ABSTRACT

A MULTIVARIATE APPROACH TO TESTING FOR HOMOGENEITY OF CORRELATED PROPORTIONS

By

Barry M. Katz

A design which frequently appears in behavioral science re-
search is one in which a single group of subjects is classified into
unordered categories by multiple raters. A test of homogeneity of
classification among the raters can be used to determine rater bias.
Analyzing attitudinal change over time, over objects, or over treat-
ments is also frequently of concern in the behavioral sciences. These
illustrations are specific cases of the more general mixed categorical
data model of order-d which is characterized by a situation in which
n randomly chosen subjects from some homogeneous population are
measured on an r-valued categorical dependent variable under d dif-
ferent conditions. The hypothesis of general interest is whether the
distribution of responses is the same under the d different condi-
tions. Koch and Reinfurt (1971) stated that such a hypothesis best
examines the relative effects of the d conditions on the dependent
measure of interest.

The probable lack of independence among measures or samples
precludes the valid use of such standard techniques as the chi-square
test of homogeneity. A technique which accounts for the correlated
nature of the resulting distributions is needed if the hypothesis of
homogeneity is to be tested validly.

The probability model associated with the mixed categorical data model of order-d is an $r \times r \times ... \times r$ contingency table of d dimensions whose cell frequencies are characterized by a multinomial distribution of r^d cell probabilities and a sample size of n. Testing the hypothesis of homogeneity of the d correlated distributions then reduces to testing for marginal homogeneity in the d dimensional table.

Four different statistical approaches to the problem of testing for homogeneity of the marginal distributions in the mixed categorical data model were examined and compared along theoretical lines. The four approaches examined were the X_S^2 statistic of Stuart (1955), a quadratic form in the differences of the marginal proportions; the X_M^2 statistic of Madansky (1963), a statistic based on the likelihood ratio criterion; the X_{GSK}^2 statistic of Koch and Reinfurt (1971), a statistic based on weighted least squares; the X_I^2 statistic of Ireland et al. (1969), a statistic based on minimum discrimination information estimation.

All four approaches were shown to belong to the same general class of large sample chi-square statistics. Each of the techniques was shown to be based on the use of BAN estimators and the four techniques were shown to be asymptotically equivalent. An explicit algebraic relationship between X_S^2 and X_{GSK}^2 was also demonstrated.

A fifth statistic, χ^2, algebraically equal to X_{GSK}^2, was proposed. The χ^2 statistic has the advantage that it does not require a knowledge of linear models for its understanding as does the X_{GSK}^2 statistic. The limiting distribution of χ^2 was shown to be chi-square with $(d-1)(r-1)$ degrees of freedom. A detailed set of

computational formulas were derived and a program written in Fortran IV to calculate the χ^2 statistic was given.

The development of two different techniques for generating confidence intervals for contrasts involving the marginal proportions was also given. One of the procedures, a simultaneous procedure, was developed along the lines of the results of Scheffé (1959) and Goodman (1964). A second technique was developed based on the Bonferroni inequality.

The behavior of the χ^2 statistic in the finite sample situation was examined by the method of simulation. The data were generated in groups of 2000 samples of a given size (n). For this study the values of n were chosen to yield average expected cell frequencies ($\bar{n}$) of 3, 5, 10, 20, 40, and 60. For each of the 3 X 3, 4 X 4, and 5 X 5 contingency tables, five different null distributions were considered and three different non-null distributions were considered. For the 3 X 3 X 3 table four null and three non-null distributions were considered.

Empirical estimates of the actual significance level and power of the χ^2 procedure were found by counting the number of rejections out of 2000 using the theoretical cutoff levels of α = .01, .05, .10 of the central chi-square distribution and were compared to their respective theoretical values. Empirical estimates of the actual significance level and power of χ^{2*}, a slight variant of the χ^2 statistic were also found. The χ^2 and χ^{2*} statistics are related by the equation

$$\chi^{2*} = \frac{\chi^2}{1 + \frac{1}{n}\chi^2} \quad .$$

It was found that for $\bar{n} \geq 10$ both statistics approximated their respective asymptotic behaviors quite well. When $\bar{n} = 5$ for those null distributions in which most of the off-diagonal cell expectations were two or greater, χ^2 was quite liberal while χ^{2*} approximated its limiting distribution quite closely. For null distributions in which most of the off-diagonal cell expectations were one or less, χ^{2*} was extremely conservative while χ^2 was slightly conservative. For $\bar{n} = 3$, values of actual alpha were found to be more extreme than those values found at $\bar{n} = 5$. In cases where the χ^2 procedure was found to be liberal at $\bar{n} = 5$, χ^2 was more liberal at $\bar{n} = 3$ and in cases where the χ^2 procedure was found to be conservative at $\bar{n} = 5$, χ^2 was more conservative at $\bar{n} = 3$. The χ^{2*} statistic was more conservative at $\bar{n} = 3$ than it was at $\bar{n} = 5$.

The χ^2 procedure was found to be somewhat more powerful than the χ^{2*} procedure for $\bar{n} = 5$. A series of guidelines were set forth based on the findings of the simulation study.

A MULTIVARIATE APPROACH TO TESTING FOR
HOMOGENEITY OF CORRELATED PROPORTIONS

By

Barry M. Katz

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Personnel Services,
and Educational Psychology

1975

TO THE MEMORY OF MY FATHER

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Questions about the homogeneity of multiple populations are common in behavioral research. When continuous variables are used, the questions may focus on population homogeneity with respect to mean level of performance or variability in performance. When categorical variables are used, the questions deal with the equivalence of the populations with respect to a variable of classification.

If multiple populations have been sampled independently and categorized independently with respect to the classification variable, then standard data analytic techniques can be used to determine whether the respective populations are homogeneous. On the other hand, if the samples have been matched (either by pairing or blocking of subjects across the samples or by repeated measurement on a single sample of subjects), the corresponding data analytic techniques to examine the question of homogeneity are not well-known in the behavioral sciences nor have their small sample properties been investigated.

This dissertation develops data analysis techniques to examine the question of homogeneity in matched populations, studies the small sample properties of these techniques, and compares these techniques with other competing approaches that have recently appeared

in the statistical literature.

To orient the reader to the design and analysis techniques presented in this thesis, the independent sample and the matched sample designs are briefly compared and the need for a separate matched sample test is explained.

The chi-square test of homogeneity is a categorical data analytic technique used to analyze data from two-way contingency tables in which one margin is fixed and the other margin is not. The corresponding experimental situation involves choosing a random sample of subjects from each of d different underlying populations and classifying each subject into one of r categories. The sample sizes chosen need not be equal but there must be independence both within and between the samples. The hypothesis of interest is whether the d populations are homogeneous with respect to the r categories of the variable of classification. The chi-square test of homogeneity is analogous in some respects to the one-way ANOVA used in the analysis of continuous data. A data example to illustrate the chi-square test of homogeneity follows.

A survey of community needs was conducted in a western Michigan county. The county was subdivided into four geographic regions of varying size, and samples of size 23, 25, 93, and 235 were randomly chosen from the adult populations of the respective four regions. Each respondent was asked to rate the degree to which he felt it was necessary to solve racial problems. The three categories of response were: (1) urgent, (2) important, (3) un-important. The survey yielded the following data given in Table 1-1.

TABLE 1-1

Data Illustration for the Chi-Square Test of Homogeneity

Regions		(1) Urgent	(2) Important	(3) Unimportant	
Region A	(1)	2	11	10	23
Region B	(2)	9	13	3	25
Region C	(3)	20	46	27	93
Region D	(4)	86	104	45	235
		117	174	85	

The row totals for these data are fixed quantities corresponding to the sample sizes used in the study, while the column totals are random, with their size determined by the respondents' replies. The hypothesis of interest is whether the four regions are homogeneous with respect to the distribution on the dependent variable. If P_{qj} denotes the probability of someone from region q responding with category j, then the null hypothesis of homogeneity can be stated in the form

$$H_0: \quad P_{1j} = P_{2j} = P_{3j} = P_{4j} \quad \text{for} \quad j = 1,2,3. \tag{1-1}$$

More generally if independent random samples are taken from d different populations and each subject is classified into one of r possible categories, the null hypothesis of homogeneity of the resulting d distributions can be written as

$$H_0: \quad P_{1j} = P_{2j} = \ldots = P_{dj} \quad \text{for} \quad j = 1,2,\ldots,r . \tag{1-2}$$

Let n_{qj} denote the number of observations from sample q which fall in category j. Let n_q, n_{0j}, n, $\hat{P}_{0j}$, and $\hat{P}_{qj}$ be defined as follows

$$n_q = \sum_{j=1}^{r} n_{qj} \quad \text{for} \quad q = 1,2,\ldots,d \tag{1-3}$$

$$n_{0j} = \sum_{q=1}^{d} n_{qj} \quad \text{for} \quad j = 1,2,\ldots,r \tag{1-4}$$

$$n = \sum_{q=1}^{d} n_q \tag{1-5}$$

$$\hat{P}_{0j} = \frac{n_{0j}}{n} \quad \text{for} \quad j = 1,2,\ldots,r \tag{1-6}$$

$$\hat{P}_{qj} = \frac{n_{qj}}{n_q} \quad \text{for} \quad j = 1,2,\ldots,r; \; q = 1,2,\ldots,d \tag{1-7}$$

where n_q is the size of the qth sample, n_{0j} is the total number of observations falling into category j from the combined samples, and n is the size of the combined samples. The formula in (1-6) is an estimate of the probability of an observation falling into category j under the assumption that this probability is the same for each of the populations; that is $\hat{P}_{0j}$ is computed under the assumption that the null hypothesis in (1-2) is true. The formula in (1-7) is an estimate of the probability that an observation from population q will fall into category j.

A statistic which can be used to test the hypothesis given in (1-2) is

$$C^2 = \sum_{q=1}^{d} \sum_{j=1}^{r} \frac{(n_{qj} - n_q \hat{P}_{0j})^2}{n_q \hat{P}_{0j}} \; . \tag{1-8}$$

The statistic in (1-8) has the general structure

$$c^2 = \sum_{q=1}^{d} \sum_{j=1}^{r} \frac{(O_{qj} - E_{qj})^2}{E_{qj}} \tag{1-9}$$

where O_{qj} represents the actual number of observations in the qth sample falling into category j and E_{qj} represents the number of observations to be expected to fall into category j for the qth sample under the assumption of the null hypothesis given in (1-2). Large values of the c^2 statistic imply that the actual cell frequencies vary widely from those expected under the null hypothesis. Thus the null hypothesis in (1-2) is rejected for large values of the c^2 statistic. When the sample sizes are suitably large it is known that the c^2 statistic has approximately a chi-square distribution with $(d-1)(r-1)$ degrees of freedom under the null hypothesis. For tests performed at the α level of significance, H_0 is rejected whenever the value of the c^2 statistic exceeds the $(1-\alpha)$th quantile of the chi-square distribution with $(d-1)(r-1)$ degrees of freedom.

An alternate way of writing the statistic c^2, which sometimes appears in the literature is given in (1-10)

$$c^2 = \sum_{q=1}^{d} \sum_{j=1}^{r} \frac{n_q(\hat{P}_{qj} - \hat{P}_{0j})^2}{\hat{P}_{0j}} . \tag{1-10}$$

To use the chi-square distribution validly as the reference distribution for the c^2 statistic, one of the assumptions which must be made is that responses across samples are made independently. One of the reasons for this assumption is that the c^2 statistic

does not make any allowances for covariation between random variables of the form $\hat{P}_{qj}$ and $\hat{P}_{q'k}$ for $q \neq q'$. That is, the c^2 statistic assumes that

$$Cov(\hat{P}_{qj}, \hat{P}_{q'k}) = 0 \quad \text{for all} \quad q \neq q', \ j,k = 1,2,\ldots,r. \quad (1-11)$$

If responses between samples are not independent, the random variables of the form $\hat{P}_{qj}$ and $\hat{P}_{q'k}$ will be correlated. Since the c^2 statistic makes no provision for this correlation an invalid test results.

The chi-square test of homogeneity is considered for the special case where $d = r = 2$ to illustrate the need for between sample independence. The 2×2 table corresponds to an experimental situation in which samples of sizes n_1 and n_2 are randomly selected from two respective underlying populations. The subjects are then classified on some dichotomous variable. The null hypothesis of homogeneity can be put in the form

$$H_0: \quad P_{1j} = P_{2j} \quad \text{for} \quad j = 1,2 \quad (1-12)$$

or equivalently in the form

$$H_0: \quad P_{11} = P_{21} \quad (1-13)$$

since the equality in (1-13) implies the equality in (1-12) for the case when $j = 2$.

To test H_0 the c^2 statistic given in (1-10) is computed for the special case when $d = r = 2$ and is given by

$$c^2 = \sum_{q=1}^{2} \sum_{j=1}^{2} \frac{n_q (\hat{P}_{qj} - \hat{P}_{0j})^2}{\hat{P}_{0j}} . \tag{1-14}$$

Making use of the equalities given in (1-15), which hold for the special case under consideration,

$$\hat{P}_{12} = 1 - \hat{P}_{11}, \quad \hat{P}_{22} = 1 - \hat{P}_{21}, \quad \hat{P}_{02} = 1 - \hat{P}_{01} \tag{1-15}$$

it can be shown through some algebraic simplification that the expression for c^2 in (1-14) can be rewritten in the form

$$c^2 = \frac{(\hat{P}_{11} - \hat{P}_{21})^2}{\frac{\hat{P}_{01}(1 - \hat{P}_{01})}{n_1} + \frac{\hat{P}_{01}(1 - \hat{P}_{01})}{n_2}} . \tag{1-16}$$

Because $c^2 \approx \chi^2$ with 1 degree of freedom

$$C = \frac{(\hat{P}_{11} - \hat{P}_{21})}{\sqrt{\frac{\hat{P}_{01}(1 - \hat{P}_{01})}{n_1} + \frac{\hat{P}_{01}(1 - \hat{P}_{01})}{n_2}}} \approx N(0,1). \tag{1-17}$$

The statistic in (1-17) can be written in the form

$$C = \frac{(\hat{P}_{11} - \hat{P}_{21}) - E_{H_0}(\hat{P}_{11} - \hat{P}_{21})}{\sqrt{\frac{\hat{P}_{01}(1 - \hat{P}_{01})}{n_1} + \frac{\hat{P}_{01}(1 - \hat{P}_{01})}{n_2}}} \approx N(0,1) \tag{1-18}$$

because $E_{H_0}(\hat{P}_{11} - \hat{P}_{21})$ is 0 by the unbiasedness of the estimators $\hat{P}_{11}$ and $\hat{P}_{21}$ and the specification of the null hypothesis, $P_{11} = P_{21}$. The C statistic thus has the general form

$$\frac{\hat{\theta} - E_{H_0}(\hat{\theta})}{\sqrt{\hat{V}ar_{H_0}(\hat{\theta})}} \qquad\qquad (1-19)$$

where $\hat{\theta}$ is an estimate of a parameter θ. Consequently the estimated variance of $\hat{P}_{11} - \hat{P}_{21}$ under the null hypothesis is given by

$$\hat{V}ar_{H_0}(\hat{P}_{11} - \hat{P}_{21}) = \frac{\hat{P}_{01}(1 - \hat{P}_{01})}{n_1} + \frac{\hat{P}_{01}(1 - \hat{P}_{01})}{n_2}$$

$$= \hat{V}ar_{H_0}\hat{P}_{11} + \hat{V}ar_{H_0}\hat{P}_{21} . \qquad (1-20)$$

The result in (1-20) is a special case of

$$Var(\hat{P}_{11} - \hat{P}_{21}) = Var(\hat{P}_{11}) + Var(\hat{P}_{21}) - 2\ Cov(\hat{P}_{11}, \hat{P}_{21}) \qquad (1-21)$$

when $Cov(\hat{P}_{11}, \hat{P}_{21}) = 0$. Thus the normal curve statistic (1-18) and the chi-square statistic (1-14) treat $\hat{P}_{11}$ and $\hat{P}_{21}$ as un-correlated random variables since (1-20) implies $\hat{Cov}_{H_0}(\hat{P}_{11}, \hat{P}_{21})$ is taken to be 0. For the general case of d samples measured on a variable of r categories the result in (1-11) is assumed to hold. Thus valid use of the chi-square test of homogeneity requires between sample independence because data from dependent samples are correlated. Use of C^2 for dependent samples assumes incorrectly that condition (1-11) is satisfied and results in an invalid test.

The mixed categorical data model of order-d involves a situation in which n randomly chosen subjects or blocks from some homogeneous population are measured on an r-valued categorical dependent variable under d different conditions. The hypothesis

of general interest is whether the distribution of responses is
the same under the d different conditions. Koch and Reinfurt
(1971) stated that such a hypothesis best examines the relative
effects of the d conditions on the dependent measure of interest.

Because of the nature of the design, the chi-square test
of homogeneity is most likely inappropriate to analyze the data from
such a model. It is doubtful that the observations or responses
made by the same subject are independent of one another. Since the
chi-square test of homogeneity does not make any allowances for
correlated responses, what is needed is a test of homogeneity for
correlated samples that builds the correlation into the procedure.

The use of the mixed categorical data model of order-d and
a subsequent test for the homogeneity of the resulting d correlated
distributions are required in the following types of situations.

    i.   Panel studies in which d different attitudinal
policy questions are asked of the subjects with
each question having the same r categories of
response. A test for homogeneity of the d distribu-
tions of response would determine if the group held
the same attitude for each of the policy questions.

    ii.   Longitudinal studies in which each of the subjects
in a group is measured at d different time intervals
on the same r-valued categorical variable. A test
for homogeneity of the d distributions of response
would determine if the group as a whole changed its
mode of classification over time.

iii. Studies in which a single group of subjects is classified into r categories by d different raters. A test for homogeneity of the d distributions would determine if the d raters viewed the group as a unit in the same manner.

iv. Studies in which the same group of subjects is asked the same question which has r categories of response under d different treatment conditions. A test for homogeneity of the d distributions would determine the relative effects of the d treatments.

v. Studies in which the subjects in each of d matched samples are each classified into one of r categories. A test for homogeneity of the d distributions would determine the degree to which the d groups are the same in terms of the variable of classification.

Design situation (1) can be illustrated by part of a survey reported by Marascuilo and McSweeney (1969). The survey concerned the attitude of Berkeley residents toward the integration of the city's schools. A sample of the city's adult population was made and the subjects were asked to respond to the following questions.

(1) For some grade schools, [it has been] suggested that lines be changed so that the percentage of nonwhite and white children in these schools would be more like the percentage for the entire school system

____ I agree ____ I disagree ____ I am not sure

(2) For [two of the three] Junior High Schools, [it has been] suggested that school lines be changed so that the percentage of nonwhite and white children in these schools would be more like the percentage for the entire school system

_____ I agree _____ I disagree _____ I am not sure

(3) If more day-care centers and nursery schools are set up, [it has been] suggested that they be set up to permit a greater integration of the races than is found at present

_____ I agree _____ I disagree _____ I am not sure

A test for the homogeneity of the distributions of the three questions would provide a test of whether the adult population of Berkeley viewed each of the issues in the same manner. Such an analysis would be an improvement upon the pairwise analysis used by Marascuilo and McSweeney (1969) in the absence of a simultaneous test for three correlated distributions on a polychotomous variable of three categories.

From the same study reported by Marascuilo and McSweeney (1969) the following question was asked of the group.

[It has been] suggested that more day-care centers and nursery schools be set up to let more children attend.

_____ I agree _____ I disagree _____ I am not sure

If such a question were asked over d increments of time a test for the homogeneity of the d distributions would examine the attitude of the Berkeley adult community to this issue as a function of time. This is an example of situation (ii).

