$

#### 4.3.2 Properties of the Estimators

Since the MLE's,  $p_{ij}$ ,  $i, j = 1, 2$ , have not been expressed in closed form, only asymptotic properties of these estimators will be investigated. Belonging to the class of BAN estimators, the MLE's are asymptotically unbiased as well as consistent and asymptotically efficient.

The asymptotic covariance matrix is given by the inverse of the information matrix, namely,

$$I^{-1} = \frac{\pi_{22}}{n_{00}} \begin{bmatrix} 1 + \frac{\pi_{22}}{\pi_{11}} + \alpha_r + \alpha_c & 1 + \alpha_r & 1 + \alpha_c \\ & 1 + \frac{\pi_{22}}{\pi_{12}} + \alpha_r & 1 \\ \text{(symmetric)} & & 1 + \frac{\pi_{22}}{\pi_{21}} + \alpha_c \end{bmatrix}^{-1} \quad (4.3.2.1)$$

$$= \frac{1}{N_Y} \begin{bmatrix} \pi_{11}(1-\pi_{11}) + \gamma_{r_1} + \gamma_{c_1} + \gamma & -\pi_{11}\pi_{12} - \gamma_{r_1} - \gamma & -\pi_{11}\pi_{21} - \gamma_{c_1} - \gamma \\ & \pi_{12}(1-\pi_{12}) + \gamma_{r_1} + \gamma_{c_2} + \gamma & -\pi_{12}\pi_{21} + \gamma \\ \text{(symmetric)} & & \pi_{21}(1-\pi_{21}) + \gamma_{r_2} + \gamma_{c_1} + \gamma \end{bmatrix}$$

where

$$\alpha_r = \frac{n_{0*}}{n_{00}} \frac{\pi_{22}}{\pi_{10}\pi_{20}}$$

$$\alpha_c = \frac{n_{*0}}{n_{00}} \frac{\pi_{22}}{\pi_{01}\pi_{02}}$$

$$\gamma_{r_i} = \frac{n_{0*}}{n_{00}} \frac{\pi_{i1}\pi_{i2}}{\pi_{i0}} \quad i = 1, 2$$

$$\gamma_{c_j} = \frac{n_{*0}}{n_{00}} \frac{\pi_{1j}\pi_{2j}}{\pi_{0j}} \quad j = 1, 2$$

$$N_Y = N + n_{oo} \gamma \sum_{i=1}^2 \sum_{j=1}^2 (1/\pi_{ij})$$

with

$$\gamma = \frac{n_{o*} n_{*o}}{n_{oo}} \frac{\prod_{i=1}^2 \prod_{j=1}^2 (\pi_{ij})}{\prod_{i=1}^2 (\pi_{io}) \prod_{j=1}^2 (\pi_{oj})}$$

A general form expressing the elements of both (4.2.3.4) and (4.3.2.1) as well as the covariances in the usual situation (i.e., without supplemented margins) is given by

$$\text{cov}(p_{ij}, p_{i'j'}) = \frac{1}{N_Y} [ \pi_{ij} (\delta_{ii'} \delta_{jj'} - \pi_{i'j'}) \tag{4.3.2.2}$$

$$- \delta_{ii'} (1 - 2\delta_{jj'}) \frac{n_{o*} \pi_{ij} \pi_{i(3-j)}}{n_{oo} \pi_{io}}$$

$$- \delta_{jj'} (1 - 2\delta_{ii'}) \frac{n_{*o} \pi_{ij} \pi_{(3-i)j}}{n_{oo} \pi_{oj}}$$

$$+ (1 - 2\delta_{ii'}) (1 - 2\delta_{jj'}) \gamma ]$$

$$i, i', j, j' = 1, 2$$

where

$$\delta_{kl} = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{if } k \neq l \end{cases}$$

= Kronecker delta

and with  $N_{\gamma}$ ,  $\gamma$  as defined in (4.3.2.1);  $n_{o^*} = 0 = n_{*o}$  for the unsupplemented case while  $n_{*o} = 0 \neq n_{o^*}$  for supplementation on the row response only.

#### 4.4 Extension of the Theory of Supplemented Margins

##### 4.4.1 The Case of One Margin Supplemented

The major difficulty in extending the results from the  $2 \times 2$  case to the  $r \times c$  case with one margin supplemented is one of notation. In general, if certain correspondences between the two cases are made, the extension is rather straightforward although most algebraically and notationally tedious. One such set of correspondences arises from the investigation of certain properties of the estimates of  $\pi_{ij}$ . This set includes the following:

<u>r x c case</u>	<u>2 x 2 case</u>	
(i) $\pi_{rc}$	$\pi_{22}$	
(ii) $(\pi_{rj_1}, \pi_{rj_2}, \dots, \pi_{rj_{c-2}})$ where $j_\alpha \neq c$	$\pi_{21}$	
(iii) $(\pi_{ij_1}, \pi_{ij_2}, \dots, \pi_{ij_{c-1}})$ where $i = 1, 2, \dots, r-1$ $j_\alpha \neq j$	$\pi_{1j'}$	(4.4.1.1)
(iv) $1 - \pi_{io}$	$\pi_{i'o}$ where $i' \neq i$	
(v) $1 - \pi_{oj}$	$\pi_{oj'}$ where $j' \neq j$	

The previous intuitive estimate,  $\hat{\pi}_{ij}$ ,  $i = 1, 2, \dots, r$ ;  $j = 1, 2, \dots, c$ , extended to the  $r \times c$  case is again the MLE,  $p_{ij}$ . Also, using conditional means and variances, the  $p_{ij}$  are unbiased with approximate variances given by the following:

$$\begin{aligned} \text{var}(p_{ij}) &= \frac{\pi_{ij}(1-\pi_{ij})}{N} + \frac{n_{o*}}{Nn_{oo}} \frac{\pi_{ij}(\pi_{io}-\pi_{ij})}{\pi_{io}} \\ &+ \frac{n_{o*}}{N^2 n_{oo}^2} \frac{\pi_{ij}(\pi_{io}-\pi_{ij})(1-\pi_{io})[(N-1)\pi_{io}+1]}{\pi_{io}^3} \end{aligned} \quad (4.4.1.2)$$

$i = 1, 2, \dots, r; \quad j = 1, 2, \dots, c .$

Alternatively, let

$$\underline{n}'_G = (\underline{n}'_1, \underline{n}'_2, \dots, \underline{n}'_r, \underline{n}'_*)$$

$1 \times [r(c+1) - 2]$

where

$$\underline{n}'_i = (n_{i1}, n_{i2}, \dots, n_{ic}) \quad i = 1, 2, \dots, r-1$$

$1 \times c$

$$\underline{n}'_r = (n_{r1}, n_{r2}, \dots, n_{r,c-1})$$

$1 \times (c-1)$

$$\underline{n}'_* = (n_{1*}, n_{2*}, \dots, n_{r-1,*})$$

$1 \times (r-1)$

$$\underline{n}'_{o*} = (n_{o*}, n_{o*}, \dots, n_{o*})$$

$1 \times c$

with

$$\underline{\pi}'_G = (\underline{\pi}'_1, \underline{\pi}'_2, \dots, \underline{\pi}'_r)$$

$1 \times (rc-1)$

where

$$\begin{matrix} \underline{\pi}'_i & = & (\pi_{i1}, \pi_{i2}, \dots, \pi_{ic}) & & i = 1, 2, \dots, r-1 \\ \text{1xc} & & & & \end{matrix}$$

$$\begin{matrix} \underline{\pi}'_r & = & (\pi_{r1}, \pi_{r2}, \dots, \pi_{r,c-1}) \\ \text{1x(c-1)} & & & & \end{matrix}$$

$$\begin{matrix} \underline{\pi}'_0 & = & (\pi_{10}, \pi_{20}, \dots, \pi_{r-1,0}) \\ \text{1x(r-1)} & & & & \end{matrix}$$

$$X = \begin{bmatrix} n_{00} I_c & & & & & & \underline{Q} \\ & n_{00} I_c & & & & & \\ & & \ddots & & & & \\ & & & \ddots & & & \\ \underline{Q}' & & & & n_{00} I_c & & \\ & & & & & n_{00} I_{c-1} & \\ \underline{n}'_{-0*} & & & & & & \underline{Q} \\ & & n'_{-0*} & & & & \\ & & & \ddots & & & \\ \underline{Q}' & & & & & & \\ & & & & & \underline{n}'_{-0*} & \\ & & & & & & 0 \end{bmatrix}$$

where

$I_\eta$  is a  $(\eta \times \eta)$  identity matrix

$\underline{Q}, \underline{Q}$  are vectors and matrices, respectively, of 0's.

Then the weighted least squares estimates,  $\tilde{\pi}_{ij}$ , obtained from the model

$$E(\underline{n}_G) = X \cdot \underline{\pi}_s$$

are given by

$$\tilde{\underline{\pi}}_s = (X' \tilde{V}^{-1} X)^{-1} X' \tilde{V}^{-1} \underline{n}_G \quad (4.4.1.3)$$

where

$$\tilde{V}^{-1} = \begin{bmatrix} \frac{1}{n_{00}} (D_{\tilde{\pi}_s}^{-1} + \frac{1}{n_{rc}} J_1) & 0 \\ 0' & \frac{1}{n_{0*}} (D_{\tilde{\pi}_0}^{-1} + \frac{1}{n_{r0}} J_2) \end{bmatrix}$$

[r(c+1)-2] x [r(c+1)-2]

with  $D_{\tilde{\pi}_s}$ ,  $D_{\tilde{\pi}_0}$ , and  $J_i$  defined correspondingly as in (4.2.2.8) and

(4.2.2.9). The considerable algebra could be carried out at a later time if warranted. It might be noted that the information matrix is again identical to the asymptotic covariance matrix of the weighted least squares estimates for several special  $r \times c$  cases that were considered.

It is of interest to note that all of the preceding theory for the case of supplementation on one margin (including two-way margins, etc.) can be extended rather readily to higher dimensional tables by associating all of the unsupplemented dimensions with the B response considered previously. For example, for a  $2 \times 2 \times 2$  table with the first response supplemented by  $n_{0**}$  additional observations, the

estimation problem is essentially that for a  $2 \times 4$  table with  $n_{0**}$  observations on the first dimension.