In the Miller et al. (1970) study a sample of 84 infants were tested on a series of tasks involving object concept development in the sensory-motor period. Two observers categorized what they believed to be each of the children's overall performance on a series of Piagetian-tasks. The categories of classification were (0) failure, (1) partial pass, and (2) pass. If the researchers were interested in determining whether the two observers applied equally stringent standards to the group, a test for homogeneity of the distributions of ratings for each of the observers would answer such a question. This illustration is an example of the design situation (iii).

Yoshinaga (1974) investigated the constancy of teachers' choices of a disciplinary strategy -- positive reinforcement, social modeling, and punishment -- when increasing increments of information were given about the child involved in the disciplinary incident. The teachers were first given a written description of the disciplinary incident and asked to choose a strategy to deal with the incident. The teachers were then given some biographical information about the child and again asked to choose a disciplinary strategy. Finally, the teachers were presented with additional biographical information about the child and again asked to choose a strategy. The researcher was interested in knowing if teachers change their strategy to handle disciplinary problems as more is known about the child. A test for homogeneity of the resulting three distributions answers such a research question. This is an illustration of design situation (iv).

In Laumann's (1973) study of social interaction and social
mobility, a husband and the fathers of both the husband and wife were
each classified into one of a number of social classes. This design
can be considered as consisting of three matched samples and is an
example of design situation (v). By testing for homogeneity of the
three correlated distributions, the researcher can make inferences
concerning the social interaction of various social classes as well
as make inferences about generational changes in social class. The
classification of a husband, his father, and his father-in-law are
most probably not made independently. The matched triples of
husband, father, and father-in-law can be considered as blocks in a
randomized block design.

The previous examples show the different ways in which the
mixed categorical data model can be used in behavioral science re-
search. The studies of Marascuilo and McSweeney (1969), Miller
et al. (1970), and Yoshinaga (1974) illustrate data analytic situa-
tions which demanded a statistic to test for homogeneity of d
correlated distributions. Such a statistic can be applied in a variety
of settings which require categorical data analysis of either matched
samples or repeated measures data.

This dissertation focuses on testing for homogeneity of dis-
tributions on a polychotomous variable for non independent samples.
The utility of such a technique in educational and behavioral science
research has been pointed out in this introductory chapter. The
dissertation has three major sections. The first section provides a
review and synthesis of some of the more significant literature re-
lating to the problem. Several methods for dealing with the problem

are presented and compared in this review. The second major section
provides a detailed development of a large sample statistic which
can be used to test for homogeneity of correlated distributions.
The statistic has the appeal of simplicity and can easily be related
to techniques which are familiar to behavioral researchers who have
had some exposure to statistical methods. In addition a development
of post-hoc techniques to be used in conjunction with the test
statistic is provided. The third major portion of the dissertation
reports a simulation study of the small sample behavior of the
statistic developed in the second section. The distributional pro-
perties of the statistic both under the null and alternative hypothesis
are considered for a number of small sample cases. The purposes of
this third section are to assess the degree to which the asymptotic
results hold in settings which are frequently encountered in data
analytic situations and to establish a series of guidelines which
can help the potential user of the technique in determining its
appropriateness. A final chapter is included which ties the body of
the dissertation together in the form of a summary and conclusions
reached as a result of the work.

CHAPTER II

TESTING FOR HOMOGENEITY

MULTIVARIATE VERSUS UNIVARIATE APPROACH

The mixed categorical data model of order-d is characterized by a situation in which n randomly chosen subjects or blocks from some homogeneous population are measured on an r-valued categorical dependent variable under d different conditions. The hypothesis of interest is whether the distribution of responses is the same under the d different conditions. Because it is doubtful that responses made by the same subject or by matched subjects are independent of one another, a multivariate model is assumed to fit the data.

The data for this model are represented in an $r \times r \times ... \times r$ contingency table of d dimensions. The set of d responses made by each subject is counted as a single observation in the contingency table. The set of responses made by a subject can be characterized by a vector $(j_1, j_2, ..., j_d)$ where $j_g = 1, 2, ..., r$ for $g = 1, 2, ..., d$. In all there are r^d such possible vectors or modes of response which characterize the r^d celled contingency table. Let $P_{j_1, j_2, ..., j_d}$ represent the probability of a response profile $(j_1, j_2, ..., j_d)$ or equivalently that of an observation falling into cell $(j_1, j_2, ..., j_d)$ of the contingency table. A multinomial distribution with parameters $\{P_{j_1, j_2, ..., j_d}\}$ and n is assumed

15

to fit the cell frequencies $\{n_{j_1,j_2,\ldots,j_d}\}$ with

$$\sum_{(j_1,j_2,\ldots,j_d)} P_{j_1,j_2,\ldots,j_d} = 1, \quad P_{j_1,j_2,\ldots,j_d} > 0, \quad \text{and}$$

$$\sum_{(j_1,j_2,\ldots,j_d)} n_{j_1,j_2,\ldots,j_d} = n \ .$$

(2-1)

The distinction between a univariate and multivariate approach is a consequence of the manner in which an observation is defined. A univariate approach would count the d responses made by each of the subjects as d separate observations. In all there would be nd observations which could be represented by a d × r contingency table. Such a model is appropriate for a design in which the d responses made by a subject are assumed to be mutually independent. As was indicated in Chapter I the chi-square test of homogeneity for this design is the appropriate procedure to test whether the distribution of responses is the same under the d different conditions. This is contrasted with the multivariate approach which views the d responses made by each of the subjects as a single response profile which is characterized by a response vector of d components. It is the response vectors which serve as the conceptual units of observation. Under the multivariate approach there are thus n and not nd observations. The n observations can be represented in an r × r ×...× r contingency table of d dimensions. Such a model is appropriate for a design in which the d responses made by a subject are assumed to be correlated as is often the case with the repeated measures or matched-sample data.

In Chapter I it was pointed out that when observations over the repeated measures are correlated, estimates of the proportions which characterize the d distributions of response are also correlated. The multivariate approach examines the joint distribution of responses over the repeated measures. A test of whether the distribution of responses is the same under the d different conditions reduces to testing whether the marginal distributions in the r × r ×...× r contingency table are homogeneous. Because the multivariate approach makes use of the joint distribution the nature of the correlation between the estimates of the proportions which characterize the d distributions of response can be determined. Such information could clearly not be determined using the d × r contingency table which characterizes the univariate approach. The multivariate approach, which is employed throughout the dissertation, takes account of the correlated nature of the responses by focusing on the joint distribution.

Let M_{qb} represent the probability that a subject is classified by category b under the qth condition. M_{qb} is a marginal probability which can be written as

$$M_{qb} = \sum_{(j_1,j_2,\ldots,j_d) \text{ with } j_q=b} \cdots \sum P_{j_1,j_2,\ldots,j_d} \quad \text{for} \quad \begin{array}{l} q = 1,2,\ldots,d \\ b = 1,2,\ldots,r \end{array} \quad (2\text{-}2)$$

The formula given in (2-2) for the marginal probability indicates that all joint probabilities which correspond to response vectors having response category b under the qth condition are summed.

The hypothesis of interest is that the distribution of responses is the same under the d different conditions. This

hypothesis of marginal homogeneity can be written as

$$H_0: \quad M_{1b} = M_{2b} = \ldots = M_{db} \quad \text{for} \quad b = 1,2,\ldots,r. \qquad (2\text{-}3)$$

A number of different techniques have appeared in the literature for testing H_0. Each of these techniques assumes the multivariate model as set forth in this chapter. The remainder of the chapter is divided into two main sections. The first section presents a number of different techniques to test H_0 while the second section provides a comparison among the techniques.

<div align="center">TEST PROCEDURES</div>

Notation

In presenting the different approaches to testing H_0 a common notation is used. The multivariate categorical model as described in the previous section is assumed for each of the techniques described together with the notation introduced in that section. Let $\hat{P}_{j_1,j_2,\ldots,j_d}$ and $\hat{M}_{qb}$ denote the unrestricted estimates of the cell and marginal probabilities respectively with

$$\hat{P}_{j_1,j_2,\ldots,j_d} = \frac{n_{j_1,j_2,\ldots,j_d}}{n}$$

$$\hat{M}_{qb} = \frac{\sum \ldots \sum_{(j_1,j_2,\ldots,j_d) \text{ with } j_q = b} n_{j_1,j_2,\ldots,j_d}}{n} . \qquad (2\text{-}4)$$

For certain of the techniques it is necessary to estimate cell probabilities under the constraints of the null hypothesis. Such estimates will be denoted as $\tilde{P}_{j_1,j_2,\ldots,j_d}$ with the method of estimation defined within the content of the technique presented.

Stuart's Statistic

The technique developed by Stuart (1955) is a large sample statistic which tests for homogeneity of two correlated distributions. The statistic is confined to the mixed categorical data model of order-2 whose data can be represented by an $r \times r$ contingency table. The null hypothesis can be written as

$$H_0: \quad M_{1b} = M_{2b} \quad \text{for} \quad b = 1,2,\ldots,r. \tag{2-5}$$

Stuart argued that the cell frequencies n_{j_1,j_2} have a limiting multivariate normal distribution. Stuart then defined the variate $d_i = n(\hat{M}_{2i} - \hat{M}_{1i})$ for $i = 1,2,\ldots,r$ and argued that since the d_i are linear functions of the cell frequencies, the d_i will also have a limiting multivariate distribution but of rank $(r-1)$ because of the constraint $\sum_{i=1}^{r} d_i = 0$.

Stuart then suggested a statistic which could be used to test H_0. The statistic has the form

$$X_S^2 = \underline{d}' \hat{\Sigma}_{\underline{d}}^{-1} \underline{d}$$

$$\tag{2-6}$$

with $\underset{1\times(r-1)}{\underline{d}'} = [d_1,d_2,\ldots,d_{r-1}], \qquad \underset{(r-1)\times(r-1)}{\hat{\Sigma}_{\underline{d}}} = ((\hat{\sigma}_{ij})).$

The matrix $\hat{\Sigma}_{\underline{d}}$ is a consistent estimator of the variance-covariance matrix for the random vector $\underline{d}$ under the assumption that H_0 is true with the $\hat{\sigma}_{ij}$ being defined as

$$\hat{\sigma}_{ij} = \widehat{\text{Cov}}(d_i,d_j|H_0) = -n(\hat{P}_{i,j} + \hat{P}_{j,i}) \qquad (i \neq j)$$

$$\tag{2-7}$$

$$\hat{\sigma}_{ii} = \widehat{\text{Var}}(d_i|H_0) = n(\hat{M}_{2i} + \hat{M}_{1i} - 2\hat{P}_{i,i}) \quad .$$

Because the vector $\underline{d}$ is approximately multivariate normal

of full rank $(r-1)$ and $\hat{\Sigma}_{\underline{d}}$ is a consistent estimator of the

variance-covariance matrix of $\underline{d}$ under H_0, Stuart claimed that

for large n, the statistic X_S^2 has a chi-square distribution with

$(r-1)$ degrees of freedom under the null hypothesis of marginal

homogeneity. Stuart then defined the rejection region for the large-

sample test as the upper tail of the chi-square distribution with

$(r-1)$ degrees of freedom. Because any $(r-1)$ of the d_i uniquely

determine the remaining one, Stuart argued that the value of the

X_S^2 statistic remains invariant under the choice of which d_i to

eliminate in the formation of the $\underline{d}$ vector.

Likelihood Ratio Statistic

Madansky (1963) developed a large sample statistic, based

upon the likelihood ratio criterion of Neyman and Pearson (1928),

which tests the null hypothesis of marginal homogeneity in the

general $r \times r \times ... \times r$ contingency table of d dimensions. Using

Madansky's technique the null hypothesis is stated as

$$H_0: \quad M_{qb} = M_{1b} \quad \text{for} \quad \begin{array}{l} q = 2,...,d \\ b = 1,2,...,r \end{array} . \tag{2-8}$$

Let $\underline{P}$ represent the vector of cell probabilities with

$$\underset{1 \times r^d}{\underline{P}'} = [P_{j_1,j_2,...,j_d}] . \tag{2-9}$$

The likelihood function for the multinomial distribution which char-

acterizes the cell frequencies is

$$L(\underline{P}) = \frac{n!}{\prod\limits_{j_1,j_2,\ldots,j_d} n_{j_1,j_2,\ldots,j_d}!} \prod\limits_{j_1,j_2,\ldots,j_d} (P_{j_1,j_2,\ldots,j_d})^{n_{j_1,j_2,\ldots,j_d}} . \tag{2-10}$$

The likelihood ratio statistic is given by

$$\lambda = \frac{L(\tilde{\underline{P}})}{L(\hat{\underline{P}})} . \tag{2-11}$$

The $\tilde{\underline{P}}$ vector of probabilities is that estimate of $\underline{P}$ which maximizes the quantity in (2-10) subject to the restrictions specified by the null hypothesis and the additional constraint that the sum of the components in $\underline{P}$ is 1. The $\hat{\underline{P}}$ vector of probabilities is that estimate of $\underline{P}$ which maximizes the likelihood function in (2-10) subject only to the constraint that the sum of the cell probabilities is 1.

Madansky specified the constraints which define the null hypothesis. In all there are $(d-1)(r-1)$ linearly independent constraints which do not depend on the added constraint that

$$\sum_{j_1,j_2,\ldots,j_d} P_{j_1,j_2,\ldots,j_d} = 1 . \tag{2-12}$$

The constraints which are implied by the null hypothesis can be expressed as

$$\sum_{j_1,\ldots,j_{q-1},j_{q+1},\ldots,j_d} [P_{j_1,j_2,\ldots,j_{q-1},b,j_{q+1},\ldots,j_d} \tag{2-13}$$

$$- P_{b,j_2,\ldots,j_{q-1},j_1,j_{q+1},\ldots,j_d}] = 0; \quad b = 1,2,\ldots,r-1$$
$$q = 2,\ldots,d .$$

The $(d-1)(r-1)$ constraints given in (2-13) are equivalent to the statement that $M_{qb} = M_{1b}$ for $q = 2,\ldots,d$ and $b = 1,2,\ldots,r-1$.

But this, together with the constraint given in (2-12), implies

the null hypothesis of homogeneity as stated in (2-8).

Maximizing the function $L(\underline{P})$ in (2-10) under the hypo-

thesis of homogeneity is equivalent to maximizing the function

$$H(\underline{P}) = \sum_{(j_1, j_2, \ldots, j_d)} n_{j_1, j_2, \ldots, j_d} \log P_{j_1, j_2, \ldots, j_d} \qquad (2\text{-}14)$$

subject to the $(d-1)(r-1) + 1$ constraints given in (2-12) and

(2-13). If the Lagrangian multipliers λ_0 and μ_{qb} (with

$q = 2, \ldots, d$ and $b = 1, \ldots, r-1$) are introduced, the problem re-

duces to finding that value of $\underline{P}$ which maximizes the function

$$L^*(\underline{P}) = H(\underline{P}) - \lambda_0 \left[\sum_{(j_1, \ldots, j_d)} P_{j_1, \ldots, j_d} - 1 \right] - \sum_{q=2}^{d} \sum_{b=1}^{r-1} \mu_{qb} \phi_{qb} \qquad (2\text{-}15)$$

where

$$\phi_{qb} = \sum_{j_1, \ldots, j_{q-1}, j_{q+1}, \ldots, j_d} [P_{j_1, j_2, \ldots, j_{q-1}, b, j_{q+1}, \ldots, j_d}$$

$$- P_{b, j_2, \ldots, j_{q-1}, j_1, j_{q+1}, \ldots, j_d}] \ .$$

Differentiating $L^*(\underline{P})$ with respect to each of the $P_{j_1, j_2, \ldots, j_d}$,

and setting the derivatives equal to zero, Madansky obtained r^d

equations of the form

$$\frac{n_{j_1, \ldots, j_d}}{P_{j_1, \ldots, j_d}} - \lambda_0 - \sum_{q=2}^{d} (\mu_{qj_q} - \mu_{qj_1}) = 0 \ . \qquad (2\text{-}16)$$

The estimates of the $\{P_{j_1, \ldots, j_d}\}$ can then be expressed in the

form

$$\tilde{P}_{j_1,\ldots,j_d} = \frac{n_{j_1,\ldots,j_d}}{\lambda_0 + \sum_{q=2}^{d} (\mu_{qj_q} - \mu_{qj_1})} \, . \qquad (2\text{-}17)$$

In order to find the actual values of the $\{\tilde{P}_{j_1,\ldots,j_d}\}$ the corresponding expressions given in (2-17) are substituted into the constraint equations (2-12) and (2-13). Since the resulting equations are nonlinear in the unknown Lagrangian multipliers an iterative method is needed to solve these equations. Madansky described a linear approximation method for handling the problem. A detailed algorithm is also given by the author. The resulting $\tilde{P}_{j_1,\ldots,j_d}$ which are obtained using this procedure are the maximum likelihood estimates of the $P_{j_1,\ldots,j_d}$ under the condition specified by the null hypothesis of marginal homogeneity.

The unrestricted maximum likelihood estimates of the $P_{j_1,\ldots,j_d}$ subject only to the constraint given in (2-12) are the observed relative frequencies given by

$$\hat{P}_{j_1,\ldots,j_d} = \frac{n_{j_1,\ldots,j_d}}{n} \, . \qquad (2\text{-}18)$$

The likelihood ratio statistic is then given by

$$\lambda = \prod_{j_1,\ldots,j_d} \left[\frac{\tilde{P}_{j_1,\ldots,j_d}}{\hat{P}_{j_1,\ldots,j_d}} \right]^{n_{j_1,\ldots,j_d}} \, . \qquad (2\text{-}19)$$

The value of λ always falls between 0 and 1. A small value of λ indicates that the data provide evidence that the null hypothesis is false. A test statistic of the form

$$\chi^2_M = -2 \log \lambda \qquad (2\text{-}20)$$

was given by Madansky. Large values of X_M^2 lead to a rejection of the null hypothesis. For large sample sizes the X_M^2 statistic is approximately distributed as a chi-square random variable with $(d-1)(r-1)$ degrees of freedom.

Minimum Discrimination Information Statistic

Ireland, Ku, and Kullback (1969) considered the problem of testing for marginal homogeneity within the context of the mixed categorical data model of order-2. The authors proposed a statistic based upon minimum discrimination information estimation (MDIE) of the cell probabilities of an observed $r \times r$ contingency table under the null hypothesis of marginal homogenity.

Before discussing the actual procedure used by the authors for the particular problem under consideration, a more general discussion of the MDIE technique and how it can be used to produce estimates of cell probabilities under a general null hypothesis is given. An associated hypothesis testing procedure is also discussed.

Let $\{P_{j_1,j_2}\}$ characterize the cell probabilities in an $r \times r$ contingency table such that

$$\sum_{j_1,j_2} P_{j_1,j_2} = 1, \quad P_{j_1,j_2} > 0 \quad j_1 = 1,\ldots,r; \; j_2 = 1,\ldots,r . \quad (2-21)$$

Such a table will be denoted as the $\underline{P}$-table. Let $\{\hat{P}_{j_1,j_2}\}$ be the observed set of proportions based upon n observations from an $r \times r$ contingency table whose classifications are common to the $\underline{P}$-table and such that

$$\sum_{j_1,j_2} \hat{P}_{j_1,j_2} = 1, \quad \hat{P}_{j_1,j_2} > 0 \quad j_1 = 1,\ldots,r; \; j_2 = 1,\ldots,r. \quad (2-22)$$

Such a table will be referred to as the $\hat{\underline{P}}$-table.