#### 4.4.2 The Case of Several Margins Supplemented

As was the case with supplementation for one margin, the major difficulty in extending the theory for the  $2 \times 2$  case with both margins supplemented to the  $r \times c$  case is notation. With the proper correspondences, it is largely an exercise in algebra and will not be carried out at this point.

The extension to higher dimensional tables is not nearly as obvious as it was for the case with supplementation over a single margin. Examples of the types of designs for supplementation in higher dimensional tables can be illustrated by the  $2 \times 2 \times 2$  case with responses A, B, and C. The margins available for supplementation include the one-way margins for A, B, and C as well as the two-way margins AB, AC, and BC. Examples of the possible supplementation designs include the following:

- (i) A (i.e., one margin supplemented)
- (ii) A, B (i.e., two one-way margins supplemented)
- (iii) AB (i.e., one two-way margin supplemented)
- (iv) A, AB (i.e., a one-way margin and an overlapping two-way margin supplemented)
- (v) A, B, AB (i.e., two one-way margins and a completely overlapping two-way margin supplemented)
- (vi) AB, AC (i.e., two two-way margins supplemented).

#### 4.5 Tests of Hypotheses

As noted in Chapter 3, testing various hypotheses for the more general case of supplemental data using a modification of the GSK program is essentially a two-stage procedure. This two-stage procedure will be illustrated here for  $2 \times 2$  tables with supplementation using the results derived in this chapter. However, as stated previously, the general formulation of the test procedure applies directly to  $r \times c$  tables and also to supplementation in higher dimensional tables.

The first stage in the two-stage test procedure consists of the following:

- (i) Input the observed frequencies such as are given in Table 4.2.1.1 for one margin supplemented (Case I) or in Table 4.5.1 for the case of both margins supplemented (Case II).
- (ii) Obtain estimates,  $\hat{\pi}_{ij}$ ,  $i, j = 1, 2$ , where  $\hat{\pi}_{22} = 1 - \hat{\pi}_{11} - \hat{\pi}_{12} - \hat{\pi}_{21}$ , corresponding to the unrestricted MLE for the unsupplemented case:

$$(1) \text{ Case I: Use the MLE's, } \hat{\pi}_{ij} = p_{ij} = \frac{N_{i0} \cdot n_{ij}}{N \cdot n_{i0}}$$

as given in (4.2.2.7).

- (2) Case II: Use either the iterated MLE's obtained from (4.3.1.3) or the iterated WLSE's obtained from (4.3.1.7). If a modification of the iteration program given in Chapter 7 is used, the latter estimate would be recommended since its iteration converges much more rapidly than that for the MLE.

(iii) Obtain an estimate,  $\hat{V}$ , of the asymptotic covariance matrix using the estimates derived in (ii):

(1) Case I: Use the obvious extension (which includes variances and covariances related to  $p_{22}$ ) of the inverse of the information matrix given in (4.2.3.4) with  $p_{ij}$  replacing  $\pi_{ij}$ ,  $i, j = 1, 2$ ;

(2) Case II: Proceed as in Case I but with the extension of (4.3.2.1) with  $\tilde{\pi}_{ij}$  replacing  $\pi_{ij}$ ,  $i, j = 1, 2$ . (See (4.3.2.2) for the details.)

The second stage consists of the following:

- (i) Input the A, K, X and C matrices required for the analysis as in the existing GSK program (see Section 3.5).
- (ii) Use the adjusted estimates of the cell probabilities as obtained in (ii) of the first stage in place of the usual estimates given in (1.4.1.2) along with the estimate of the asymptotic covariance matrix obtained in (iii) of the first stage in place of the usual estimate given in (1.4.1.6). With these adjusted estimates incorporating the information provided by the supplemented margins, proceed using the GSK program to obtain tests of marginal symmetry, uniformity, etc. Note that the program must be modified not only to incorporate the first stage but also to delete the usual steps for obtaining  $p_{ij}$ ,  $i, j = 1, 2$ , as in (1.4.1.2) and  $\hat{V}$  as in (1.4.1.6).

It should be noted that, since all rc parameters are used, it may be necessary, in order to avoid singularities, to effect a reparameterization by the proper selection of the A matrix.

As an example of the steps required for the two-stage testing procedure, consider the following "no factor, two response" situation with supplementation on both margins (see Table 4.5.1). The hypothesis

Table 4.5.1 Observed frequencies for a 2 x 2 contingency table with both margins supplemented

Response A Levels	Response B Levels		Sub-total ( $n_{i0}$ )	Supplementation on A ( $n_{i*}$ )	Total ( $N_{i0}$ )
	1	2			
1	1 ( $n_{11}$ )	2 ( $n_{12}$ )	3 ( $n_{10}$ )	5 ( $n_{1*}$ )	8 ( $N_{10}$ )
2	4 ( $n_{21}$ )	5 ( $n_{22}$ )	9 ( $n_{20}$ )	9 ( $n_{2*}$ )	18 ( $N_{20}$ )
Sub-total ( $n_{0j}$ )	5 ( $n_{01}$ )	7 ( $n_{02}$ )	12 ( $n_{00}$ )	14 ( $n_{0*}$ )	26 ( $N_T$ )
Supplementation on B ( $n_{*j}$ )	4 ( $n_{*1}$ )	6 ( $n_{*2}$ )	10 ( $n_{*0}$ )		
Total ( $N_{0j}$ )	9 ( $N_{01}$ )	13 ( $N_{02}$ )	22 ( $N_c$ )		36 ( $N$ )

of interest is one of marginal symmetry, i.e.,

$$H_0: \pi_{10} = \pi_{01} .$$

The test proceeds as follows:

Stage 1:

- (i) Input Table 4.5.1 in the proper format.
- (ii) Iterate for the weighted least squares estimates (see (7.3)) yielding

$$\begin{aligned} \tilde{\pi}_{11} &= 0.1010 & \tilde{\pi}_{12} &= 0.2069 \\ \tilde{\pi}_{21} &= 0.3046 & \tilde{\pi}_{22} &= 0.3875 . \end{aligned} \quad (4.5.1)$$

- (iii) Insert the estimates from (4.5.1) into the extended version of (4.3.2.1) yielding

$$\hat{V}_{4 \times 4} = \begin{bmatrix} 0.0059 & -0.0031 & -0.0030 & 0.0000 \\ & 0.0085 & -0.0003 & -0.0033 \\ & & 0.0109 & -0.0076 \\ \text{(symmetric)} & & & 0.0125 \end{bmatrix} . \quad (4.5.2)$$

Stage 2:

- (i) Input the appropriate A matrix, namely  $A = (0, 1, -1, 0)$  (cf. Grizzle et al. (1969), p. 493).
- (ii) Using (4.5.1) and (4.5.2), the revised GSK program would then compute the appropriate test statistic,  $X_{MS}^2$ , which, under  $H_0$ , is asymptotically  $\chi^2$  with D.F. = 1 where

$$\begin{aligned}
X_{MS}^2 &= (\underline{A\tilde{\pi}})' (\hat{A\tilde{V}A'})^{-1} (\underline{A\tilde{\pi}}) \\
&= (-.0977)^2 / (.0200) \\
&= 0.48
\end{aligned} \tag{4.5.3}$$

indicating marginal symmetry. Note that, due to the simplicity of (4.5.3) for this 2 x 2 case,  $\hat{A\tilde{V}A'}$  is explicitly given by

$$\begin{aligned}
\hat{\text{var}}(\tilde{\pi}_{12}) + \hat{\text{var}}(\tilde{\pi}_{21}) - 2 \hat{\text{cov}}(\tilde{\pi}_{12}, \tilde{\pi}_{21}) &= \\
&= \frac{1}{N_{\tilde{\gamma}}} [\tilde{\pi}_{12} + \tilde{\pi}_{21} - (\tilde{\pi}_{12} - \tilde{\pi}_{21})^2 + \frac{n_{o*}}{n_{oo}} (\frac{\tilde{\pi}_{11}\tilde{\pi}_{12}}{\tilde{\pi}_{1o}} + \frac{\tilde{\pi}_{21}\tilde{\pi}_{22}}{\tilde{\pi}_{2o}}) \\
&\quad + \frac{n_{*o}}{n_{oo}} (\frac{\tilde{\pi}_{11}\tilde{\pi}_{21}}{\tilde{\pi}_{o1}} + \frac{\tilde{\pi}_{12}\tilde{\pi}_{22}}{\tilde{\pi}_{o2}})] \\
&= 0.0201
\end{aligned} \tag{4.5.4}$$

providing a check for (4.5.3).

Tests for independence, total symmetry, equality of scored means, and uniformity (e.g.,  $\pi_{1o} = 1/2$ ) are also appropriate for this example. When the theory of estimation for supplemented margins is developed for more general situations (e.g., the "multi-factor, multi-response" case with supplementation), the full generality of the modified (or two-stage) GSK program can be utilized.

## 5. SUMMARY AND SUGGESTIONS FOR FURTHER RESEARCH

### 5.1 Summary

This research has utilized primarily the weighted least squares method of estimating functions of individual cell probabilities for contingency table representations of categorical data situations along with minimum Neyman  $\chi^2$  tests or equivalently tests using the quadratic forms arising in the familiar linear regression analyses. Grizzle et al. (1969) provide the theoretical framework for this approach. Maximum likelihood estimation is used where feasible for purposes of comparison. The minimum discrimination information and minimum  $\chi^2$  methods of estimation along with the Pearson  $\chi^2$ , Neyman-Pearson likelihood ratio, and minimum discrimination information test statistics are only briefly mentioned since a primary goal of this research is to investigate a most general method of analyzing categorical data with the linear model approach being the method selected.

In order to formulate appropriate hypotheses to be tested using the Grizzle, Starmer, Koch (GSK) approach, the underlying factor-response structure of the table should be identified. As noted in Bhapkar and Koch (1968a), the types of multi-dimensional contingency tables (and appropriate hypotheses) include "no factor, multi-response" tables (problems of independence); "uni-factor, multi-response" tables (effect of the factor on the responses, association between the responses); "multi-factor, uni-response" tables (homogeneity, "no interaction" between factors and the way they affect the response); "multi-factor, multi-response" tables (relationships among the responses and in the way the factors combine to affect the responses).