A distance-like measure from the $\underline{P}$-table to the $\hat{\underline{P}}$-table known as the discrimination information denoted by $I(\underline{P}:\hat{\underline{P}})$ can be defined as

$$I(\underline{P}:\hat{\underline{P}}) = \sum_{j_1,j_2} P_{j_1,j_2} \log \frac{P_{j_1,j_2}}{\hat{P}_{j_1,j_2}} \; . \tag{2-23}$$

The discrimination information is a distance-like function in the sense that

$$\begin{array}{ll} \text{(i)} & I(\underline{P}:\underline{P}) = 0 \\ & \tag{2-24} \\ \text{(ii)} & I(\underline{P}:\hat{\underline{P}}) > 0 \quad \text{for} \quad \underline{P} \neq \hat{\underline{P}} \; . \end{array}$$

It should be pointed out however that in general the values $I(\underline{P}:\hat{\underline{P}})$ and $I(\hat{\underline{P}}:\underline{P})$ are not the same.

Let W represent the family of all $\underline{P}$-tables which satisfy the constraints of some null hypothesis H_0. The method of MDIE then consists of finding that $\underline{P} \in W$ which minimizes the discrimination information in (2-23). In a sense the technique consists of finding that $\underline{P}$-table which satisfies the constraints of the null hypothesis and at the same time most closely resembles the observed table $\hat{\underline{P}}$. The $\underline{P} \in W$ which is closest in distance to the observed table will be denoted as $\tilde{\underline{P}}$. The $\{\tilde{P}_{j_1,j_2}\}$ are then the MDIE of the cell probabilities under the constraints of the null hypothesis H_0. Taylor (1953) has shown that estimates obtained in this manner are BAN in the sense of Neyman (1949). A discussion of BAN estimators is given in more detail later in the chapter.

In order to test a null hypothesis a statistic of the form

$$X_I^2 = 2n \sum_{j_1,j_2} \tilde{P}_{j_1,j_2} \log \frac{\tilde{P}_{j_1,j_2}}{\hat{P}_{j_1,j_2}} \tag{2-25}$$

can be used. Using the results of Kullback (1968) and Neyman (1949), it can be shown that the statistic X_I^2 has a limiting chi-square distribution under the null hypothesis. The degrees of freedom are given by the number of linearly independent constraints put on the components of $\underline{P}$ which are specified by the null hypothesis H_0 and which are independent of the constraint that the sum of the components in $\underline{P}$ is 1. The X_I^2 statistic was called the minimum discrimination information statistic (MDIS) by Ireland, Ku, and Kullback (1969). Large values of the statistic indicate that the observed probabilities differ considerably from those cell probabilities which would be expected under the null hypothesis. Hence large values of X_I^2 would reject the null hypothesis.

Now consider the specific problem investigated by Ireland et al. (1969) who were interested in testing for the marginal homogeneity of two distributions in an $r \times r$ table. The procedure which these authors used follows the general methodology which has just been described.

The specific null hypothesis under consideration is

$$H_0: \quad M_{1b} = M_{2b} \quad \text{for} \quad b = 1,2,\ldots,r . \tag{2-26}$$

The linearly independent constraints specified by H_0 which are independent of $\sum_{i,j} P_{i,j} = 1$ are then expressed as

$$\sum_{j=1}^{r} P_{i,j} = \sum_{k=1}^{r} P_{k,i} \quad \text{for} \quad i = 1,2,\ldots,r-1 . \tag{2-27}$$

The first part of the procedure consists of finding that $\underline{P}$ vector whose components satisfy the constraints given in (2-27) and which minimizes the discrimination information

$$I(\underline{P}:\hat{\underline{P}}) = \sum_{i,j} P_{i,j} \log \frac{P_{i,j}}{\hat{P}_{i,j}} . \tag{2-28}$$

If the Lagrangian multipliers γ_0 and α_i, $i = 1,\ldots,r-1$ are introduced, the problem reduces to minimizing

$$\sum_{i=1}^{r} \sum_{j=1}^{r} P_{i,j} \log \frac{P_{i,j}}{\hat{P}_{i,j}} + \sum_{i=1}^{r-1} \alpha_i \left[\sum_{j=1}^{r} P_{i,j} - \sum_{k=1}^{r} P_{k,i} \right]$$

$$+ \gamma_0 \left[\sum_{i=1}^{r} \sum_{j=1}^{r} P_{ij} - 1 \right] \tag{2-29}$$

with respect to the $P_{i,j}$. Expressions for the $\{\tilde{P}_{i,j}\}$ can be obtained in terms of the unknown Lagrangian multipliers γ_0 and α_i, $i = 1,\ldots,r-1$ by differentiating the expressions in (2-29) with respect to each of the $P_{i,j}$ and solving the resulting r^2 equations together with the constraint equations given in (2-27) and by $\sum_{i,j} P_{i,j} = 1.$

Rather than solving for the unknown Lagrangians the authors developed a convergent iterative procedure which leads directly to the estimates $\{\tilde{P}_{i,j}\}$. Approximations to the $\{\tilde{P}_{i,j}\}$ can be achieved to any desired level of accuracy. A proof that the approximations do in fact converge to the desired $\{\tilde{P}_{i,j}\}$ was also given by the authors. Once the $\{\tilde{P}_{i,j}\}$ are found, or at least approximated, the MDIS to test the null hypothesis (2-26)

is given by

$$X_I^2 = 2n \sum_{i,j} \tilde{P}_{i,j} \log \frac{\tilde{P}_{i,j}}{\hat{P}_{i,j}} \quad . \tag{2-30}$$

Values of X_I^2 larger than the $(1-\alpha)$th quantile of the tabled chi-square distribution with $(r-1)$ degrees of freedom will reject the null hypothesis of marginal homogeneity for tests performed at the α level of significance.

Weighted Least Squares Approach

Koch and Reinfurt (1971) derived a statistic which tests for homogeneity of the marginal distributions within the context of the general case of the mixed categorical data model of order-d. The development of the test statistic makes use of a general methodology given in Grizzle, Starmer, and Koch (1969). This general methodology involves the derivation of test statistics in terms of weighted least squares analysis of certain linear models. A brief outline of the general methodology given by Grizzle et al. (1969) is discussed, followed by the Koch and Reinfurt (1971) adaptation of the technique to the problem of testing for homogeneity of the d marginal distributions in the $r \times r \times ... \times r$ contingency table.

Let $\underline{P}$ represent the vector of cell probabilities which characterizes the $r \times r \times ... \times r$ contingency table of d dimensions and let $\underline{\hat{P}}$ represent the corresponding vector of observed proportions. Let $F_m(\underline{P})$ be any function of the elements of $\underline{P}$ that has partial derivatives up to the second order with respect to the $\{P_{j_1, ..., j_d}\}$, $m = 1, 2, ..., u$. Also let

$$F_m(\hat{\underline{P}}) = F_m(\underline{P}) \quad \text{evaluated at} \quad \underline{P} = \hat{\underline{P}}$$

$$\left[\underset{u \times 1}{\underline{F}}(\underline{P})\right]' = [F_1(\underline{P}), F_2(\underline{P}), \ldots, F_u(\underline{P})] \tag{2-31}$$

$$\left[\underset{u \times 1}{\underline{F}}(\hat{\underline{P}})\right]' = [F_1(\hat{\underline{P}}), F_2(\hat{\underline{P}}), \ldots, F_u(\hat{\underline{P}})] \ .$$

It is assumed that the functions $F_m(\underline{P})$, m = 1,2,...,u are linearly independent of one another and of the constraint that the sum of the components of the $\underline{P}$ vector is 1.

Let

$$\underset{u \times 1}{\underline{F}}(\underline{P}) = \underset{u \times v}{X} \underset{v \times 1}{\underline{B}} \ , \tag{2-32}$$

where X is a known matrix of coefficients and $\underline{B}$ is a vector of unknown parameters. A general equation of the form given in (2-32) can be used to specify a linear model which is hypothesized to characterize the data. Grizzle et al. (1969) proposed a statistic to test the fit of the data to the linear model proposed in (2-32). The test statistic is given by

$$X_{GSK}^2 = (\underline{F}(\hat{\underline{P}}) - X \tilde{\underline{B}})'S^{-1}(\underline{F}(\hat{\underline{P}}) - X \tilde{\underline{B}}) \tag{2-33}$$

where S is the sample estimate of the variance-covariance matrix of $F(\hat{\underline{P}})$ and $\tilde{\underline{B}}$ is that value of the $\underline{B}$ vector which minimizes the quantity

$$(\underline{F}(\hat{\underline{P}}) - X \underline{B})'S^{-1}(\underline{F}(\hat{\underline{P}}) - X \underline{B}) \ . \tag{2-34}$$

The estimate of the vector of unknown parameters which is given by $\tilde{\underline{B}}$ is the weighted least squares estimate and the test statistic is simply the residual sum of squares or sum of squares error due

to lack of fit. If X is of rank v, the test statistic X^2_{GSK} has approximately a chi-square distribution with u - v degrees of freedom for large sample sizes under the assumption that the model in (2-32) fits. Large values of the statistic provide evidence that the model does not fit.

Consider the Koch and Reinfurt (1971) adaptation of the general weighted least squares methodology to test for homogeneity of d marginal distributions in a mixed categorical data model of order-d. The null hypothesis of marginal homogeneity can be stated as

$$H_0: \quad M_{1b} = M_{2b} = \ldots = M_{db} \quad \text{for} \quad b = 1,2,\ldots,r \quad . \quad (2-35)$$

In terms of a linear model approach the hypothesis in (2-35) could be stated in the form

$$H_0: \quad M_{qb} = B_b, \quad q = 1,2,\ldots,d; \quad b = 1,2,\ldots,r-1 \quad (2-36)$$

where the B_b are unknown parameters which are estimated from the data. The linear model can be written in matrix terms as

$$
\begin{bmatrix}
M_{11} \\
M_{21} \\
\cdots \\
M_{d1} \\
M_{12} \\
M_{22} \\
\cdots \\
M_{d2} \\
\cdot \\
\cdot \\
\cdot \\
M_{1r-1} \\
M_{2r-1} \\
\cdots \\
M_{dr-1}
\end{bmatrix}
=
\begin{bmatrix}
1 & 0 & \cdots & 0 \\
1 & 0 & \cdots & 0 \\
 & & \cdots & \\
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
 & & \cdots & \\
0 & 1 & \cdots & 0 \\
 & & \cdot & \\
 & & \cdot & \\
 & & \cdot & \\
0 & 0 & \cdots & 1 \\
0 & 0 & \cdots & 1 \\
 & & \cdots & \\
0 & 0 & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
B_1 \\
B_2 \\
\cdot \\
\cdot \\
\cdot \\
B_{r-1}
\end{bmatrix}
\qquad (2\text{-}37)
$$

$$
\begin{array}{ccc}
\underline{F} & (\underline{P}) = & X \quad \underline{B} \quad \cdot \\
d(r\text{-}1)\times 1 & & d(r\text{-}1)\times(r\text{-}1) \quad (r\text{-}1)\times 1
\end{array}
$$

Testing the null hypothesis of marginal homogeneity as given

in (2-35) can be accomplished by testing the fit of the model given

in (2-36) to the observed data. A significant test of fit statistic

(2-33) indicates that the hypothesized linear model in (2-36) does

not fit the data. This is equivalent to rejecting the null hypo-

thesis of marginal homogeneity as is stated in (2-35). The test

statistic X^2_{GSK} has an asymptotic chi-square null distribution

with $d(r\text{-}1) - (r\text{-}1) = (d\text{-}1)(r\text{-}1)$ degrees of freedom.

COMPARING TEST PROCEDURES

This section of the chapter is devoted to a discussion of the relationships among the four procedures presented in the previous section. The relationships developed are, in some cases, based upon modifications and elaborations of previous research. Before any direct comparisons are made among the four procedures under consideration, a series of theoretical results are presented which serve as the basis for the subsequent comparisons.

Throughout this section the symbol P_i will be used in place of $P_{j_1,\ldots,j_d}$ for the sake of notational simplicity and the multinomial probability model will be assumed throughout.

Neyman and BAN Estimators

The mixed categorical data model of order-d, as considered in this dissertation, assumes an $r \times r \times \ldots \times r$ contingency table of d dimensions whose cell frequencies are characterized by the probability distribution

$$\phi = \frac{n!}{\prod\limits_i n_i!} \prod\limits_i P_i^{n_i} \tag{2-38}$$

such that $\sum\limits_i P_i = 1$ and $\sum\limits_i n_i = n$.

Let

$$f_m(\underline{P}) = 0 \ , \quad m = 1,2,\ldots,t \tag{2-39}$$

define t linearly independent constraints on the components of $\underline{P}$ which are independent of the constraint $\sum\limits_i P_i = 1$. It is assumed that $f_m(\underline{P})$, $m = 1,2,\ldots,t$, possess continuous partial derivatives up to the second order with respect to the P_i and

that there is at least one solution such that $P_i > 0$ for all i.

The problem of estimating the components of the $\underline{P}$ vector given the model in (2-38) and the constraints (2-39) was considered by Neyman (1949). Neyman defined a class of estimators known as best asymptotically normal (BAN) estimators which possess the following properties:

If $\tilde{P}_i$ is a BAN estimator of P_i then

(i) $\tilde{P}_i$ is a consistent estimator of the parameter P_i. This means that as n, the sample size tends to infinity, the estimator $\tilde{P}_i$ approaches the parameter P_i.

(ii) $\sqrt{n}(\tilde{P}_i - P_i)$ is asymptotically normal with zero mean and asymptotic variance σ^2, where σ^2 is independent of n. This means that as n tends to infinity, for any real number t,

$$P\left\{ \frac{\sqrt{n}(\tilde{P}_i - P_i)}{\sigma} < t \right\} \rightarrow P\{Z < t\}, \text{ where } Z \sim N(0,1) . \quad (2\text{-}40)$$

(iii) If $\tilde{P}_i^*$ is any estimator of P_i satisfying (i) and (ii) but with σ^* taking the place of σ in (ii) then

$$\sigma^* \geq \sigma . \quad (2\text{-}41)$$

(iv) $\tilde{P}_i$ has continuous partial derivatives with respect to the observed proportions.

Neyman (1949) showed that BAN estimators for the components of $\underline{P}$ under the required restrictions can be obtained by either minimizing the quantities

(a) $\displaystyle\sum_i \frac{(n_i - nP_i)^2}{nP_i}$ or (b) $\displaystyle\sum_i \frac{(n_i - nP_i)^2}{n_i}$ (2-42)

or by maximizing the quantity

(c) $\displaystyle\frac{n!}{\prod_i n_i!} \prod_i P_i^{n_i}$ (2-43)

with respect to the P_i under the given constraints. Estimates

obtained using criterion (a) are called minimum chi-square estimates.

Estimates obtained using criterion (b) are called modified minimum

chi-square estimates while those estimates computed using criterion

(c) are called maximum likelihood estimates. Maximum likelihood

estimators constitute a subclass of BAN estimators. The motiva-

tion for Neyman's (1949) paper was to find a class of estimators

more general than that of maximum likelihood but possessing the

same desirable asymptotic properties as maximum likelihood estimators.

It was hoped that in certain cases some of the BAN estimators would

be easier to compute than maximum likelihood estimators. Neyman

showed that if the $f_m(\underline{P})$, $m = 1,2,\ldots,t$ given in (2-39) are

linear in the P_i, then a set of modified minimum chi-square

estimates could be obtained by solving only a system of linear equa-

tions.

Neyman demonstrated that a null hypothesis defined by the

t linearly independent constraints given in (2-39) can be tested

by using either the chi-square statistic, modified chi-square

statistic, or likelihood ratio test statistic defined respectively as

$$\chi^2 = \sum_i \frac{(n_i - n\tilde{P}_i)^2}{n\tilde{P}_i}$$

and

$$\chi_1^2 = \sum_i \frac{(n_i - n\tilde{P}_i)^2}{n_i} \tag{2-44}$$

and

$$-2 \log \lambda = 2 \sum_i n_i (\log n_i - \log n \tilde{P}_i) \ .$$

Neyman proved that each of the statistics in (2-44), using any set of BAN estimators defined in (2-42, 2-43) $\{\tilde{P}_i\}$ has a limiting chi-square distribution with t degrees of freedom under the null hypothesis as the sample size, n, approaches infinity.

All three test criteria given in (2-44) are consistent for the null hypothesis being tested. That is under any admissible form of the alternative the power of each of the procedures tends to one as n approaches infinity for every fixed level of significance. Neyman showed that the three test criteria in (2-44) are asymptotically equivalent in the sense that the probability of the respective tests' contradicting each other tends to zero as n approaches infinity for every admissible hypothesis specifying either the null or alternative. The Neyman results were reported by Bhapkar (1966).

Mitra (1958) has shown that under a suitable sequence of alternatives tending to the null hypothesis at a suitable rate, the χ^2 statistic given in (2-44) has a limiting noncentral chi-square distribution if the $\{\tilde{P}_i\}$ are maximum likelihood estimators.

Bhapkar (1966) conjectured that Mitra's results should hold for any of the statistics in (2-44) using any system of BAN

estimators. This conjecture was based on the asymptotic equivalence of the chi-square statistics in (2-44) and the fact that any BAN estimators possess the asymptotic properties of the maximum likelihood estimators used by Mitra in his proofs.

The results of Neyman discussed in this subsection together with certain results which are now discussed serve as the basis for comparisons among the four procedures outlined in the second section of this chapter.

Some Results by Bhapkar

In this subsection, a major result by Bhapkar (1961) which links the modified minimum chi-square estimators to a quadratic form of the unbiased estimators of the $f_m(\underline{P})$ is presented. The result serves as a theoretical basis for the Grizzle, Starmer, and Koch (1969) methodology which uses a weighted least squares approach. The result also provides some common ground for comparing the four procedures presented in the second section.

Let the probability model in (2-38) be assumed. Also assume that n is large enough so that $n_i > 0$ for all i. Consider a null hypothesis defined by t linearly independent constraints on the $\{P_i\}$ (independent of $\Sigma_j P_i = 1$) of the form

$$H_0: \quad f_m(\underline{P}) = \sum_i f_{mi} P_i = 0 \qquad m = 1, 2, \ldots, t . \tag{2-45}$$

The f_{mi} are known constants such that the equations in (2-45), together with $\Sigma_i P_i = 1$, have at least one set of solutions for which the P_i's are all positive.

Let

$$\hat{f}_m = f_m(\underline{\hat{P}}) = \sum_i f_{mi}\ \hat{P}_i \qquad m = 1,2,\ldots,t$$

$$\underset{1 \times t}{\hat{\underline{f}}'} = [\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_t] \qquad\qquad (2\text{-}46)$$

$$\text{with} \quad \hat{P}_i = \frac{n_i}{n} \quad \text{for all } i\ .$$

The $\hat{f}_m$, $m = 1,2,\ldots,t$, are the unbiased estimates of the corresponding $f_m(\underline{P})$.

Using the multinomial probability model defined in (2-38) and the linearity of the covariance operator it can be shown that

$$\text{Cov}(\hat{f}_m, \hat{f}_{m'}) = \frac{1}{n} \sum_i f_{mi} f_{m'i}\ P_i(1 - P_i) - \frac{1}{n} \sum_{i \neq i'} f_{mi} f_{m'i'}\ P_i P_{i'}$$

$$\qquad\qquad (2\text{-}47)$$

$$= \frac{1}{n} \sum_i f_{mi} f_{m'i}\ P_i - \frac{1}{n} f_m(\underline{P}) f_{m'}(\underline{P}) = \phi_{mm'}\ .$$

Let the variance-covariance matrix of $\hat{\underline{f}}$ be denoted as Φ where

$$\underset{t \times t}{\Phi} = ((\phi_{mm'})) \quad m, m' = 1,2,\ldots,t\ . \qquad\qquad (2\text{-}48)$$

Let G denote the sample variance-covariance matrix of $\hat{\underline{f}}$ formed by replacing the P_i in Φ by their respective sample estimates $\hat{P}_i$ with

$$\underset{t \times t}{G} = ((g_{mm'})) \quad m, m' = 1,2,\ldots,t\ . \qquad\qquad (2\text{-}49)$$

The G matrix is a consistent estimator of Φ, the true variance-covariance matrix of $\hat{\underline{f}}$, both under the null hypothesis and alternative.