An additional model is recognized in this research and is referred to as a "mixed categorical data model" due to its analogy with the well-known mixed model in the analysis of variance. The experimental situation involves exposing each of  $n$  subjects to each of  $d$  levels of a given factor and classifying each of the  $d$  responses into one of  $r$  categories. The resulting data are represented in an  $r \times r \times \dots \times r$  contingency table of  $d$  dimensions.

The hypothesis of principal interest is equality of one-dimensional marginal distributions. If the  $r$  categories can be quantitatively scaled, attention is focused on the hypothesis of equality of the mean scores over the  $d$  first order marginals. Test statistics are developed in terms of minimum Neyman  $\chi^2$  or equivalently weighted least squares analysis of underlying linear models.

The analogy of the test for equality of mean scores in mixed categorical data models of order 2 to the matched pairs  $t$ -test is illustrated for a well-used example on eye-testing for employees in Royal Ordnance factories. It is indicated that tests for mixed categorical data models of higher order bear a strong resemblance to the Hotelling  $T^2$  procedures used with continuous data in mixed models. In addition, the application of the GSK approach to split plot contingency tables is illustrated with an example from Lessler (1962), and the extension of this research to complex mixed categorical data models is outlined.

In mixed categorical data models as well as the models given in Bhapkar and Koch (1968a), the assumption that the data is complete in the sense that every experimental unit is classified according to each

of the  $d$  dimensions of the table is often violated. The remainder of this research deals with methods of utilizing this "incomplete" (or "supplemental") data by extending the GSK approach as given in Grizzle et al. (1969).

It is assumed throughout that there is no interaction between subjects and the presence of supplemental data (i.e., for the responses classified, the joint marginal distributions for subjects with complete data and for those with supplemental data are the same). Conditional on the number of individuals with various types of complete or supplemental data being fixed by design a priori, the basic probability model is the product of several multinomials with marginal and/or cell probabilities as fundamental parameters.

The most detailed investigation is carried out for  $2 \times 2$  tables. For the case with supplementation on one margin, the logical intuitive estimates of the cell probabilities are shown to be the MLE's as well as the iterated weighted least squares estimates (WLSE's). Using conditional expectations, the MLE's are shown to be unbiased. Being MLE's, they are consistent and asymptotically efficient. The asymptotic covariance matrix is then obtained by inverting the information matrix. Since this is exactly the asymptotic covariance matrix of the WLSE's, if the iterated WLSE's are unique, they must be identical to the MLE's.

When both margins are supplemented, the MLE's and WLSE's must be obtained by iteration. For the computer program that was written, the WLSE's iteration converged considerably faster than did the one for the MLE's. It is noted that the iterative proportional fitting technique

of Mosteller (1968), Goodman (1968) and Bishop (1969) does not readily give MLE's for this case. Again, the asymptotic covariance matrix is identical for the MLE's and WLSE's. A general representation for the elements of the covariance matrix is given which yields, as special cases, the covariances of the estimates for the cases with supplementation on one margin and the usual case of complete data (i.e., no supplementation).

The extension of the theory to general  $r \times c$  tables as well as to higher dimensional tables is outlined.

The incorporation of these results into the general GSK framework for hypothesis testing is indicated as a two-stage test procedure. The first stage consists of the following:

- (i) Input the observed cell and marginal frequencies.
- (ii) Obtain estimates of the cell probabilities corresponding to the unrestricted MLE's for the case with complete data:
  - (1) If the table is fairly simple, obtain the MLE's. (This is the closest analogue to the unrestricted MLE's for the complete data version and is most consistent with tests using Wald statistics.)
  - (2) For most cases, compute the WLSE's by iterative techniques.
  - (3) If interested in estimating functions of the cell probabilities (say,  $A\pi$ ), obtain WLSE's of these expressions. (This is particularly appropriate in the mixed model situation.)

- (iii) Obtain an estimate,  $\hat{V}$  (or  $\hat{A}\hat{V}\hat{A}'$  for (3)), of the asymptotic covariance matrix using the estimates derived in (ii).

The second stage consists of the following:

- (i) Input the A, K, X, and C matrices required for the analysis as in the existing program.
- (ii) Use the adjusted (for supplemental data) estimates of the cell probabilities as obtained in (ii) of the first stage in place of the usual estimates along with the adjusted estimate of the asymptotic covariance matrix in (iii) of the first stage in place of the usual GSK estimate. Proceed with the testing of various hypotheses exactly as in the existing GSK framework.

A number of examples are presented to illustrate the more general categorical data models utilizing supplemental data.

## 5.2 Suggestions for Further Research

The present research uncovered a number of areas which might warrant further investigation. These include the following:

- (i) Generalization of the GSK program to efficiently handle the case of supplemental data.
- (ii) Development of the theory for supplemented margins in general multi-factor, multi-response situations.
- (iii) Consideration of iterative proportional fitting techniques for maximum likelihood estimation when more than one margin is supplemented.
- (iv) Investigation of various supplementation designs with a consideration of the cost of measurement.

- (v) Comparison of small sample properties of tests for contingency tables with and without supplementation.
- (vi) Power comparisons of tests for contingency tables with and without supplementation.
- (vii) Consideration of the effect of supplementation on various measures of association (e.g., Kendall's  $\tau$ , Goodman-Kruskal G for two-way tables).
- (viii) Investigation of the analysis of generalized mixed models using the GSK approach.

## 6. LIST OF REFERENCES

- Afifi, A. A. and Elashoff, R. M. 1966. Missing observations in multivariate statistics I. Review of the literature. *Journal of the American Statistical Association* 61: 595-605.
- Afifi, A. A. and Elashoff, R. M. 1967. Missing observations in multivariate statistics II. Point estimation in simple linear regression. *Journal of the American Statistical Association* 62: 10-29.
- Afifi, A. A. and Elashoff, R. M. 1969a. Missing observations in multivariate statistics III. Large sample analysis of simple linear regression. *Journal of the American Statistical Association* 64: 337-58.
- Afifi, A. A. and Elashoff, R. M. 1969b. Missing observations in multivariate statistics IV. A note on simple linear regression. *Journal of the American Statistical Association* 64: 359-65.
- Bartlett, M. S. 1935. Contingency table interactions. *Journal of the Royal Statistical Society (Supplement 2)*: 248-52.
- Berkson, J. 1954. Estimation by least squares and by maximum likelihood. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* 1: 1-11. University of California Press, Berkeley and Los Angeles.
- Berkson, J. 1968. Application of minimum logit  $\chi^2$  estimate to a problem of Grizzle with a notation on the problem of no interaction. *Biometrics* 24: 75-95.
- Bhapkar, V. P. 1961. Some tests for categorical data. *Annals of Mathematical Statistics* 32: 72-83.
- Bhapkar, V. P. 1965. Categorical data analogs of some multivariate tests. *S. N. Roy Memorial Volume*, University of North Carolina Press, Chapel Hill, N. C.
- 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: 228-35.
- Bhapkar, V. P. 1968. On the analysis of contingency tables with a quantitative response. *Biometrics* 24: 329-38.
- Bhapkar, V. P. and Koch, G. G. 1968a. Hypotheses of 'no interaction' in multi-dimensional contingency tables. *Technometrics* 10: 107-23.

- Bhapkar, V. P. and Koch, G. G. 1968b. On the hypotheses of 'no interaction' in contingency tables. *Biometrics* 24: 567-94.
- Birch, M. W. 1963. Maximum likelihood in three-way contingency tables. *Journal of the Royal Statistical Society (B)* 25: 220-33.
- Bishop, Y. M. M. 1969. Full contingency tables, logits, and split contingency tables. *Biometrics* 25: 383-99.
- Bishop, Y. M. M. and Fienberg, S. E. 1969. Incomplete two-dimensional contingency tables. *Biometrics* 25: 119-28.
- Blumenthal, S. 1968. Multinomial sampling with partially categorized data. *Journal of the American Statistical Association* 63: 542-51.
- Cochran, W. G. 1950. The comparison of percentages in matched samples. *Biometrika* 37: 256-66.
- Cochran, W. G. 1952. The  $\chi^2$  test of goodness of fit. *Annals of Mathematical Statistics* 23: 315-45.
- Cochran, W. G. 1954. Some methods of strengthening the common  $\chi^2$  test. *Biometrics* 10: 417-51.
- Cochran, W. G. 1963. *Sampling Techniques*. John Wiley and Sons, New York.
- Friedman, M. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association* 32: 675-99.
- Good, I. J. 1963. Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *Annals of Mathematical Statistics* 34: 911-34.
- Goodman, L. A. 1968. The analysis of cross-classified data: Independence, quasi-independence, and interactions in contingency tables with or without missing entries. *Journal of the American Statistical Association* 63: 1091-131.
- Goodman, L. A. 1970. The multivariate analysis of qualitative data: Interactions among multiple classifications. *Journal of the American Statistical Association* 65: 226-56.
- Grizzle, J. E., Starmer, C. F. and Koch, G. G. 1969. Analysis of categorical data by linear models. *Biometrics* 25: 489-504.
- Ireland, C. T. and Kullback, S. 1968a. Contingency tables with given marginals. *Biometrika* 55: 179-88.

- Ireland, C. T. and Kullback, S. 1968b. Minimum discrimination information estimation. *Biometrics* 24: 707-13.
- Ireland, C. T., Ku, H. H. and Kullback, S. 1969. Symmetry and marginal homogeneity of an  $r \times r$  contingency table. *Journal of the American Statistical Association* 64: 1323-41.
- Kastenbaum, M. A. and Lamphiear, D. E. 1959. Calculation of chi-square to test the no three-factor interaction hypothesis. *Biometrics* 15: 107-15.
- Kleinbaum, D. G. 1970. A general method for obtaining test criteria for multivariate linear models with more than one design matrix and/or incomplete in response variates. Unpublished Ph.D. thesis, Department of Statistics, University of North Carolina at Chapel Hill.
- Koch, G. G. and Sen, P. K. 1968. Some aspects of the statistical analysis of the 'mixed model'. *Biometrics* 24: 27-48.
- Ku, H. H. and Kullback, S. 1968. Interaction in multidimensional contingency tables: An information theoretic approach. *Journal of Research of the National Bureau of Standards - Mathematical Sciences* 72B: 159-99.
- Kullback, S. 1959. *Information Theory and Statistics*. John Wiley and Sons, New York.
- Kullback, S., Kupperman, M. and Ku, H. H. 1962. Tests for contingency tables and Markov chains. *Technometrics* 4: 573-608.
- Lessler, K. J. 1962. The anatomical and cultural dimensions of sexual symbols. Unpublished Ph.D. thesis, Michigan State University, East Lansing.
- Lewis, J. A. 1968. A program to fit constants to multiway tables of quantitative and quantal data. *Applied Statistics* 17: 33-41.
- Mantel, N. 1963. Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association* 58: 690-700.
- Mantel, N. and Haenszel, W. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 22: 719-48.
- Miettinen, O. S. 1968. The matched pairs design in the case of all-or-none responses. *Biometrics* 24: 339-52.
- Miettinen, O. S. 1969. Individual matching with multiple controls in the case of all-or-none responses. *Biometrics* 25: 339-55.