Theorem 2.1. $\quad$ Min $\quad \chi_1^2 = \hat{\underline{f}}' \, G^{-1} \, \hat{\underline{f}}$. $\qquad\qquad$ (2-50)
$\qquad\qquad$ Subject to H_0

The expression on the left side of the equality in (2-50) represents the value of the modified chi-square statistic (2-44) used in conjunction with modified minimum chi-square estimators, computed under the constraints of the null hypothesis. Theorem 2.1 shows that the χ_1^2 method to test the linear hypothesis in (2-45) is algebraically equivalent to a test statistic based upon the asymptotic normality of the unbiased estimators of the $f_m(\underline{P})$, whose variance-covariance matrix is estimated by the sample variance-covariance matrix. The quadratic form in (2-50) bears some resemblance to Stuart's statistic (2.6). The explicit relationship is specified later in this chapter. A modification and elaboration of the proof of Theorem 2.1 originally formulated by Bhapkar (1961) is given in Appendix A.

Bhapkar's result given by Theorem 2.1 serves as the theoretical basis for the Grizzle, Starmer and Koch (GSK) methodology presented earlier in the chapter. The GSK methodology is concerned with testing the null hypothesis that a specified linear model characterizes the data:

$$H_0: \quad \underset{u \times 1}{\underline{F}\,(\underline{P})} = \underset{u \times v}{X} \quad \underset{v \times 1}{\underline{B}} \qquad\qquad (2\text{-}51)$$

where X is a prespecified matrix of coefficients with full rank $v \le u$, and $\underline{B}$ is an unknown $(v \times 1)$ vector of parameters.

Bhapkar (1966) stated that there exists a $[(u - v) \times u]$ C matrix of full rank which is orthogonal to X such that

$$C \ \underline{F}(\underline{P}) \ = \ C \ X \ \underline{B} \ = \ \underset{(u-v) \times 1}{\underline{0}} \qquad . \qquad (2\text{-}52)$$

The model in (2-51) implies the $u - v$ constraint equations given

in (2-52). These constraints represent $u - v$ linearly independent

constraints on the components of the $\underline{P}$ vector.

As a concrete illustration of this abstract argument con-

sider the case of testing for marginal homogeneity in an $r \times r$

table where $r = 3$. The null hypothesis can be formulated in terms

of testing the fit of the linear model when the Koch and Reinfurt

(1971) approach based on GSK methodology is used.

$$\begin{bmatrix} M_{11} \\ M_{21} \\ M_{12} \\ M_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \qquad\qquad (2\text{-}53)$$

$$\underset{4 \times 1}{\underline{F}} \ (\underline{P}) \ = \ \underset{4 \times 2}{X} \ \underset{2 \times 1}{\underline{B}} \quad .$$

If the C matrix is taken to be

$$\underset{2 \times 4}{C} \ = \ \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \quad ,$$

C and X are orthogonal,

$$C \ X \ = \ \underset{2 \times 2}{0}$$

and the linear model in (2-53) gives rise to the constraint equations

$$C \ \underline{F}(\underline{P}) \ = \ \underline{0}$$

which can be written as

$$M_{11} - M_{21} = 0$$

$$M_{12} - M_{22} = 0 \ . \tag{2-54}$$

For this illustration testing for marginal homogeneity can be accomplished either by testing the fit of the model in (2-53) using the weighted least squares approach as given in (2-32 – 2-34) or by testing the null hypothesis as it is formulated in (2-54) in terms of linear constraints placed upon the components of the $\underline{P}$ vector. The constraint equations given by (2-54) can actually be written out in terms of the components of $\underline{P}$ by writing the marginal probabilities in terms of the P_i as given in (2-2). Testing the null hypothesis as it is formulated in (2-54) can be accomplished by using any of the techniques given in (2-44).

Bhapkar (1966) showed that testing a null hypothesis of the form given in (2-52) using the χ_1^2 statistic with modified minimum chi-square estimators computed under H_0 results in a test statistic which is algebraically equal to the statistic used to test the fit of the corresponding linear model given by (2-51). Bhapkar thus demonstrated that by thinking of the null hypothesis either in terms of a linear model (2-51) or in terms of a group of constraint equations which the linear model defines (2-52), two algebraically equivalent test procedures result.

If the null hypothesis is defined in terms of a linear model

$$H_0: \ \underset{u \times 1}{\underline{F}} \ (\underline{P}) = \underset{u \times v}{X} \ \underset{v \times 1}{\underline{B}} \tag{3-55}$$

the corresponding statistic to test H_0 is given by the sum of squares residual

$$\chi^2_{GSK} = (\underline{F}(\hat{\underline{P}}) - X\ \underline{\tilde{B}})'S^{-1}(\underline{F}(\hat{\underline{P}}) - X\ \underline{\tilde{B}}) \tag{2-33}$$

with $\underline{\tilde{B}}$ a weighted least estimator of $\underline{B}$ (2-34) and S a consistent estimator of the variance-covariance matrix of $\underline{F}(\hat{\underline{P}})$ under both the null and alternative hypotheses.

If the null hypothesis is defined in terms of the constraint equations

$$H_0: \quad \underset{(u-v)\times 1}{\underline{f}}(\underline{P}) = \underset{(u-v)\times u}{C}\ \underline{F}(\underline{P}) = \underset{(u-v)\times 1}{\underline{0}} \tag{2-52}$$

which are induced by the linear model, then it can be tested using the statistic

$$\chi^2_1 = \sum_i \frac{(n_i - n\tilde{P}_i)^2}{n_i} \tag{2-43}$$

where the $\{\tilde{P}_i\}$ are the modified minimum chi-square estimators computed under the constraints given in (2-52) and the additional constraint that $\sum_i P_i = 1$.

Bhapkar's (1966) result states that the statistics χ^2_{GSK} and χ^2_1 are algebraically equal whenever the latter is defined. Each has a limiting chi-square distribution with $u - v$ degrees of freedom under the null hypothesis. In addition by Theorem 2.1 the result

$$\chi^2_{GSK} = \underline{\hat{f}}'\ G^{-1}\ \underline{\hat{f}} \tag{2-56}$$

follows. The $(u - v) \times (u - v)$ matrix G is a consistent estimator of the variance-covariance matrix of $\underline{\hat{f}}$ under both the null and alternative hypotheses with

$$G = C S C' . \qquad (2\text{-}57)$$

χ_1^2 Statistic and Maximum Likelihood Estimators

It was shown by Theorem 2.1 that if a null hypothesis of the form (2-45)

$$H_0: \quad f_m(\underline{P}) = \sum_i f_{mi} P_i = 0 \quad m = 1,2,\ldots,t$$

is tested using the χ_1^2 statistic in conjunction with modified minimum chi-square estimators computed under the null hypothesis, a statistic results which is equal to the quadratic form

$$\chi_1^2 = \hat{\underline{f}}' \ G^{-1} \hat{\underline{f}}$$

with $\hat{\underline{f}}$ defined by (2-46) and G defined in (2-47 - 2-48).

Define

$$_0G = ((_0g_{mm'})) \quad m,m' = 1,2,\ldots,t$$

$$(2\text{-}58)$$

$$\text{with} \quad _0g_{mm'} = \frac{1}{n} \sum_i f_{mi} f_{m'i} \hat{P}_i \ .$$

The $_0G$ matrix is a consistent estimator of the variance-covariance matrix of $\hat{\underline{f}}$ provided the null hypothesis is true. From (2-47) and the definition of $_0G$ in (2-58) the relationship between G and $_0G$ can be written as

$$G = {_0G} - \frac{1}{n} \hat{\underline{f}} \, \hat{\underline{f}}' \ . \qquad (2\text{-}59)$$

In the computation of $_0G$, the zero vector is taken as an estimate of $\underline{f}$, which results in a consistent estimate only under the condition that the null hypothesis (2-45) holds.

Theorem 2.2. If the null hypothesis

$$H_0 : \quad f_m(\underline{P}) = \sum_i f_{mi} P_i = 0 \quad m = 1, 2, \ldots, t$$

is tested using the χ_1^2 statistic with approximate maximum likelihood estimators computed under H_0 then

$$\chi_1^{2*} \doteq \underline{\hat{f}}' {}_o G^{-1} \underline{\hat{f}} \tag{2-60}$$

where χ_1^{2*} is used to denote the modified chi-square statistic computed with approximate maximum likelihood estimators. The proof of Theorem 2.2, which is an extension and elaboration of a proof given by Bennett (1968), appears in Appendix B.

It can be shown that when $\underline{\hat{f}}$ is taken to be the vector $\underline{d}$ defined in (2-6) the quadratic form in (2-60) is the test statistic given by Stuart (1955) which was the first procedure introduced (2-5 - 2-7) to test for homogeneity of marginal distributions.

Explicit Relationships Among Statistics

The four statistics discussed in section two of the chapter are now compared from the standpoint of the research and results cited in (2-38 through 2-60). Under consideration are the χ_S^2 statistic of Stuart (1955), the χ_M^2 statistic of Madansky (1963) based upon the likelihood ratio criterion, the χ_I^2 statistic of Ireland et al. (1969) based upon minimum discrimination information estimation, and the χ_{GSK}^2 statistic of Koch and Reinfurt (1971) based upon fitting a specified linear model by weighted least squares.

There is an algebraic relationship between the X^2_{GSK} and X^2_S statistics. For the case of marginal homogeneity of the $r \times r$ table the linear model can be specified as

$$
H_0: \quad
\begin{bmatrix}
M_{11} \\
M_{21} \\
M_{12} \\
M_{22} \\
\cdot \\
\cdot \\
\cdot \\
M_{1r-1} \\
M_{2r-1}
\end{bmatrix}
=
\begin{bmatrix}
1 & 0 & \ldots & 0 \\
1 & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 \\
0 & 1 & \ldots & 0 \\
 & & \cdot & \\
 & & \cdot & \\
 & & \cdot & \\
0 & 0 & \ldots & 1 \\
0 & 0 & \ldots & 1
\end{bmatrix}
\begin{bmatrix}
B_1 \\
B_2 \\
\cdot \\
\cdot \\
\cdot \\
B_{r-1}
\end{bmatrix}
\qquad (2\text{-}61)
$$

using the Koch and Reinfurt approach. This model leads to the resulting constraint equations

$$
\begin{array}{lll}
M_{21} - M_{11} = 0 & & n(M_{21} - M_{11}) = 0 \\
M_{22} - M_{12} = 0 & \text{or equivalently} & n(M_{22} - M_{12}) = 0 \quad (2\text{-}62) \\
H_0: \quad \cdot & & \cdot \\
\qquad \cdot & & \cdot \\
\qquad \cdot & & \cdot \\
M_{2r-1} - M_{1r-1} = 0 & & n(M_{2r-1} - M_{1r-1}) = 0
\end{array}
$$

Let $d_i = n(\hat{M}_{2i} - \hat{M}_{1i})$, $i = 1,2,\ldots,r-1$

$$
\underset{1 \times (r-1)}{\underline{d}'} = [d_1, d_2, \ldots, d_{r-1}] \ .
$$

The result that $X^2_{GSK} = \underline{\hat{f}}' \, G^{-1} \underline{\hat{f}}$ with $\underline{\hat{f}}$ taken as $\underline{d}$ and the relationship between G and $_oG$ given in (2-59) with $_oG$ taken as $\hat{\Sigma}_{\underline{d}}$ defined by (2-7) can be used to write the X^2_{GSK} statistic to test H_0 as

$$\chi^2_{GSK} = \underline{d}'[\hat{\Sigma}_{\underline{d}} - \frac{1}{n}\underline{d}\,\underline{d}']^{-1}\underline{d} \tag{2-63}$$

For notational simplicity let A equal $\hat{\Sigma}_{\underline{d}}$

Claim: By Ireland et al. (1969)

$$[A - \frac{1}{n}\underline{d}\,\underline{d}']^{-1} = A^{-1} + \frac{A^{-1}\underline{d}\,\underline{d}'\,A^{-1}}{[n - \underline{d}'A^{-1}\underline{d}]} \tag{2-64}$$

Proof of Claim

$$\text{Show} \quad [A - \frac{1}{n}\underline{d}\,\underline{d}']\left[A^{-1} + \frac{A^{-1}\underline{d}\,\underline{d}'A^{-1}}{[n - \underline{d}'A^{-1}d]}\right] = I \tag{2-65}$$

where I is the identity matrix of dimensions $(r-1) \times (r-1)$.

The product of the two matrices in (2-65) can be written as

$$AA^{-1} - \frac{1}{n}\underline{d}\,\underline{d}'A^{-1} + \frac{AA^{-1}\underline{d}\,\underline{d}'A^{-1}}{n - \underline{d}'A^{-1}\underline{d}} - \frac{\underline{d}\,\underline{d}'A^{-1}\underline{d}\,\underline{d}'A^{-1}}{n(n - \underline{d}'A^{-1}\underline{d})}$$

$$= AA^{-1} + \frac{(-n\underline{d}\,\underline{d}'A^{-1}) + \underline{d}\,\underline{d}'A^{-1}\underline{d}'A^{-1}\underline{d} + n\underline{d}\,\underline{d}'A^{-1} - \underline{d}\,\underline{d}'A^{-1}\underline{d}\,\underline{d}'A^{-1}}{n(n - \underline{d}'A^{-1}\underline{d})}$$

$$= AA^{-1} + \frac{\underline{d}\,\underline{d}'A^{-1}\underline{d}A^{-1}\underline{d} - \underline{d}(\underline{d}'A^{-1}\underline{d})\underline{d}'A^{-1}}{n(n - \underline{d}'A^{-1}\underline{d})}$$

$$= AA^{-1} + \frac{\underline{d}\,\underline{d}'A^{-1}\underline{d}A^{-1}\underline{d} - \underline{d}\,\underline{d}'A^{-1}(\underline{d}'A^{-1}\underline{d})}{n(n - \underline{d}'A^{-1}\underline{d})}$$

$$= AA^{-1} + \underset{(r-1)\times(r-1)}{0} = I \quad .$$

In a similar manner it can be shown that

$$\left[A^{-1} + \frac{A^{-1}\underline{d}\,\underline{d}'A^{-1}}{[n - \underline{d}'A^{-1}\underline{d}]}\right][A - \frac{1}{n}\underline{d}\,\underline{d}'] = I \quad .$$

Substituting the result in (2-64) into (2-63) one obtains

$$X^2_{GSK} = \underline{d}' \left[A^{-1} + \frac{A^{-1}\underline{d}\ \underline{d}'A^{-1}}{[n - \underline{d}'A^{-1}\underline{d}]} \right] \underline{d} = \underline{d}'A^{-1}\underline{d} + \frac{\underline{d}'A^{-1}\underline{d}\ \underline{d}'A^{-1}\underline{d}}{[n - \underline{d}'A^{-1}\underline{d}]}$$

$$= \frac{n\ \underline{d}'A^{-1}\underline{d} - (\underline{d}'A^{-1}\underline{d})^2 + (\underline{d}'A^{-1}\underline{d})^2}{[n - \underline{d}'A^{-1}\underline{d}]} = \frac{\underline{d}'A^{-1}\underline{d}}{1 - \dfrac{\underline{d}'A^{-1}\underline{d}}{n}}$$

$$= \frac{\underline{d}'\ \hat{\Sigma}_{\underline{d}}^{-1}\ \underline{d}}{1 - \dfrac{1}{n}\underline{d}'\hat{\Sigma}_{\underline{d}}^{-1}\underline{d}} \ ; \tag{2-66}$$

but $\underline{d}'\ \hat{\Sigma}_{\underline{d}}^{-1}\underline{d}$ is the Stuart statistic. Hence from (2-66) the re-

lationship between X^2_{GSK} and X^2_S can be stated as

$$X^2_{GSK} = \frac{X^2_S}{1 - \dfrac{1}{n}X^2_S} \quad \text{or} \quad X^2_S = \frac{X^2_{GSK}}{1 + \dfrac{1}{n}X^2_{GSK}} . \tag{2-67}$$

As $n \to \infty$ under H_0, $X^2_{GSK} \to X^2_S$ although $X^2_{GSK} \geq X^2_S$ for all finite

n. When n is large and the components of $\underline{d}$ are small the X^2_{GSK}

and X^2_S statistics will be close in value. The X^2_{GSK} provides a

more powerful test under forms of the alternative hypothesis since

it uses a variance-covariance matrix whose variance terms are

smaller in value than those used by X^2_S. In addition to the algebraic

relationships given in (2-67), there is an underlying relationship

among all four statistics. Each of the techniques either is itself

or reduces to a large sample chi-square test based upon the use of

BAN estimators. The X^2_S statistics of Stuart can be shown to be

approximately equal to the modified chi-square statistic X^2_1 (2-44)

used in conjunction with approximate maximum likelihood estimators.

It was also shown that the X^2_{GSK} statistic of Koch and Reinfurt

is equal to the modified chi-square statistic X^2_1 used in conjunction

with modified minimum chi-square estimators. The X^2_M statistic of

Madansky is the likelihood ratio test statistic (2-44) used in conjunction with maximum likelihood estimators. According to Neyman (1949) each of these three techniques possesses the same asymptotic properties and are all asymptotically equivalent. Ireland et al. (]969) pointed out that by using results of Taylor (1953) and Neyman (1949) it can be demonstrated that minimum discrimination information estimators are also BAN and the X_I^2 statistic thus belongs to the same general class of statistics as X_{GSK}^2, X_M^2 and X_S^2.

The technique of estimating a set of cell probabilities $\{P_i\}$ under the constraints of a given null hypothesis using MDIE consists of finding that $\underline{P}$ vector whose components satisfy the null hypothesis and which minimizes the function

$$I(\underline{P},\hat{\underline{P}}) = \sum_i P_i \log \frac{P_i}{\hat{P}_i} \ . \tag{2-68}$$

If, instead of minimizing the function in (2-68), a $\underline{P}$-vector is chosen which minimizes the function

$$I(\hat{\underline{P}},\underline{P}) = \sum_i \hat{P}_i \log \frac{\hat{P}_i}{P_i} \ , \tag{2-69}$$

an estimation procedure results which is also referred to as minimum discrimination estimation by Kullback (1968). The estimates found using (2-68) and (2-69) are, in general, different.

When (2-69) is rewritten as

$$\begin{aligned}
I(\hat{\underline{P}},\underline{P}) &= \sum_i \frac{n_i}{n} \log \frac{n_i}{n} - \sum_i \frac{n_i}{n} \log P_i \\
&= \text{constant}_1 - \frac{1}{n} \sum_i n_i \log P_i \\
&= \text{constant}_1 - \frac{1}{n} \log \prod_i P_i^{n_i}
\end{aligned} \tag{2-70}$$

it is seen that the value of $\underline{P}$ which minimizes $I(\hat{\underline{P}},\underline{P})$ is the same as that value of $\underline{P}$ which maximizes the quantity

$$\prod_i P_i^{n_i} = L(\underline{P}) + constant_2 \qquad (2\text{-}71)$$

where $L(\underline{P})$ is the likelihood function defined in (2-10). The equivalence of minimum discrimination estimation and maximum likelihood estimation thus follows from this last result, provided the distance-like measure defined by (2-69) is used. The technique of MDIE has appeared in the literature using both definitions of distance, (2-68) and (2-69). When (2-69) is used, the resulting χ_I^2 statistic which is given by

$$\chi_I^2 = 2n \sum_i \hat{P}_i \log \frac{\hat{P}_i}{\tilde{P}_i} \qquad (2\text{-}72)$$

is equal to the likelihood ratio test statistic given by (2-44) since $\{\tilde{P}_i\}$ are maximum likelihood estimators. When the distance-like measure defined in (2-68) is used, the χ_I^2 statistic is generally not the same as the likelihood ratio statistic. Berkson (1972) noted that the differences between the two estimation procedures defined by the two different distance functions are analogous to the differences between the modified minimum chi-square and minimum chi-square estimation procedures defined in (2-42) respectively. The point of comparison is the fact that $I(\underline{P},\hat{\underline{P}})$ and $I(\hat{\underline{P}},\underline{P})$ interchange the observed and estimated probabilities as do the respective chi-square procedures. Berkson then stated that the χ_I^2 statistic using estimation procedure (2-68) is to the χ_1^2

statistic (2-44) as the χ^2_I statistic using estimation procedure (2-69) is to the χ^2 statistic (2-44). The analogy is a conceptual rather than a mathematical statement. It is conjectured that Ireland et al. (1969) chose to use the distance-like function given by (2-68) rather than that given by (2-69) to define their statistic because of the difficulty in computing maximum likelihood estimators for the problem of testing for marginal homogeneity. The use of the (2-69) definition would require the computation of maximum likelihood estimators.