- Miettinen, O. S. 1970. Estimation of relative risk from individually matched series. *Biometrics* 26: 75-87.
- Mosteller, F. 1968. Association and estimation in contingency tables. *Journal of the American Statistical Association* 63: 1-28.
- Neyman, J. 1949. Contribution to the theory of the  $\chi^2$  test. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley and Los Angeles. Pp. 239-73.
- Pearson, E. S. 1947. The choice of statistical tests illustrated on the interpretation of data classed in a 2x2 table. *Biometrika* 34: 139-67.
- Rao, C. R. 1963. Criteria of estimation in large samples, pp. 345-62. In C. R. Rao (ed.), *Contributions to Statistics*. Pergamon Press, Oxford, New York.
- Roy, S. N. and Kastenbaum, M. A. 1956. On the hypothesis of no interaction in a multiway contingency table. *Annals of Mathematical Statistics* 27: 749-57.
- Scheffe, H. 1959. *The Analysis of Variance*. John Wiley and Sons, New York.
- Stuart, A. 1955. A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika* 42: 412-16.
- Wald, A. 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* 54: 426-82.
- Wilks, S. S. 1946. Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *Annals of Mathematical Statistics* 17: 257-81.

## 7. APPENDIX

This chapter contains the iteration scheme for the MLE's ( $p_{ij}$ ) and WLSE's ( $\tilde{\pi}_{ij}$ ) for the  $2 \times 2$  case with both margins supplemented. The expressions which must be iterated are as follows:

$$p_{ij} = \frac{1}{N} \left( n_{ij} + n_{i*} \frac{p_{ij}}{p_{io}} + n_{*j} \frac{p_{ij}}{p_{oj}} \right) \quad i, j = 1, 2 \quad (7.1)$$

$$\begin{aligned} \tilde{\pi}_{ij} = \frac{1}{\tilde{Y}} \left[ n_{ij} + n_{i*} \frac{\tilde{\pi}_{ij}}{\tilde{\pi}_{io}} + n_{*j} \frac{\tilde{\pi}_{ij}}{\tilde{\pi}_{oj}} + \frac{n_{oo}}{n_{oo}} \frac{n_{ij} \tilde{\pi}_{i'j'} - n_{i'j} \tilde{\pi}_{ij}}{\tilde{\pi}_{io}} \right. \\ \left. + \frac{n_{*o}}{n_{oo}} \frac{n_{ij} \tilde{\pi}_{i'j'} - n_{i'j} \tilde{\pi}_{ij}}{\tilde{\pi}_{oj}} + \tilde{Y} \left\{ \frac{n_{ij}}{\tilde{\pi}_{ij}} + \frac{\left( \frac{n_{oo}}{n_{*o}} n_{*j} - n_{i'j'} \right)}{\tilde{\pi}_{i'j'}} \right. \right. \\ \left. \left. + \frac{\left( \frac{n_{oo}}{n_{o*}} n_{i*} - n_{ij'} \right)}{\tilde{\pi}_{ij'}} + \frac{\left( \frac{n_{oo}}{n_{*o}} n_{*j} - \frac{n_{oo}}{n_{o*}} n_{i'*} + n_{i'j'} \right)}{\tilde{\pi}_{i'j'}} \right\} \right] \end{aligned} \quad (7.2)$$

$$i, j = 1, 2 \quad i \neq i' \quad j \neq j'$$

where

$$\tilde{N}_{\tilde{Y}} = N + n_{oo} \tilde{Y} \sum_{i=1}^2 \sum_{j=1}^2 \left( \frac{1}{\tilde{\pi}_{ij}} \right)$$

$$\tilde{Y} = \frac{n_{o*} n_{*o} \prod_{i=1}^2 \prod_{j=1}^2 (\tilde{\pi}_{ij})}{n_{oo}^2 \prod_{i=1}^2 (\tilde{\pi}_{io}) \prod_{j=1}^2 (\tilde{\pi}_{oj})}$$