This somewhat lengthy discussion of the X^2_I statistic and two associated estimation procedures has been given in order to relate this statistic to the better known statistics in (2-44) which have served as the basis for comparing the techniques discussed in the second section of the chapter. It should be pointed out that the comparisons made are all based upon large sample theory.

In Chapter III an explicit statement of a statistic which is algebraically equivalent to X^2_{GSK} is given. The statistic has the advantage of not requiring a knowledge of linear models as is the case for X^2_{GSK}. A development of the large sample distribution of the statistic is given. In addition post hoc procedures which are used to locate sources of significance are developed and illustrated.

CHAPTER III

THE χ^2 STATISTIC AND ASSOCIATED POST HOC PROCEDURES

THE χ^2 STATISTIC

In Chapter II four different techniques to test for homogeneity of the marginal distributions in the mixed categorical data model were given. The X_I^2 statistic of Ireland et al. (1969) and the X_M^2 statistic of Madansky (1963) each require iterative procedures to obtain estimates of the cell probabilities computed under the null hypothesis. The X_{GSK}^2 statistic of Koch and Reinfurt (1971) does not require the estimation of the individual cell probabilities but does require the estimation of a vector of parameters specified by the linear model which is to be fitted to the data. In the Koch and Reinfurt procedure the parameters of the model are estimated using a weighted least squares technique and the estimates are computed using a straightforward chain of matrix multiplications and inversions. Although the X_{GSK}^2 statistic is less computationally difficult than either the X_I^2 or X_M^2 statistics, it does require a certain knowledge of linear models and weighted least squares analysis which may not be possessed by potential users of the technique.

The statistic developed in this chapter has the advantage of both computational simplicity relative to the X_M^2 and X_I^2 statistics and conceptual simplicity relative to the X_{GSK}^2

50

statistic in that a knowledge of linear models is not required in order to understand the technique. If the null hypothesis of marginal homogeneity is stated in terms of t linearly independent constraint equations, each linearly independent of the constraint $\sum_i P_i = 1$, and is written in the form

$$H_0: \quad f_m = \sum_i f_{mi} P_i = 0 \quad \text{for} \quad m = 1, \ldots, t \qquad (3\text{-}1)$$

a statistic to test H_0 can be written as

$$x^2 = \hat{\underline{f}}' \, G^{-1} \, \hat{\underline{f}}$$

with $\hat{\underline{f}}$ and G as defined in (2-46) and (2-49) of Chapter II respectively. The structure of the null hypothesis and associated test statistic in (3-1) serve as the general framework for developing an explicit statistic to test for marginal homogeneity in the mixed categorical data model of order-d.

Model and Hypothesis

Throughout the discussion it is assumed that d repeated measures are taken on each of n subjects and the measures are based on the same r-level categorical dependent variable. For notational simplicity the numbers $1, 2, \ldots, r$ are used to index the levels of the dependent variable.

The data from such a design can be represented as an $r \times r \times \ldots \times r$ contingency table of d dimensions with a multinomial distribution assumed to fit the observed set of r^d frequencies. Any cell in the contingency table can be represented

I

as an ordered d-tuple each of whose components represents a particular level of the dependent variable on a particular measure. In general any cell can be represented in the form

$$(j_1, j_2, \ldots, j_d), \quad j_i = 1, 2, \ldots, r \quad \forall i = 1, 2, \ldots, d.$$

Following the notation introduced in Chapter II let $n_{j_1, \ldots, j_d}$ denote the number of observations which fall in cell $(j_1, \ldots, j_d)$. Let $P_{j_1, \ldots, j_d}$ denote the probability of a subject's set of responses falling into cell $(j_1, \ldots, j_d)$ with the cell probabilities summing to one. Let M_{qb} denote the marginal probability that a subject is classified by category b on the qth measure with

$$M_{qb} = \sum_{\substack{(j_1, \ldots, j_d) \\ \text{with } j_q = b}} \cdots \sum P_{j_1, \ldots, j_d}, \quad \begin{array}{l} q = 1, \ldots, d \\ b = 1, \ldots, r \end{array} \qquad (3\text{-}2)$$

An unbiased estimator of $P_{j_1, \ldots, j_d}$ is

$$\hat{P}_{j_1, \ldots, j_d} = \frac{n_{j_1, \ldots, j_d}}{n} \qquad (3\text{-}3)$$

and an unbiased estimator of M_{qb} is

$$\hat{M}_{qb} = \sum_{\substack{(j_1, \ldots, j_d) \\ \text{with } j_q = b}} \cdots \sum \hat{P}_{j_1, \ldots, j_d}$$

with $\displaystyle\sum_{b=1}^{r} M_{qb} = \sum_{b=1}^{r} \hat{M}_{qb} = 1 \quad \forall q = 1, \ldots, d$.

Define

$$\bar{M}_{.b} = \frac{1}{d} \sum_{q=1}^{d} M_{qb}, \quad \hat{\bar{M}}_{.b} = \frac{1}{d} \sum_{q=1}^{d} \hat{M}_{qb} \quad \forall b = 1,\ldots,r . \qquad (3\text{-}4)$$

Using the definitions of $\bar{M}_{.b}$ and $\hat{\bar{M}}_{.b}$ and the constraints on the marginal probabilities given in (3-3) it follows that

$$\sum_{b=1}^{r} (M_{qb} - \bar{M}_{.b}) = \sum_{b=1}^{r} (\hat{M}_{qb} - \hat{\bar{M}}_{.b}) = 0 \quad \forall q = 1,\ldots,d \qquad (3\text{-}5)$$

and

$$\sum_{q=1}^{d} (M_{qb} - \bar{M}_{.b}) = \sum_{q=1}^{d} (\hat{M}_{qb} - \hat{\bar{M}}_{.b}) = 0 \quad \forall b = 1,\ldots,r . \qquad (3\text{-}6)$$

The null hypothesis of equality of the d marginal distributions can be stated as

$$H_0: \quad M_{1b} = M_{2b} = \ldots = M_{db} \quad \forall b = 1,\ldots,r \quad \text{or equivalently as}$$

$$H_0: \quad \underset{(dr \times 1)}{\underline{V}} = \underset{(dr \times 1)}{\underline{0}} \qquad (3\text{-}7)$$

where

$$\underset{(1 \times d)}{\underline{V}'_b} = [M_{1b} - \bar{M}_{.b}, \; M_{2b} - \bar{M}_{.b}, \ldots, \; M_{db} - \bar{M}_{.b}] \quad \forall b = 1,\ldots,r \qquad (3\text{-}8)$$

and

$$\underset{(1 \times dr)}{\underline{V}'} = [\underline{V}'_1, \; \underline{V}'_2, \ldots, \; \underline{V}'_r] \quad .$$

The corresponding sample estimates are given by

$$\hat{\underline{V}}'_b = [\hat{M}_{1b} - \hat{\bar{M}}_{.b}, \; \hat{M}_{2b} - \hat{\bar{M}}_{.b}, \ldots, \hat{M}_{db} - \hat{\bar{M}}_{.b}] \qquad (3\text{-}9)$$

and

$$\hat{\underline{V}}' = [\hat{\underline{V}}'_1, \; \hat{\underline{V}}'_2, \ldots, \; \hat{\underline{V}}'_r] \quad .$$

In the next subsection a statistic for testing H_0: $\underline{V} = \underline{0}$ is developed and its large sample distribution is derived. The test statistic is a quadratic form which involves a reduced form of the vector $\hat{\underline{V}}$, $\hat{\underline{V}}^*$, and the sample variance-covariance matrix $\hat{\Sigma}_{\hat{\underline{V}}^*}$ which is a consistent estimator of $\Sigma_{\hat{\underline{V}}^*}$. The test statistic has the form

$$\chi^2 = \hat{\underline{V}}^{*'} \, \hat{\Sigma}^{-1}_{\hat{\underline{V}}^*} \, \hat{\underline{V}}^* \tag{3-10}$$

which is shown under H_0 and for large n to be approximately chi-square with $(d-1)(r-1)$ degrees of freedom.

Large Sample Distribution of χ^2 Under H_0

Because each subject is measured d times, the outcome for subject i can be represented as falling into cell $(j_{1i}, j_{2i}, \ldots, j_{di})$. Define a set of Bernoulli random variables on the sample space of outcomes as follows:

$$X_{a_1, a_2, \ldots, a_d}((j_1, j_2, \ldots, j_d)) = 1 \quad \text{if} \quad a_k = j_k \quad \forall k = 1, \ldots, d$$
$$= 0 \quad \text{otherwise .} \tag{3-11}$$

In all there will be r^d such Bernoulli random variables with one being defined for each value of a_k, $a_k = 1, 2, \ldots, r$ and $k = 1, 2, \ldots, d$. One can now represent the d responses for each subject i as a vector of the form

$$
\underset{r^d \times 1}{\underline{X}_i} = \begin{bmatrix} X_{1,1,\ldots,1}((j_{1i},j_{2i},\ldots,j_{di})) \\[6pt] X_{1,1,\ldots,2}((j_{1i},j_{2i},\ldots,j_{di})) \\[2pt] \vdots \\[2pt] X_{1,1,\ldots,r}((j_{1i},j_{2i},\ldots,j_{di})) \\[2pt] \vdots \\[2pt] X_{j_1,j_2,\ldots,j_d}((j_{1i},j_{2i},\ldots,j_{di})) \\[2pt] \vdots \\[2pt] X_{r,r,\ldots,r}((j_{1i},j_{2i},\ldots,j_{di})) \end{bmatrix} \quad . \qquad (3\text{-}12)
$$

The vector in (3-12) contains $r^d - 1$ zeros and a single 1 which indicates the cell into which subject i's set of observations falls. The random vector $\underline{X}_i$ has a multinomial distribution with parameters $\{P_{j_1,\ldots,j_d}\}$ and n = 1.

Let

$$
\underset{1 \times r^d}{\underline{P}'} = [P_{j_1,\ldots,j_d}] \quad \text{and}
$$

$$
\underset{1 \times r^d}{\hat{\underline{P}}'} = [\hat{P}_{j_1,\ldots,j_d}] \quad . \qquad (3\text{-}13)
$$

By the definition of $\underline{X}_i$, $\hat{\underline{P}}$ can be written in the form

$$
\hat{\underline{P}} = \frac{1}{n} \sum_{i=1}^{n} \underline{X}_i \quad . \qquad (3\text{-}14)
$$

By the definition of the mixed categorical data model it is assumed that the n subjects are a random sample from a homogeneous population. Then it follows that $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$ are independent, and identically distributed. By a form of the multivariate central limit theorem given in Rao (1965)

$$\sqrt{n}\ (\hat{\underline{P}} - \underline{P}) \overset{\mathcal{D}}{\to} \underline{W} \sim N_{r^d}(\underline{0}\ \Sigma) \tag{3-15}$$

where $\Sigma = \Sigma_{\underline{X}_1}$. For a fixed n, where n is taken large, it follows from (3-15) that $\hat{\underline{P}}$ is approximately $N_{r^d}(\underline{P}, \frac{1}{n}\Sigma)$· and this approximation improves as n approaches infinity. Because a multinomial distribution is assumed to characterize the cell frequencies

$$\underset{\substack{\hat{\underline{P}} \\ (r^d \times r^d)}}{\Sigma} = \frac{1}{n}\ [D_{\underline{P}} - \underline{P}\ \underline{P}'] \tag{3-16}$$

where $D_{\underline{P}}$ is an $(r^d \times r^d)$ diagonal matrix with elements of the $\underline{P}$ vector on the main diagonal.

Let

$$\underset{1 \times d}{\hat{\underline{M}}'_b} = [\hat{M}_{1b}, \hat{M}_{2b}, \ldots, \hat{M}_{db}] \quad \forall b = 1, \ldots, r \quad \text{and}$$

$$\underset{1 \times dr}{\hat{\underline{M}}'} = [\hat{\underline{M}}'_1, \hat{\underline{M}}'_2, \ldots, \hat{\underline{M}}'_r]\ . \tag{3-17}$$

It is possible to find a matrix $\underset{dr \times r^d}{K_d}$, the elements of which are zeros and ones such that $K\ \hat{\underline{P}} = \hat{\underline{M}}$. This can be done because $\hat{\underline{M}}$ is a vector of estimates of marginal probabilities and $\hat{\underline{P}}$ is a vector of estimates of joint probabilities which, when summed appropriately, yield the marginal estimates.

The following proposition from multivariate statistics is used to derive the approximate sampling distribution of $\hat{\underline{V}}$ for large n.

<u>Proposition 1.</u> If $\underset{k \times 1}{\underline{Y}} \sim N_k(\underline{u}, \Sigma_{\underline{Y}})$ and if $\underset{m \times k}{L}$ is a matrix of constants then $L\underline{Y} \sim N_m(L\underline{u}, L\ \underset{\hat{\underline{Y}}}{\Sigma}\ L')$.

Using the facts that $\hat{\underline{P}} \sim N_{r^d}(\underline{P}, \Sigma_{\underline{P}})$ for large n, and $K\hat{\underline{P}} = \hat{\underline{M}}$ from Proposition 1 it follows that the distribution of $\hat{\underline{M}}$ can be

approximated by

$$\hat{\underline{M}} \sim N_{dr}(L \underline{P}, L \Sigma L') \quad . \qquad (3\text{-}18)$$

It is now possible to find a matrix $\underset{dr \times dr}{A}$ such that $\hat{\underline{V}} = A \hat{\underline{M}}$ where $\hat{\underline{V}}$ is the vector defined by (3-8). Matrix A, which is a block-diagonal matrix, has the form

$$A = \begin{bmatrix} D_1 & & & \\ & D_2 & & 0 \\ & & \ddots & \\ 0 & & & D_r \end{bmatrix} \quad \text{where} \quad \underset{d \times d}{D_m} = ((d_{ij})) \quad d_{ij} = -\frac{1}{d} \quad \text{for} \quad i \neq j$$

$$= \frac{d-1}{d} \quad \text{for} \quad i = j \qquad (3\text{-}19)$$

$$\forall m = 1, 2, \ldots, r \quad .$$

By Proposition 1, the distribution of $\hat{\underline{V}} = A \hat{\underline{M}}$ can be approximated by a multivariate normal of the form $N_{dr}(A L \underline{P}, A L \Sigma L'A')$ when n is large. Because $\hat{\underline{V}}$ is an unbiased estimator of $\underline{V}$, the null hypothesis $\underline{V} = \underline{0}$ can be stated as $E(\hat{\underline{V}}) = \underline{0}$; hence under H_0, $\hat{\underline{V}}$ is approximately distributed as

$$\hat{\underline{V}} \sim N_{dr}(\underline{0}, \Sigma_{\hat{V}}), \quad \text{with} \quad \Sigma_{\hat{V}} = A L \Sigma L'A' \quad . \qquad (3\text{-}20)$$

The following result from multivariate statistics provides the basis for constructing a statistic to test H_0: $\underline{V} = \underline{0}$ or equivalently to test H_0: $E(\hat{\underline{V}}) = \underline{0}$:

Proposition 2. If $\underset{k \times 1}{\underline{Y}} \sim N_k(\underline{u}, \Sigma_Y)$ and Σ_Y is of full rank, then

$$(\underline{Y} - \underline{u})'\Sigma_Y^{-1}(\underline{Y} - \underline{u}) \sim \chi_k^2.$$

Let $\hat{\Sigma}_{\hat{V}}$ denote a sample estimate of $\Sigma_{\hat{V}}$ found by replacing $\underline{P}$ in $\Sigma_{\hat{V}}$ by $\hat{\underline{P}}$. Because $\hat{\Sigma}_{\hat{V}}$ is a consistent estimator of $\Sigma_{\hat{V}}$ a statistic to test H_0 is given by

$$\chi^2 = (\hat{\underline{V}} - \underline{0})' \hat{\Sigma}_{\hat{\underline{V}}}^{-1} (\hat{\underline{V}} - \underline{0}) = \hat{\underline{V}}' \hat{\Sigma}_{\hat{\underline{V}}}^{-1} \hat{\underline{V}} \ . \tag{3-21}$$

The statistic in (3-21) relies on (3-20) and the substitution of $\hat{\Sigma}_{\hat{\underline{V}}}^{-1}$, a consistent estimator of $\Sigma_{\hat{\underline{V}}}^{-1}$. If $\hat{\Sigma}_{\hat{\underline{V}}}$ were of full rank then χ^2 would have an approximate χ^2 distribution with dr

degrees of freedom under H_0. Large values of χ^2 would provide

evidence against H_0 and hence the upper tail of χ^2 could serve

as the rejection region. The only drawback to the preceding

argument is that $\Sigma_{\hat{\underline{V}}}$ and $\hat{\Sigma}_{\hat{\underline{V}}}$ are both of deficient rank.

The rank of $\Sigma_{\hat{\underline{V}}}$ is determined by the number of linearly

independent components in the random vector $\hat{\underline{V}}$. The components of

$\hat{\underline{V}}$ are themselves random variables which can be thought of as

elements in the inner product space L_2 , the space of all random

variables with finite second moments with an inner product defined

as follows:

$$\forall U_i, U_j \in L_2 \quad , \quad (U_i, U_j) = \text{Cov}(U_1, U_j) \ . \tag{3-22}$$

Claim: There are $(d-1)(r-1)$ linearly independent components in $\hat{\underline{V}}$.

The legitimacy of this claim rests on the restrictions

introduced in (3-5 and 3-6). The restriction $\sum_{q=1}^{d} (\hat{M}_{qb} - \overline{\hat{M}}_{.b}) = 0$

$\forall b = 1, \ldots, r$ implies that r linearly dependent vectors

are introduced, one for each value of b. For a fixed value of b,

once any $d-1$ terms in $\sum_{q=1}^{d} (\hat{M}_{qb} - \overline{\hat{M}}_{.b})$ are known, the other term

is determined and redundancy exists in the definition of $\hat{\underline{V}}$. The

restriction $\sum_{b=1}^{r} (\hat{M}_{qb} - \overline{\hat{M}}_{.b}) = 0$ $\forall q = 1, \ldots, d$ implies that d

linearly dependent vectors are introduced, one for each value of

q. Knowing any $r-1$ terms in $\sum_{b=1}^{r} (\hat{M}_{qb} - \overline{\hat{M}}_{.b})$ determines the

remaining term. Again, a redundancy is introduced. In all,

d + r - 1 redundant components are introduced since one component

is counted twice. Consequently there will be dr - (d + r - 1) =

(d - 1)(r - 1) linearly independent components in $\hat{\underline{V}}$.