The Call-A-Computer program written for simultaneously iterating (7.1) and (7.2) is as follows:

```

010  DIMENSION PA(2,2),N(2,2),D(2,2),MR(2),MC(2),A(2,2),B(2),
020  + C(2),X(2,2),Y(2),Z(2),PX(2,2)
030  INPUT,((PA(I,J),J=1,2),I=1,2),((N(I,J),J=1,2),I=1,2),
040  + (MR(I),I=1,2),(MC(J),J=1,2),NT,IR,IC,NTT,((PX(I,J),J=1,2),
050  + I=1,2)
060  WIR=IR; WIC=IC; WNT=NT
070  WR=WIR/WNT; WS=WIC/WNT
080  6 DO 2 I=1,2; DO 2 J=1,2
090  A(I,J)=PA(I,J); 2 X(I,J)=PX(I,J)
100  DO 3 I=1,2; B(I)=A(I,1)+A(I,2)
110  3 Y(I)=X(I,1)+X(I,2)
120  DO 4 J=1,2
130  C(J)=A(1,J)+A(2,J)
140  4 Z(J)=X(1,J)+X(2,J)
150  E=(WR*WS*X(1,1)*X(1,2)*W(2,1)*X(2,2))/(Y(1)*Y(2)*Z(1)*Z(2))
160  F=NTT+(NT*E*(X(1,1)*X(1,2)*X(2,1)+X(1,1)*X(1,2)*X(2,2)+
170  + X(1,1)*X(2,1)*X(2,2)+X(1,2)*X(2,1)*X(2,2))/(X(1,1)*X(1,2)*
180  + X(2,1)*X(2,2))
190  DO 5 I=1,2; DO 5 J=1,2
200  PA(I,J)=(N(I,J)+MR(I)*A(I,J)/B(I)+MC(J)*A(I,J)/C(J))/NTT
210  K=3-I; L=3-J
220  PX(I,J)=(N(I,J)+MR(I)*X(I,J)/Y(I)+MC(J)*X(I,J)/Z(J)+
230  + WR*(N(I,J)*X(I,L)-N(I,L)*X(I,J))/Y(I)+WS*(N(I,J)*X(K,J)-
240  + N(K,J)*X(I,J))/Z(J)+E*(N(I,J)/X(I,J)+(MC(J)/WS-N(K,J))/
250  + X(K,J)+(MR(I)/WR-N(I,L))/X(I,L)+(MC(J)/WS-MR(K)/WR+
260  + N(K,L))/X(K,L))/F
270  5 D(I,J)=PA(I,J)-PX(I,J)
280  PRINT,↑ "ML=",((PA(I,J),J=1,2),I=1,2)," LS=",((PX(I,J),
290  + J=1,2),I=1,2)," D=",((D(I,J),J=1,2),I=1,2)
300  GO TO 6; STOP; END

```

Using Table 4.5.1 as an example and starting with the unrestricted MLE's without supplementation as initial estimates of the  $p_{ij}$  and  $\tilde{\pi}_{ij}$ ,  $i, j = 1, 2$  (i.e.,  $p_{11}^{(0)} = 1/12 = 0.0833 = \tilde{\pi}_{11}^{(0)}$ ;  $p_{12}^{(0)} = 0.1667 = \tilde{\pi}_{12}^{(0)}$ ;  $p_{21}^{(0)} = 0.3333 = \tilde{\pi}_{21}^{(0)}$ ;  $p_{22}^{(0)} = 0.4167 = \tilde{\pi}_{22}^{(0)}$ ), the iteration proceeds as shown in Table 7.1.

Table 7.1 Iterative solution for the MLE's ( $p_{ij}$ ) and WLSE's ( $\tilde{\pi}_{ij}$ ) for Table 4.5.1

Stage in Iteration	$p_{11}$	$\tilde{\pi}_{11}$	$p_{12}$	$\tilde{\pi}_{12}$	$p_{21}$	$\tilde{\pi}_{21}$	$p_{22}$	$\tilde{\pi}_{22}$	$D_M^a$
1	.0963	.1016	.1958	.2063	.3111	.3041	.3968	.3881	.01050
2	.0998	.1011	.2037	.2068	.3058	<u>.3046</u>	.3906	<u>.3875</u>	.00310
3	.1008	<u>.1010</u>	.2059	<u>.2069</u>	.3047	.3046	.3886	.3875	.00110
4	.1011	.1010	.2065	.2069	.3045	.3046	.3880	.3875	.00044
5	.1011	.1010	.2067	.2069	.3045	.3046	.3877	.3875	.00020
6	.1011	.1010	.2068	.2069	.3045	.3046	.3876	.3875	.00010
7	.1011	.1010	.2068	.2069	.3045	.3046	.3876	.3875	.00006
8	.1011	.1010	.2068	.2069	.3045	.3046	.3876	.3875	.00004
9	.1011	.1010	<u>.2069</u>	.2069	.3045	.3046	.3876	.3875	.00003
10	<u>.1010</u>	.1010	.2169	.2069	.3045	.3046	.3876	.3875	.00002
11	.1010	.1010	.2069	.2069	.3045	.3046	<u>.3875</u>	.3875	.00002
12	.1010	.1010	.2069	.2069	<u>.3046</u>	.3046	.3875	.3875	.00001

$$^a D_M = \max |p_{ij} - \tilde{\pi}_{ij}|.$$

For this example, the iterated cell probabilities are given by the following:

$$\begin{aligned} p_{11} &= 0.1010 = \tilde{\pi}_{11} & p_{12} &= 0.2069 = \tilde{\pi}_{12} \\ p_{21} &= 0.3046 = \tilde{\pi}_{21} & p_{22} &= 0.3875 = \tilde{\pi}_{22} \end{aligned} \quad (7.3)$$

In this case (and in each of a number of other examples), the iteration for the WLSE's converged considerably faster than did the iteration for the MLE's.