No information will be lost be deleting any (d + r - 1)

redundant components. The goal is to form a new vector $\hat{\underline{V}}^*$ from

$\hat{\underline{V}}$ by choosing any (d-1)(r-1) linearly independent components

from $\hat{\underline{V}}$. $\hat{\underline{V}}^*$ will be approximately $\hat{\underline{V}}^* \sim N_{(d-1)(r-1)}(0, \Sigma_{\hat{\underline{V}}^*})$

under H_0 since any marginal distribution of a multivariate

normal distribution is itself multivariate normal. The matrix

$\Sigma_{\hat{\underline{V}}^*}$ is then of full rank and for n large, $\hat{\Sigma}_{\hat{\underline{V}}^*}$ will also be of

full rank. As a consequence of Proposition 2, a statistic χ^2

can be defined which, for large n, is approximately distributed

under H_0 as

$$\chi^2 = \hat{\underline{V}}^{*'} \hat{\Sigma}_{\hat{\underline{V}}^*}^{-1} \hat{\underline{V}}^* \sim \chi^2_{(d-1)(r-1)} . \qquad (3\text{-}23)$$

The vector $\hat{\underline{V}}^*$ is formed from $\hat{\underline{V}}$ by choosing any (d-1)(r-1)

linearly independent components of $\hat{\underline{V}}$. Define for each b,

b = 1,2,...,r

$$\underset{1\times(d-1)}{\hat{\underline{V}}_b^{*'}} = [\hat{M}_{1b} - \bar{\hat{M}}_{.b},\dots,\hat{M}_{m_b-1\ b} - \bar{\hat{M}}_{.b}, \hat{M}_{m_b+1} - \bar{\hat{M}}_{.b},\dots, \hat{M}_{db} - \bar{\hat{M}}_{.b}]$$
$$(3\text{-}24)$$

where m_b is any integer such that $1 \le m_b \le d$ ∀b and

$$\underset{1\times(d-1)(r-1)}{\hat{\underline{V}}^{*'}} = [\hat{\underline{V}}_1^{*'},\dots,\hat{\underline{V}}_{h-1}^{*'}, \hat{\underline{V}}_{h+1}^{*'},\dots,\hat{\underline{V}}_r^{*'}] \qquad (3\text{-}25)$$

for any h such that $1 \le h \le r$. The specific $\hat{\underline{V}}^*$ which is formed

is determined by the choice of the m_b's and h. If $_1\hat{\underline{V}}^*$ and

$_2\hat{\underline{V}}^*$ represent two different vectors formed by two different

choices of $(d-1)(r-1)$ linearly independent components from $\hat{\underline{V}}$,

then

$$\chi^2 = {}_1\hat{\underline{V}}^{*'}\hat{\Sigma}^{-1}_{{}_1\hat{\underline{V}}^*}\, {}_1\hat{\underline{V}}^* = {}_2\hat{\underline{V}}^{*'}\hat{\Sigma}^{-1}_{{}_2\hat{\underline{V}}^*}\, {}_2\hat{\underline{V}}^* \ . \tag{3-26}$$

Invariance of χ^2

The invariance of χ^2 under mode of formation of $\hat{\underline{V}}^*$ is

demonstrated in this subsection. The invariance follows as a con-

sequence of the following theorem.

Theorem 3.1. Suppose that $\underset{1 \times t}{\underline{W}'} = [W_1, W_2, \ldots, W_t]$ such that

$W_i \in L_2$ $\forall$ i, $1 \leq i \leq t$ and also suppose that the W_i's generate

a subspace S of L_2 such that $\dim(S) = k$ where $k < t$. Now let

$\underset{1 \times k}{\underline{X}'} = [X_1, X_2, \ldots, X_k]$ where $\underline{X}$ is formed by choosing a set of k

linearly independent components from $\underline{W}$ and let

$\underset{1 \times k}{\underline{Y}'} = [Y_1, Y_2, \ldots, Y_k]$ where $\underline{Y}$ is formed by choosing a different

set of k linearly independent components from $\underline{W}$. Suppose $\hat{\Sigma}_{\underline{X}}$ is

a function from $(L_2)^k$ onto the class of k × k nonsingular

variance-covariance matrices such that $\hat{\Sigma}_{T\underline{X}} = T\hat{\Sigma}_{\underline{X}}T'$ for all k × k

real matrices T. Then

$$\underline{X}'\hat{\Sigma}^{-1}_{\underline{X}}\underline{X} = \underline{Y}'\hat{\Sigma}^{-1}_{\underline{Y}}\underline{Y} \ . \tag{3-27}$$

Proof. By definition of a basis $\{X_1, X_2, \ldots, X_k\}$ and $\{Y_1, Y_2, \ldots, Y_k\}$

each forms a basis for S. Hence $\underline{Y} = T\underline{X}$ where T is a k × k

matrix of full rank. This result follows because one basis can be

expressed as a nonsingular linear transformation of another basis

for the same space S. By substitution

$$\underline{Y}'\hat{\Sigma}_{\underline{Y}}^{-1} \ \underline{Y} = \underline{X}'T'\hat{\Sigma}_{T\underline{X}}^{-1} \ T\underline{X} \ . \tag{3-28}$$

By hypothesis

$$\hat{\Sigma}_{T\underline{X}} = T \ \hat{\Sigma}_{\underline{X}} \ T' \ . \tag{3-29}$$

Substituting (3-29) into (3-28)

$$\underline{Y}'\hat{\Sigma}_{\underline{Y}}^{-1}\underline{Y} = \underline{X}'T'(T\hat{\Sigma}_{\underline{X}}T')^{-1}T\underline{X} = \underline{X}'T'(T')^{-1}\hat{\Sigma}_{\underline{X}}^{-1}T^{-1}T\underline{X} \ . \tag{3-30}$$

From (3-30) the final result

$$\underline{Y}'\hat{\Sigma}_{\underline{Y}}^{-1} \ \underline{Y} = \underline{X}'\hat{\Sigma}_{\underline{X}}^{-1} \ \underline{X} \tag{3-31}$$

is obtained.

The invariance of χ^2 under mode of formation of $\underline{\hat{V}}^*$ follows directly from Theorem 3.1. In the specific case under consideration, two vectors $_1\underline{\hat{V}}^*$ and $_2\underline{\hat{V}}^*$ whose components generate the same subspace S of dimension $(d-1)(r-1)$ are formed from the vector $\underline{\hat{V}}$ by choosing any $(d-1)(r-1)$ linearly independent components from $\underline{\hat{V}}$. In terms of the more general theorem, $\underline{\hat{V}}$, $_1\underline{\hat{V}}^*$, and $_2\underline{\hat{V}}^*$ play the roles of $\underline{W}$, $\underline{X}$, and $\underline{Y}$ respectively. The result in (3-31) with the appropriate substitutions made verifies the invariance of the χ^2 statistic under mode of formation of $\underline{\hat{V}}^*$.

Computational Form of χ^2

In this subsection the necessary formulas needed in the computation of the χ^2 statistic are provided.

The majority of the work in calculating the χ^2 statistic given by $\underline{\hat{V}}^{*'}\hat{\Sigma}_{\underline{\hat{V}}^*}^{-1} \underline{\hat{V}}^*$ goes into finding the variance-covariance

matrix. Instead of calculating $\hat{\Sigma}_{\underline{\hat{V}}^*}$ directly, it is easier to write

$$\underline{\hat{V}}^* = \underset{(d-1)(r-1)\times dr}{A^*} \underline{\hat{M}} \quad \text{and} \quad \hat{\Sigma}_{\underline{\hat{V}}^*} = A^* \hat{\Sigma}_{\underline{\hat{M}}} A^{*'} \; .$$

Associated with each $\underline{\hat{V}}^*$ there are numbers $m_1, m_2, \ldots, m_r$ and h defined in (3-24) and (3-25). The matrix A^* associated with the corresponding $\underline{\hat{V}}^*$ can be found by deleting rows m_1, $[m_2 + d], \ldots, [m_r + (r-1)d]$, and rows $[(h-1)d + 1]$ through dh of matrix A defined in (3-19). Row $[m_h + (h-1)d]$ will be included in the rows $[(d-1)h + 1]$ through dh which are deleted.

An illustration of the formation of A^* follows

$$\underset{9\times 1}{\underline{\hat{V}}} = \begin{bmatrix} \hat{M}_{11} - \hat{\overline{M}}_{.1} \\ \hat{M}_{21} - \hat{\overline{M}}_{.1} \\ \hat{M}_{31} - \hat{\overline{M}}_{.1} \\ \hat{M}_{12} - \hat{\overline{M}}_{.2} \\ \hat{M}_{22} - \hat{\overline{M}}_{.2} \\ \hat{M}_{32} - \hat{\overline{M}}_{.2} \\ \hat{M}_{13} - \hat{\overline{M}}_{.3} \\ \hat{M}_{23} - \hat{\overline{M}}_{.3} \\ \hat{M}_{33} - \hat{\overline{M}}_{.3} \end{bmatrix} \qquad A = \frac{1}{3}\begin{bmatrix} 2 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 2 \end{bmatrix}$$

Consider a vector $\underline{\hat{V}}^*$ of the form

$$\underline{\hat{V}}^* = \begin{bmatrix} \hat{M}_{21} - \hat{\overline{M}}_{.1} \\ \hat{M}_{31} - \hat{\overline{M}}_{.1} \\ \hat{M}_{12} - \hat{\overline{M}}_{.2} \\ \hat{M}_{22} - \hat{\overline{M}}_{.2} \end{bmatrix} \qquad h = 3, \; m_1 = 1, \; m_2 = 3 \; .$$

To find the appropriate A^* matrix using the value of $h = 3$, rows 7, 8, and 9 are deleted from A. In addition for $m_1 = 1$ and $m_2 = 3$, rows 1 and 4 are also deleted from A. The resulting A^* matrix has the form

$$A^* = \frac{1}{3} \begin{bmatrix} -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \end{bmatrix} .$$

It now remains to determine the structure of $\hat{\Sigma}_M$. For notational simplicity let $Cov(\cdot)$ and $Var(\cdot)$ denote population parameter values and let $C(\cdot)$ and $V(\cdot)$ denote consistent estimators of the former found by replacing the $P_{j_1,\ldots,j_d}$ with $\hat{P}_{j_1,\ldots,j_d}$. The formulas which follow are a consequence of the fact that a multinomial distribution characterizes the cell frequencies.

$$Var(\hat{P}_{j_1,\ldots,j_d}) = \frac{1}{n}(P_{j_1,\ldots,j_d})(1 - P_{j_1,\ldots,j_d}) \qquad (3\text{-}32)$$

$$V(\hat{P}_{j_1,\ldots,j_d}) = \frac{1}{n}(\hat{P}_{j_1,\ldots,j_d})(1 - \hat{P}_{j_1,\ldots,j_d}) \qquad (3\text{-}33)$$

$$Cov(\hat{P}_{j_1,\ldots,j_d}, \hat{P}_{j_1',\ldots,j_d'}) = -\frac{1}{n}(P_{j_1,\ldots,j_d})(P_{j_1',\ldots,j_d'}) \qquad (3\text{-}34)$$

where for at least one value of $i = 1,\ldots,d$, $j_i \neq j_i'$

$$C(\hat{P}_{j_1,\ldots,j_d}, \hat{P}_{j_1',\ldots,j_d'}) = -\frac{1}{n}(\hat{P}_{j_1,\ldots,j_d})(\hat{P}_{j_1',\ldots,j_d'}) . \qquad (3\text{-}35)$$

The following estimates of $Var(\cdot)$ and $Cov(\cdot)$ are obtained from the respective parameters in a similar fashion

$$V(\hat{M}_{qb}) = \frac{1}{n} \hat{M}_{qb}(1 - \hat{M}_{qb}) \tag{3-36}$$

$$C(\hat{M}_{qb}, \hat{M}_{qb'}) = -\frac{1}{n} \hat{M}_{qb}\hat{M}_{qb'} \quad \text{for} \quad b \neq b' \tag{3-37}$$

$$C(\hat{M}_{qb}, \hat{M}_{q'k}) = \frac{1}{n} \sum_{\substack{j_q=b \\ j_{q'}=k \\ j_i=1,\ldots,r \\ \text{for } i\neq q,q'}} (\hat{P}_{j_1,\ldots,j_d}) - \frac{1}{n} \hat{M}_{qb}\hat{M}_{q'k} \quad \text{for} \quad q \neq q' \tag{3-38}$$

Formula (3-38) results from writing the marginal proportions $\hat{M}_{qb}$ and $\hat{M}_{q'k}$ in terms of formula (3-3) and using the linearity of the covariance operator to obtain the algebraic simplification in (3-38).

Define the following:

$$\underset{d\times d}{\hat{\Sigma}}_{bb} = \hat{\Sigma}_{\underline{\hat{M}}_b} \quad \text{where} \quad \underline{\hat{M}}_b \text{ is defined in (3-17)} \quad b = 1,\ldots,r$$

$$\hat{\Sigma}_{\underline{\hat{M}}_b} = ((S_{km})) \quad S_{km} = C(\hat{M}_{kb}, \hat{M}_{mb}) \quad k \neq m \tag{3-39}$$
$$k,m=1,\ldots,d \quad S_{kk} = V(\hat{M}_{kb})$$

$$\underset{d\times d}{\hat{\Sigma}}_{ij} = C(\underline{\hat{M}}_i, \underline{\hat{M}}_j) \quad \text{for} \quad i \neq j$$

$$C(\underline{\hat{M}}_i, \underline{\hat{M}}_j) = ((a_{gh})) \quad a_{gh} = C(\hat{M}_{gi}, \hat{M}_{hj}) \tag{3-40}$$
$$g,h=1,\ldots,d$$

The variance-covariance matrix $\hat{\Sigma}_{\underline{M}}$ which is of dimension $dr \times dr$ can be expressed as a block matrix having the form

$$\hat{\Sigma}_{\hat{\underline{M}}} = \begin{bmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} & \cdots & \cdots & \hat{\Sigma}_{1r} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} & & & \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \vdots & & & & \\ \hat{\Sigma}_{r1} & \cdots & \cdots & \ddots & \hat{\Sigma}_{rr} \end{bmatrix} \qquad \cdot \qquad (3\text{-}41)$$

The form given in (3-41) is a consequence of the definition for

$\hat{\underline{M}}$ given in (3-17). The computational form of the χ^2 statistic

is then

$$\chi^2 = \underline{\hat{M}}' A^{*'} (A^* \underset{\hat{\underline{M}}}{\hat{\Sigma}} A^{*'})^{-1} A^* \ \underline{\hat{M}} \ . \qquad (3\text{-}42)$$

Conceptually the χ^2 statistic follows the general frame-

work set forth in (3-1). The null hypothesis in (3-7) can equi-

valently be stated as

$$H_0: \quad \underset{(d-1)(r-1) \times 1}{\underline{V}^*} = \underline{0} \ . \qquad (3\text{-}43)$$

The null hypothesis in (3-43) defines $(d-1)(r-1)$ linearly in-

dependent constraints on the parameters $\{P_{j_1, \ldots, j_d}\}$ each of which

is independent of the constraint that the cell probabilities sum to

one. The χ^2 statistic has the same form as the more general

statistic given in (3-1) with $\hat{\underline{V}}^*$ replacing $\underline{\hat{f}}$ and $\hat{\Sigma}_{\hat{\underline{V}}^*}$ replacing

G in (3-1). From the results of Chapter II the χ^2 statistic is

thus algebraically equivalent to both the X^2_{GSK} statistic and the

χ^2_1 statistic used in conjunction with modified minimum chi-square

estimators. The null hypothesis of marginal homogeneity is rejected

at the α level of significance whenever χ^2 is greater than the

$(1-\alpha)$ quantile of chi-square with $(d-1)(r-1)$ degrees of freedom.

Asymptotic Power of χ^2

Bhapkar (1966) conjectured that the results of Mitra (1958) concerning the power of the minimum chi-square statistic based on maximum likelihood estimators could be extended to the χ_1^2 statistic based on modified minimum chi-square estimators. Because χ^2 is algebraically equal to a χ_1^2 statistic based on modified minimum chi-square estimators, Mitra's results will be applied to the χ^2 statistic.

Let a sequence of alternatives to the null hypothesis be defined as

$$H_A^{(n)}: \quad \underline{v}^* = \frac{1}{\sqrt{n}} \underline{\gamma} \tag{3-44}$$

where $\underline{\gamma}$ is a vector of $(d-1)(r-1)$ constants not all of which are zero. As n tends to infinity the sequence of alternatives $\{H_A^{(n)}\}$ tends to the null hypothesis. For large n, the power function of the χ^2 test can be approximated by a noncentral chi-square with noncentrality parameter given by

$$\lambda = \underline{v}^{*'} \Sigma_{\hat{\underline{v}}^*}^{-1} \underline{v}^* = \frac{1}{\sqrt{n}} \underline{\gamma}' \Sigma_{\hat{\underline{v}}^*}^{-1} \frac{1}{\sqrt{n}} \underline{\gamma} \quad . \tag{3-45}$$

Because $\Sigma_{\hat{\underline{v}}^*}^{-1}$ has an n to the first power in it the expression given in (3-45) does not depend upon n.

TECHNIQUES FOR ISOLATING
SOURCES OF SIGNIFICANCE

If the χ^2 statistic is so large as to lead to rejection of the null hypothesis of marginal homogeneity, the next step in the analysis is to identify the sources of significance and estimate the magnitude of the differences. If there are only two correlated distributions it suffices to identify those categories, or combinations of categories, with unequal marginal probabilities. When more than two correlated distributions are involved, the additional problem of locating which of the distributions differ is encountered.

Define a contrast to be a function of the M_{qb} of the form

$$\Psi = \sum_{b=1}^{r} \sum_{q=1}^{d} C_{qb} M_{qb} \qquad (3\text{-}46)$$

where the C_{qb} are known constants subject to the condition that

$$\sum_{q=1}^{d} C_{qb} = 0 \qquad \forall\, b \, . \qquad (3\text{-}47)$$

Let

$$\underset{1 \times d}{\underline{C}'_b} = [C_{1b}, C_{2b}, \ldots, C_{db}] \qquad \forall\, b = 1, \ldots, r$$

and let   (3-48)

$$\underset{1 \times dr}{\underline{C}'} = [\underline{C}'_1, \underline{C}'_2, \ldots, \underline{C}'_r] \quad .$$

Define

$$\underset{1 \times dr}{\underline{M}'} = [\underline{M}'_1, \underline{M}'_2, \ldots, \underline{M}'_r] \qquad (3\text{-}49)$$

where $\underline{M}'_b = [M_{1b}, M_{2b}, \ldots, M_{db}] \quad \forall\, b = 1, \ldots, r.$

Any contrast, Ψ, can then be written in vector notation as

$$\Psi = \underline{C}' \ \underline{M} \ . \tag{3-50}$$

For suitably chosen $\underline{C}$ vectors, contrasts which provide meaningful information to the experimenter can be formed. Interval estimates of these contrasts can be made which not only help isolate where the differences between the distributions lie, but also provide estimates of the magnitude of these differences. Some examples of contrasts are given in the following illustration from Yoshinaga's (1974) study.

Yoshinaga investigated the constancy of teachers' choices of a disciplinary strategy when increasing increments of information were given about a child involved in a disciplinary incident. The teachers were given a description of the incident and asked to choose a strategy from a group of three strategies to deal with the behavior problem. The three strategies were: (1) positive reinforcement, (2) social modeling, and (3) punishment. The teachers were given some biographical information about the child and were again asked to choose a strategy. Finally the teachers were presented with further background information about the child and asked again to choose a strategy from the same group of those three strategies. The χ^2 procedure provides a test of whether the teachers as a group vary their strategy as increased amounts of information are provided about the child. A significant value of the χ^2 statistic provides evidence that the probability of a teacher choosing a given strategy may change as this teacher learns more about the child. In order to identify which of the probabilities change and under what circumstances,

the researcher can examine a series of contrasts. The contrasts are typically chosen to reflect the logical comparisons which might be of interest in the given experiment. For this experiment the marginal proportions might be compared according to the contrasting sets of conditions specified below.

(a) strategy based upon no information about the child vs. strategy based upon one source of background information

(b) strategy based upon no information about the child vs. strategy based upon two sources of background information

(c) strategy based upon one source of background information vs. strategy based upon two sources of information

(d) strategy based upon no information about the child vs. strategy based upon some information.

A set of contrasts which reflects comparisons of these types and the corresponding $\underline{C}$ vectors are listed below.

$\Psi_1 = M_{11} - M_{21}$ $\qquad$ $_1\underline{C}' = [1,-1,0,0,0,0,0,0,0]$

$\Psi_2 = M_{12} - M_{22}$ $\qquad$ $_2\underline{C}' = [0,0,0,1,-1,0,0,0,0]$

$\Psi_3 = M_{13} - M_{23}$ $\qquad$ $_3\underline{C}' = [0,0,0,0,0,0,1,-1,0]$

$\Psi_4 = M_{21} - M_{31}$ $\qquad$ $_4\underline{C}' = [0,1,-1,0,0,0,0,0,0]$

$\Psi_5 = M_{22} - M_{32}$ $\qquad$ $_5\underline{C}' = [0,0,0,0,1,-1,0,0,0]$

$\Psi_6 = M_{23} - M_{33}$ $\qquad$ $_6\underline{C}' = [0,0,0,0,0,0,0,1,-1]$

$\Psi_7 = M_{13} - M_{33}$ $\qquad$ $_7\underline{C}' = [0,0,0,0,0,0,1,0,-1]$

$\Psi_8 = M_{12} - M_{32}$ $\qquad$ $_8\underline{C}' = [0,0,0,1,0,-1,0,0,0]$

$\Psi_9 = M_{11} - M_{31}$ $\qquad$ $_9\underline{C}' = [1,0,-1,0,0,0,0,0,0]$

$\Psi_{10} = M_{11} - \frac{1}{2}M_{21} - \frac{1}{2}M_{31}$ $\qquad$ $_{10}\underline{C}' = [1,-\frac{1}{2},-\frac{1}{2},0,0,0,0,0,0]$

$\Psi_{11} = M_{12} - \frac{1}{2}M_{22} - \frac{1}{2}M_{32}$ $\qquad$ $_{11}\underline{C}' = [0,0,0,1,-\frac{1}{2},-\frac{1}{2},0,0,0]$

$\Psi_{12} = M_{13} - \frac{1}{2}M_{23} - \frac{1}{2}M_{33}$ $\qquad$ $_{12}\underline{C}' = [0,0,0,0,0,0,1,-\frac{1}{2},-\frac{1}{2}]$

Contrasts Ψ_1 through Ψ_3 are based upon comparison type (a), while contrasts Ψ_4 through Ψ_6 are based upon comparison type (c). Contrasts Ψ_7 through Ψ_9 reflect comparison type (b) while contrasts Ψ_{10} through Ψ_{12} reflect comparison type (d). Contrasts which contain only two marginal proportions, such as those given by Ψ_1 through Ψ_9 are called simple or pairwise contrasts. Contrasts containing more than two marginal proportions, as in Ψ_{10} through Ψ_{12}, are called complex contrasts.

Hypotheses of the form

$$_iH_0: \quad \Psi_i = 0 \qquad\qquad (3\text{-}51)$$

can be tested by determining whether the corresponding confidence interval for Ψ_i includes the value of zero. Should $_iH_0$ be rejected, meaning that the confidence interval for Ψ_i does not contain zero, the confidence limits provide an interval estimate of the magnitude of Ψ_i. A confidence interval procedure not only locates differences, it provides interval estimates of the magnitude of these differences. For example a confidence interval for Ψ_1 not only tests whether the proportion of teachers who choose the strategy of positive reinforcement changes with the first introduction of background information about the child, but also provides an interval estimate of the magnitude of the change, should one exist. A confidence interval for Ψ_{10} tests whether some versus no background information about the child has any effect on the probability that a teacher chooses a positive reinforcement strategy, and provides an interval estimate of the effect, should one exist.

When several hypotheses are tested at one time, the prob-
lem of determining the significance level for the experiment as
a whole becomes very complicated. If the probability of committing
a type I error is set equal to α for each individual confidence
interval, the probability of rejecting at least one true null hypo-
thesis, $_1H_0$: $\Psi_i = 0$, when several confidence intervals are gen-
erated becomes considerably larger than α. If enough intervals are
examined, it is very likely that one or more hypotheses are re-
jected even though they are all true. The error rate experiment-
wise refers to the probability that one or more erroneous statements
will be made in an experiment, where for the purposes of this dis-
cussion an erroneous statement is the rejection of a true null
hypothesis. For most data analytic situations the experimentwise
error rate needs to be controlled since the researcher typically
wants to guard against making any erroneous statements.

In the discussion which follows two methods are considered
for forming confidence intervals about the Ψ_i. Each technique
controls the magnitude of the experimentwise error rate. The
theoretical basis for each of the two techniques is now presented.

Scheffé-type Solution

The technique presented in this subsection is a simultaneous
confidence interval method based upon the works of Scheffé (1959)
and Goodman (1964) and adapted to the design and statistic under
consideration in this chapter. Let the set of all possible con-
trasts of the form Ψ be denoted by $\mathcal{F}$. The unrestricted maximum
likelihood estimator of Ψ for any $\Psi \in \mathcal{F}$ is denoted as $\hat{\Psi}$ where

$$\hat{\Psi} = \underline{C}' \ \hat{\underline{M}} \ .$$

<div align="right">(3-52)</div>

$\hat{\Psi}$ is also an unbiased estimator of Ψ for each $\Psi \in \mathcal{F}$. Let $S^2(\hat{\Psi})$ denote the consistent estimator of the variance of $\hat{\Psi}$ defined in (3-53)

$$S^2(\hat{\Psi}) = \underline{C}' \ \hat{\Sigma}_{\hat{\underline{M}}} \ \underline{C} \ .$$

<div align="right">(3-53)</div>

The theorem which follows provides the theoretical basis for the simultaneous procedure described in this subsection.

Theorem 3.2. As n tends to infinity, the probability will approach $1 - \alpha$ that simultaneously for all $\Psi \in \mathcal{F}$

$$\hat{\Psi} - S(\hat{\Psi})L \le \Psi \le \hat{\Psi} + S(\hat{\Psi})L \ .$$

<div align="right">(3-54)</div>

Here n denotes the sample size, L is the positive square root of the $100(1 - \alpha)$th percentile of the central chi-square distribution with $(d-1)(r-1)$ degrees of freedom and $S(\hat{\Psi})$ is the positive square root of $S^2(\hat{\Psi})$.

The result of Theorem 3.2 allows the experimenter to examine the confidence intervals for as many contrasts, $\Psi \in \mathcal{F}$, as desired, holding the probability of making one or more erroneous statements at α, regardless of the number of confidence intervals examined. Rather than assigning an error rate per contrast, an error rate for the entire family of contrasts, $\mathcal{F}$, is assigned. The technique is quite advantageous to use when many different contrasts are of interest, or when one is merely searching the data to locate some of the sources of significance in the event of rejection of the over-all null hypothesis of marginal homogeneity. The use of such

a technique permits the experimenter to control the experimentwise error rate and thereby guard against making one or more erroneous statements, irrespective of the number of statements made.

Proof of Theorem 3.2.

Define $\Delta_{qb} = M_{qb} - \overline{M}_{.b}$
$\quad q = 2,\ldots,d$
$\quad b = 2,\ldots,r$
$\qquad\qquad$ (3-55)

The $(d-1)(r-1)$ functions Δ_{qb}, $(q = 2,\ldots,d)$, $(b = 2,\ldots,r)$, are contrasts. The $\{\Delta_{qb}\}$ form a set of $(d-1)(r-1)$ linearly independent estimable functions of the components of the $\underline{M}$ vector and will span the space of all contrasts $\Psi \in \mathscr{F}$.

Define

$$\underset{1\times(d-1)(r-1)}{\underline{\Delta}'} = [\underline{\Delta}_2',\ldots,\underline{\Delta}_r']$$

where $\qquad\qquad$ (3-56)

$$\underset{1\times(d-1)}{\underline{\Delta}_b'} = [\Delta_{2b},\ldots,\Delta_{db}] \ .$$

Any contrast $\Psi \in \mathscr{F}$ can then be written in the form

$$\Psi = \underline{h}' \ \underline{\Delta} \qquad\qquad (3\text{-}57)$$

where $\underline{h}$ is a vector of suitably chosen constants.

Let

$$\underset{1\times(d-1)(r-1)}{\underline{D}'} = [\underline{D}_2',\ldots,\underline{D}_r'] \qquad\qquad (3\text{-}58)$$

where

$$\underline{D}_b' = [\hat{M}_{2b} - \overline{\hat{M}}_{.b},\ldots, \hat{M}_{db} - \overline{\hat{M}}_{.b}]$$

$$\qquad\qquad (3\text{-}59)$$

and $\qquad\qquad \hat{\Psi} = \underline{h}' \ \underline{D} \cdot$

It was shown earlier in the chapter that for large n, $\underline{D}$ is approximately multivariate normal with an expected value of $\underline{\Delta}$ and with a nonsingular variance-covariance matrix $\Sigma_{\underline{D}}$. The $\underline{D}$ vector is a special case of the more general $\hat{\underline{V}}^*$ vector defined earlier in the chapter. If S denotes a consistent estimator of $\Sigma_{\underline{D}}$, where S is $A^*\hat{\Sigma}_{\hat{M}}A^{*'}$ with $\underline{D} = A^*\hat{\underline{M}}$ then for large n, S is nonsingular and as n tends to infinity, the probability will approach $1 - \alpha$ that

$$(\underline{D} - \underline{\Delta})'S^{-1}(\underline{D} - \underline{\Delta}) \leq L^2 \tag{3-60}$$

where L^2 is the $100(1 - \alpha)$th percentile of the chi-square distribution with $(d-1)(r-1)$ degrees of freedom. The statement in (3-60) is the result of Proposition 2 of this chapter.

The Cauchy-Schwarz inequality states that if $\underline{U}$ and $\underline{N}$ are any two vectors in the space R^k then

$$|\underline{N}'\underline{U}| \leq \sqrt{\underline{N}'\underline{N}} \ \sqrt{\underline{U}'\underline{U}} \ . \tag{3-61}$$

The inequality given in (3-61) is used to prove a lemma which aids in the proof of Theorem 3.2.

Lemma 3.1. If J is a $k \times k$ symmetric matrix such that $J = R'R$ where R is a $k \times k$ matrix and of full rank, then for all $\underline{X}, \underline{Y} \in R^k$

$$|\underline{X}'\underline{Y}| \leq \sqrt{\underline{X}'J\underline{X}} \ \sqrt{\underline{Y}'J^{-1}\underline{Y}} \ . \tag{3-62}$$

Proof of Lemma 3.1. If $\underline{U}$ is taken to be $(R^{-1})'\underline{Y}$ and $\underline{N}$ is taken to be $R\underline{X}$ then it follows that

$$\underline{N}' \ \underline{U} = \underline{X}' \ R'(R^{-1})'\underline{Y} = \underline{X}' \ \underline{Y} \ .$$

By definition of matrix J

$$\underline{N}' \ \underline{N} = \underline{X}' \ R'R \ \underline{X} = \underline{X}'J \ \underline{X}$$

and

$$\underline{U}'\underline{U} = \underline{Y}'R^{-1}(R^{-1})'\underline{Y} = \underline{Y}'J^{-1}\underline{Y} \ .$$

If the preceding equalities are used together with the Cauchy-Schwarz inequality stated in (3-61) the desired result

$$\left|\underline{X}'\underline{Y}\right| \leq \sqrt{\underline{X}'J \ \underline{X}} \ \sqrt{\underline{Y}'J^{-1}\underline{Y}}$$

is obtained and Lemma 3.1 is proven.

A second lemma needed in the proof of Theorem 3.2 is now given.

Lemma 3.2. $(\underline{D} - \underline{\Delta})'S^{-1}(\underline{D} - \underline{\Delta}) \leq L^2$ if and only if $\left|\underline{h}'(\underline{D} - \underline{\Delta})\right| \leq L \sqrt{\underline{h}'S \ \underline{h}}$ for all $\underline{h}$ in $(d-1)(r-1)$ Euclidean space.

Proof of Lemma 3.2. The variance-covariance matrix S is of full rank for n large, thus it possesses the properties of matrix J in Lemma 3.1. If $\underline{h}$ is substituted for $\underline{X}$, $(\underline{D} - \underline{\Delta})$ for $\underline{Y}$, and S for J in Lemma 3.1 with $k = (d-1)(r-1)$ the result given in (3-63) follows.

$$\left|\underline{h}'(\underline{D} - \underline{\Delta})\right| \leq \sqrt{\underline{h}'S \ \underline{h}} \ \sqrt{(\underline{D} - \underline{\Delta})'S^{-1}(\underline{D} - \underline{\Delta})} \qquad (3-63)$$

$$\forall \underline{h} \text{ in } (d-1)(r-1) \text{ Euclidean space.}$$

Now if $(\underline{D} - \underline{\Delta})'S^{-1}(\underline{D} - \underline{\Delta}) \leq L^2$ then from (3-63) it follows that

$$|\underline{h}'(\underline{D} - \underline{\Delta})| \leq L \sqrt{\underline{h}'S\ \underline{h}}$$

$$\forall\ \underline{h} \text{ in } (d-1)(r-1) \text{ Euclidean space.}$$

This completes the proof of Lemma 3.2 going one way. Going in the other direction it is assumed that

$$|\underline{h}'(\underline{D} - \underline{\Delta})| \leq L \sqrt{\underline{h}'S\ \underline{h}} \quad \forall\ \underline{h} \text{ in } (d-1)(r-1) \text{ Euclidean space.}$$

Choosing an $\underline{h}$ equal to $(S^{-1})'(\underline{D} - \underline{\Delta})$ will result in the inequality

$$|(\underline{D} - \underline{\Delta})'S^{-1}(\underline{D} - \underline{\Delta})| \leq L \sqrt{(\underline{D} - \underline{\Delta})'S^{-1}S(S^{-1})'(\underline{D} - \underline{\Delta})}\ .$$

Because both sides of this inequality are non negative it is valid to square both sides to obtain an equivalent statement. Squaring both sides of the inequality and using the fact that symmetry of S implies that S^{-1} is also symmetric when the inverse exists, the following result is obtained:

$$[(\underline{D} - \underline{\Delta})'S^{-1}(\underline{D} - \underline{\Delta})]^2 \leq L^2(\underline{D} - \underline{\Delta})'S^{-1}(\underline{D} - \underline{\Delta})\ . \tag{3-64}$$

The expression $(\underline{D} - \underline{\Delta})'S^{-1}(\underline{D} - \underline{\Delta})$ is non negative since it is a quadratic form. If the quadratic form happens to be zero, the result that $(\underline{D} - \underline{\Delta})'S^{-1}(\underline{D} - \underline{\Delta}) \leq L^2$ is obvious since L^2 is positive. Otherwise $[(\underline{D} - \underline{\Delta})'S^{-1}(\underline{D} - \underline{\Delta})] > 0$ and dividing both sides of (3-64) by this expression yields

$$(\underline{D} - \underline{\Delta})'S^{-1}(\underline{D} - \underline{\Delta}) \leq L^2\ . \tag{3-65}$$

This completes the proof of Lemma 3.2.

As a consequence of the results in (3-65) and Lemma 3.2 the probability that

$$\underline{h}'\underline{D} - \sqrt{\underline{h}'S\ \underline{h}}\ \ L \le \underline{h}'\underline{\Delta} \le \underline{h}'\underline{D} + \sqrt{\underline{h}'S\ \underline{h}}\ \ L \qquad (3\text{-}66)$$

simultaneously for all $\underline{h}$ in $(d-1)(r-1)$ Euclidean space tends to $1 - \alpha$ as n tends to infinity. But by (3-57), each contrast $\Psi \in \mathcal{F}$, is of the form $\underline{h}'\underline{\Delta}$ and conversely every $\underline{h}'\underline{\Delta}$ is a contrast Ψ; $\underline{h}'\underline{D}$ is then $\hat{\Psi}$ while $\underline{h}'S\ \underline{h}$ is the sample variance of $\hat{\Psi}$ which is denoted as $S^2(\hat{\Psi})$ as defined in (3-53). The result given in (3-66) together with these substitutions completes the proof of Theorem 3.2.

The simultaneous confidence interval procedure is typically used as a follow up procedure when the value of the χ^2 statistic is large enough to permit rejection of H_0, the over-all null hypothesis of marginal homogeneity.

The experimenter is free to generate as many confidence intervals as desired, provided the intervals have endpoints as specified by (3-54). Hypotheses of the form $_iH_0: \Psi_i = 0$ can be tested by seeing if the corresponding confidence interval $(\hat{\Psi}_i - S(\hat{\Psi})L,\ \hat{\Psi}_i + S(\hat{\Psi})L)$ spans zero. If the interval does not span zero, a source of significance is located and an interval estimate of its magnitude is given by the bounds of the interval. In this manner sources of significance can be located and their magnitudes estimated, while at the same time the experimenter is protected against making one or more erroneous statements because a simultaneous error rate of α is set for the entire family of contrasts $\mathcal{F}$.

The simultaneous confidence interval procedure just described has the added advantage of correspondence to the test criterion associated with the χ^2 procedure. The χ^2 procedure tests the over-all null hypothesis of marginal homogenity,

$$H_0: \quad \underline{\Delta} = \underline{0} , \tag{3-67}$$

and the χ^2 statistic is algebraically equal to the quantity $\underline{D}'S^{-1}\underline{D}$. For a test performed at the α level, the test criterion is to reject H_0 if and only if

$$\underline{D}'S^{-1}\underline{D} > L^2 . \tag{3-68}$$

But from Lemma 3.2 the condition in (3-68) occurs if and only if there exists at least one $\underline{h}$ in $(d-1)(r-1)$ Euclidean space for which

$$|\underline{h}'\underline{D}| > \sqrt{\underline{h}'S^{-1}\underline{h}} \ L .$$

The over-all null hypothesis, H_0, is thus rejected using the χ^2 test criterion if and only if there is at least one contrast $\Psi \in \mathcal{F}$ for which zero is not included in the interval

$$(\hat{\Psi} - S(\hat{\Psi})L, \ \hat{\Psi} + S(\hat{\Psi})L) .$$

Rejection of the over-all null hypothesis of marginal homogeneity using the χ^2 test criterion guarantees the existence of at least one contrast $\Psi \in \mathcal{F}$ which is statistically significant when using the simultaneous confidence procedure.

The ability of the experimenter to examine as many contrasts $\Psi \in \mathcal{F}$ as he desires for a fixed experimentwise error rate is not

without drawbacks. Because the Scheffé-type procedure focuses on the entire family $\mathcal{F}$, it may not be as powerful a procedure as certain other techniques. The confidence intervals generated using the Scheffé-type solution may be wider than the corresponding intervals using other procedures and less sensitive in detecting differences. An alternative procedure to the Scheffé type technique follows.

Bonferroni-type Solution

If only a specified subset, S, of the contrasts in $\mathcal{F}$ are of interest to the experimenter and if the number of contrasts in S is sufficiently small, then the technique presented in this subsection may be preferable to the Scheffé-type solution. The Bonferroni-type solution places an error rate on each individual contrast which is examined but the error rate per contrast is determined so as to control the over-all experimentwise error rate.

Let S be a subset of $\mathcal{F}$, with S containing the finite number of contrasts which are of interest to the experimenter, denoted as $\Psi_1, \Psi_2, \ldots, \Psi_k$. Let E_i be the event that the null hypothesis

$$_iH_0: \quad \Psi_i = 0 \qquad\qquad (3-54)$$

is falsely rejected. Let

$$\alpha_i = P(E_i) \ .$$

Let E^* be the event that one or more null hypotheses of the form given in (3-54) are falsely rejected. Then

$$P(E^*) = P(E_1 \cup E_2 \cup \ldots \cup E_k) \le \sum_{i=1}^{k} P(E_i) = \sum_{i=1}^{k} \alpha_i . \qquad (3-69)$$

The inequality in (3-69) is called the Bonferroni inequality. The quantity $P(E^*)$ is the experimentwise error rate, having as an upper bound the quantity $\sum_{i=1}^{k} \alpha_i$. If k is not too large (for example 5) and the α_i are relatively small (for example .01) the approximation of this upper bound to the quantity $P(E^*)$ is quite good. Typically each α_i is chosen to be α/k where α is a specified value which serves as an upper bound for the experimentwise error rate. By dividing the entire α value among the different contrasts of interest, the experimenter is able to control the probability of making one or more erroneous statements.

Earlier it was shown that for large n, the vector of cell frequencies is approximately multivariate normal. Because any contrast, $\hat{\Psi}_i$, is a linear combination of the components of the cell frequency vector, when n is large $\hat{\Psi}_i$ will be approximately univariate normal. Because $S(\hat{\Psi}_i) \to \sigma(\hat{\Psi}_i)$ as $n \to \infty$, for n large the distribution of

$$\frac{\hat{\Psi}_i - \Psi_i}{S(\hat{\Psi}_i)} \qquad (3-70)$$

is approximately standard normal. An equivalent statement of (3-70) is that as $n \to \infty$ the probability will approach $1 - \alpha_i$ that

$$\hat{\Psi}_i - S(\hat{\Psi}_i)Z_{\frac{\alpha_i}{2}} \le \Psi_i \le \hat{\Psi}_i + S(\hat{\Psi}_i) Z_{\frac{\alpha_i}{2}} \qquad (3-71)$$

where $Z_{\alpha_i/2}$ is the $100(1 - \alpha_i/2)$th percentile of the standard normal distribution. To test hypotheses of the form $_iH_0$: $\Psi_i = 0$,

determine whether the corresponding confidence interval for Ψ_i,

$$(\hat{\Psi}_i - S(\hat{\Psi}_i)Z_{\alpha_i/2}, \quad \hat{\Psi}_i + S(\hat{\Psi}_i)Z_{\alpha_i/2})$$

spans zero.

If all $Z_{\alpha_i/2}$ are set equal to Z where Z is the $100(1 - \frac{\alpha}{2k})$th percentile of the standard normal distribution, then the confidence intervals using the Bonferroni technique (3-71) will be narrower than the corresponding intervals using the Scheffé-like procedure, (3-54), if and only if

$$Z < L$$

where L is the $100(1 - \alpha)$th percentile of the chi-square distribution with $(d-1)(r-1)$ degrees of freedom. Typically $Z \leq L$ will hold whenever the number of contrasts examined, k, is such that

$$k \leq \tfrac{1}{2}d(d - 1)r \qquad\qquad (3-72)$$

for the usual values of α (.05 or .01), Goodman (1964).

The Bonferroni technique can be used as a substitute for the omnibus test of homogeneity or as a follow up procedure to locate and estimate sources of significance should the omnibus test be significant. In either case, the choice of the contrasts to comprise set S must be made before the data are examined, otherwise a serious distortion of the error rate may occur. The experimenter must not use the data in determining which contrasts to include in set S. This is a major point of difference between the Bonferroni and Scheffé-type procedures. The Scheffé-type procedure allows the experimenter to examine any contrasts in $\mathcal{F}$ for a constant

experimentwise error rate, whether the choice of contrasts is dictated by prior hypotheses or inspection of the data.

Although both the Scheffé and Bonferroni type techniques are large sample procedures based upon asymptotic results, the robustness to nonnormality of the Ψ_i is probably greater for the Scheffé-type than for the Bonferroni-type technique. The accuracy of the Scheffé-type procedure is based upon the degree to which the quadratic form $\underline{D}'S^{-1}\underline{D}$ approaches its limiting chi-square distribution under the null hypothesis. Such a study is conducted in Chapter IV for tests performed at the nominal .01, .05, and .10 significance levels. A good fit to the limiting distribution at these cutoff values will guarantee a valid Scheffé-type confidence procedure at these levels. The validity of the Bonferroni procedure is dependent upon the degree to which the quantity

$$\frac{\hat{\Psi}_i - \Psi_i}{S(\hat{\Psi}_i)} \tag{3-70}$$

approaches its standard normal limiting distribution for each contrast Ψ_i examined. The fit of the statistic in (3-70) to its limiting standard normal distribution may not be good in the extreme tails of the normal distribution. Because a separate significance level is assigned to each test of $_iH_0$: $\Psi_i = 0$, the significance level per test is typically quite small, for example $\frac{\alpha}{k}$. If $\alpha = .05$ and $k = 10$, it would be necessary to use the .25th and the 99.75th percentiles of the standard normal distribution as the cutoff values. The fit of the statistic in (3-70) to its limiting distribution in such extremities of the distribution as the .25th and 99.75th

percentiles is questionable for small or even moderate sample sizes. A Monte Carlo investigation to examine such behavior would be extremely costly because of the large number of samples which would have to be generated to produce stable estimates of the actual significance level for tests performed at such small nominal α values. The actual significance level for each test of $\Psi_i = 0$ or equivalently the actual confidence level given to each interval in (3-71) may be quite different from its stated nominal level when n, the sample size, is small or moderate. When similar procedures are employed for the k contrasts, the resulting experimentwise error rate may be quite different from the theoretical rate of α, the value set by the experimenter.

Asymptotically the relationship between the Bonferroni and Scheffé-type techniques is made explicit in (3-72), but for small or moderate sample sizes the relationship in (3-72) may not be a valid criterion to apply because of the lack of information about the actual significance level of tests using the extreme tails of the distribution as would be the case in the Bonferroni-type procedure when many hypotheses are tested.

A data example which illustrates the χ^2 procedure and the two post hoc techniques presented in this section is now given.

DATA EXAMPLE

The following data example is based upon Laumann's (1973)

social interaction data. In Laumann's study of social mobility

and social interaction, a husband and the fathers of both the husband

and wife were each classified into one of a number of social classes.

This design can be considered as consisting of three matched samples.

By testing for homogeneity of the three matched samples on the

variable of classification, the researcher can make inferences con-

cerning the social interaction of various social classes. Because

the three samples are probably not independent, the χ^2 procedure

is used to test for homogeneity of the marginal distributions.

Laumann's original categories of classification are collapsed into

just four categories to make the analysis feasible since r^d, the

number of cells in the contingency table, would be too large had

Laumann's original categories been used.

The four categories of classification are given as:

(1) professional, technical, kindred-white collar managers,

     officials

(2) clerical - white collar

     sales - white collar

(3) craftsman - blue collar

(4) operatives - blue collar

     service - blue collar

     laborers (except farm) - blue collar

Let husbands be sample 1, husbands' fathers sample 2, and wives'

fathers sample 3. The data are represented in the $4 \times 4 \times 4$ con-

tingency table given in Table 3-1.

TABLE 3-1

Laumann's Social Interaction Data

Sample 1

(1)

Sample 2

		(1)	(2)	(3)	(4)
	(1)	44	17	4	12
Sample 3	(2)	10	3	6	2
	(3)	29	7	22	22
	(4)	13	8	21	32

(2)

Sample 2

	(1)	(2)	(3)	(4)
(1)	11	2	4	8
(2)	1	2	2	3
(3)	6	7	4	9
(4)	5	1	11	8

(3)

Sample 2

		(1)	(2)	(3)	(4)
	(1)	8	2	19	11
Sample 3	(2)	0	0	5	5
	(3)	11	2	26	35
	(4)	4	1	21	37

(4)

Sample 2

	(1)	(2)	(3)	(4)
(1)	9	2	9	10
(2)	0	0	2	6
(3)	11	0	22	32
(4)	12	4	28	39

Marginal Totals

	(1)	(2)	(3)	(4)
Sample 1	252	84	187	186
Sample 2	174	58	206	271
Sample 3	172	47	245	245

n = 709

H_0: $M_{1b} = M_{2b} = M_{3b}$ for $b = 1,2,3,4$

$$\hat{\underline{V}}^* = \begin{bmatrix} \hat{M}_{11} - \hat{\overline{M}}_{.1} \\ \hat{M}_{21} - \hat{\overline{M}}_{.1} \\ \hat{M}_{12} - \hat{\overline{M}}_{.2} \\ \hat{M}_{22} - \hat{\overline{M}}_{.2} \\ \hat{M}_{13} - \hat{\overline{M}}_{.3} \\ \hat{M}_{23} - \hat{\overline{M}}_{.3} \end{bmatrix} = \begin{bmatrix} .074283 \\ -.035731 \\ .029619 \\ -.007052 \\ -.036201 \\ -.009403 \end{bmatrix}$$

The $\hat{\underline{V}}^*$ vector is formed by choosing $m_1 = m_2 = m_3 = m_4 = 3$ and $h = 4$, where the m's and h are defined in (3-15). This choice represents one of many possible $\hat{\underline{V}}^*$ vectors which could be used. In the computation of the χ^2 statistic

$$\hat{\Sigma}^{-1}_{\hat{\underline{V}}^*} = \begin{bmatrix} 11404.6 & 6198.0 & 6501.1 & 2647.3 & 5468.9 & 2420.6 \\ 6198.0 & 13478.0 & 2991.9 & 6467.3 & 3445.4 & 4713.3 \\ 6501.1 & 2991.9 & 22447.7 & 12705.9 & 5326.5 & 2411.8 \\ 2647.3 & 6467.3 & 12705.9 & 25063.1 & 2961.5 & 4452.6 \\ 5468.9 & 3445.4 & 5326.5 & 2961.5 & 10192.6 & 4445.6 \\ 2420.6 & 4713.3 & 2411.8 & 4452.6 & 4445.6 & 8584.4 \end{bmatrix}$$

$$\chi^2 = \hat{\underline{V}}^{*'} \hat{\Sigma}^{-1}_{\hat{\underline{V}}^*} \hat{\underline{V}}^* = 71.39 \quad .$$

The null hypothesis is rejected for tests performed at the nominal level of $\alpha = .05$ since the value of χ^2 is larger than the 95th quantile of χ^2 with 8 degrees of freedom.

Confidence intervals using both the Scheffé-like and Bonferroni-techniques are computed for the following contrasts:

$$\Psi_1 = M_{11} - M_{21}$$

$$\Psi_2 = M_{11} - M_{31}$$

$$\Psi_3 = M_{21} - M_{31}$$

$$\Psi_4 = 2M_{11} - M_{21} - M_{31}$$

$$\Psi_5 = M_{14} - M_{24}$$

$$\Psi_6 = M_{14} - M_{34}$$

$$\Psi_7 = M_{24} - M_{34}$$

$$\Psi_8 = 2M_{14} - M_{24} - M_{34}$$

$$\Psi_9 = M_{22} - M_{32}$$

$$\Psi_{10} = M_{23} - M_{33}$$

TABLE 3-2

Scheffé-like Confidence Intervals at $\alpha = .05$ and

Bonferroni Confidence Intervals with $\alpha = .005$ per Contrast

Ψ	$\hat{\Psi}$	$S(\hat{\Psi})$	Scheffé-like $(\hat{\Psi} - S(\hat{\Psi})L, \hat{\Psi} + S(\hat{\Psi}) L)$	Bonferroni $(\hat{\Psi} - S(\hat{\Psi}) Z, \hat{\Psi} + S(\hat{\Psi})Z)$
Ψ_1	.1100	.0212	(.0349, .1852)*	(.0504, .1696)*
Ψ_2	.1128	.0228	(.0320, .1937)*	(.0487, .1769)*
Ψ_3	.0028	.0200	(-.0638, .0740)	(-.0534, .0590)
Ψ_4	.2228	.0270	(.1270, .3186)*	(.1531, .2925)*
Ψ_5	-.1199	.0233	(-.2026, -.0372)*	(-.1854, -.0544)*
Ψ_6	-.0832	.0227	(-.1639, -.0025)*	(-.1470, -.0194)*
Ψ_7	.0367	.0237	(-.0475, .1209)	(-.0299, .1033)
Ψ_8	-.2031	.0395	(-.3431, -.0631)*	(-.3141, -.0921)*
Ψ_9	.0155	.0137	(-.0332, .0642)	(-.0230, .0540)
Ψ_{10}	-.0550	.0245	(-.1418, .0318)	(-.1238, .0138)

* Significant contrast

Contrasts Ψ_3, Ψ_7, Ψ_9, and Ψ_{10} each compare the respective family backgrounds of husband and wife. Neither the Scheffé nor the Bonferroni procedure yields statistically significant confidence intervals. This suggests that the variable of social class has the same distribution for both father of husband and father of wife. In a global sense these results indicate that the social class backgrounds are the same for husband and wife. Contrasts Ψ_1 and Ψ_5 each examine generational differences in social class. Both contrasts are significant, and the nature of the intervals suggests upward class mobility of the son versus his father. Contrasts Ψ_2, Ψ_4, Ψ_6, and Ψ_8 each make generational comparisons as well as comparisons between the husband and the wife's father. Each of these contrasts is significant using both the Bonferroni and Scheffé-like procedures. The confidence intervals again reflect an upward class mobility of the husband relative to his father and/or father-in-law.

For this data example the Bonferroni solution yields shorter confidence intervals and thus greater power than the Scheffé-like procedure. The power of the Scheffé-like procedure relative to the Bonferroni solution would increase as more contrasts are examined and eventually surpass the power of the Bonferroni solution. For this data example the small number of contrasts examined relative to the total number of all possible contrasts is not enough to realize the advantage of a simultaneous procedure such as the Scheffé-like procedure.

The extent to which the asymptotic results of this chapter hold in the case of finite sample sizes is investigated in the next

chapter. A simulation study of the behavior of the χ^2 statistic for finite sample sizes is reported in Chapter IV. The degree to which the χ^2 statistic approaches its limiting distribution both under the null and various forms of the alternative hypothesis is studied in the next chapter.

CHAPTER IV

A MONTE CARLO STUDY OF THE χ^2 STATISTIC

In Chapter III it was shown that the χ^2 statistic has a limiting central chi-square distribution under the null hypothesis of marginal homogeneity. The nature of the chi-square approximation to the exact distribution of χ^2 for finite sample sizes has not as yet been investigated. The jump from asymptotic distribution theory to the practical use of an asymptotic result cannot be made unless the asymptotic result is examined in a context of practical circumstances. Before the use of χ^2 could become a legitimate statistical procedure, the author felt that an investigation of the distributional behavior of χ^2 for finite sample sizes was warranted.

The extent of the disparity between the nominal significance level at which a statistical test is performed and the exact level of significance is a major factor in determining the legitimacy of the procedure. In this investigation exact levels of significance associated with the nominal one, five, and ten percent levels using a central chi-square distribution for the χ^2 procedure were estimated by Monte Carlo sampling. Samples of 2000 sets of n observations of a given discrete distribution specifying a null hypothesis were generated for each of a number of parameter sets considered and for each of several different values of n.

90

The ability of a statistical procedure to detect departures from the null hypothesis is a major concern in any hypothesis testing problem. Knowledge of the power of a statistical procedure can help in designing studies which use the procedure. Exact values of power for the χ^2 procedure were estimated by Monte Carlo sampling for tests performed at the nominal one, five, and ten percent significance levels. The author was interested in studying the extent to which the empirically determined values of power approximated values which were derived using a noncentral chi-square distribution with the appropriate noncentrality parameter. Interest centered on determining the correspondence of nominal and estimated actual power values for varying sample sizes and different sets of cell probabilities, each set specifying a form of the alternative hypothesis.

The chapter is divided into five sections. The first serves as a literature review, reporting the results of some related research findings. The remaining four sections describe the investigation. The first of these four sections reports the design parameters used in the investigation. The second section describes the generation procedure for the Monte Carlo sampling. The third section reports as well as discusses the results of the investigation and the final section discusses the conclusions reached as a result of the investigation.

RELATED RESEARCH

Bhapkar (1961) demonstrated that a statistic of the form
of χ^2 was algebraically equivalent to the modified minimum chi-square statistic of Neyman (1949) denoted as χ_1^2. The χ^2
statistic can be represented in the form

$$\chi^2 = \chi_1^2 = \sum_{(j_1,j_2,\ldots,j_d)} \frac{(n_{j_1,j_2,\ldots,j_d} - n\hat{P}_{j_1,j_2,\ldots,j_d})^2}{n_{j_1,j_2,\ldots,j_d}} \qquad (4-1)$$

provided $n_{j_1,j_2,\ldots,j_d} > 0$ for all cells of the contingency table.
The $\hat{P}_{j_1,j_2,\ldots,j_d}$ form a set of best asymptotically normal
estimators (BAN) which are least squares estimators found by min-
imizing the quantity given in (4-1) subject to the constraints
which specify the null hypothesis of marginal homogeneity (Neyman,
1949). Neyman demonstrated, that under certain regularity con-
ditions, the statistic

$$\chi_1^2 = \sum_{k=1}^{c} \frac{(n_k - n\hat{p}_k)^2}{n_k}$$

has a limiting central chi-square null distribution with t degrees
of freedom where the $\hat{p}_k$ are BAN estimators found by minimizing
χ_1^2 subject to t linearly independent constraints,

$$F_i(\underline{p}) = 0 \qquad i = 1,2,\ldots,t \qquad (4-2)$$

The statistic given in (4-1) is a special case of the more general result just stated. The constraints which specify the particular null hypothesis of marginal homogeneity possess the appropriate regularity conditions stated by Neyman. Consequently, the statistic in (4-1) has a limiting central chi-square distribution under the null hypothesis of marginal homogeneity. The degrees of freedom are $(d-1)(r-1)$ since $(d-1)(r-1)$ linearly independent constraints are needed to specify the marginal homogeneity null hypothesis.

Neyman showed the asymptotic equivalence of three statistical methods to test a null hypothesis of the form given in (4-2). Use of either the Pearson χ^2,

$$\chi^2 = \sum_{k=1}^{c} \frac{(n_k - n\hat{p}_k)^2}{n\hat{p}_k} \qquad (4-3)$$

the Neyman modified minimum χ_1^2,

$$\chi_1^2 = \sum_{k=1}^{c} \frac{(n_k - n\hat{p}_k)^2}{n_k} \qquad (4-4)$$

or the likelihood ratio statistic

$$-2 \ln \lambda = 2 \sum_{k=1}^{c} n (\ln n_k - \ln(n\hat{p}_k)) \qquad (4-5)$$

leads to asymptotically equivalent results when the $\hat{p}_k$ are BAN estimators of the p_k subject to the constraints of (4-2). Neyman also established that each of these three statistics has as its

limiting null distribution the central chi-square distribution

with t degrees of freedom. Bhapkar (1966) verified that the three

statistics have the same asymptotic relative efficiency.

Research concerning the behavior of the χ^2 statistic

for finite samples may be applicable to the χ^2_1 and λ statistics

because all three statistics belong to the same general class and

all three possess the same asymptotic properties. Since much of

the work on finite sample behavior has been done with the χ^2

statistic, some of the more significant results are cited. Some

of these results may be relevant to the behavior of the χ^2

statistic for finite sample sizes since the χ^2 statistic is

algebraically equal to the χ^2_1 statistic whenever the latter is

defined and χ^2_1 is asymptotically equivalent to the χ^2 statistic.

Asymptotic equivalence does not necessarily imply that the statistics

will behave in a similar manner for finite sample sizes, but a

citation of some of the significant findings in the literature

will help in designing a Monte Carlo study to examine the behavior

of the χ^2 statistic for finite sample sizes.

There has been some disagreement as to how large the

theoretical cell frequencies, $n\hat{p}_k$, must be before the χ^2 statistic

is distributed approximately as a chi-square random variable.

Fisher (1941) recommended that no theoretical cell frequency be

less than five while Cramér (1946) recommended that expectations

should be at least ten. Cochran (1954) suggested that if relatively

fewer than 20% of the cell expectations are less than five a minimum

expectation of 1 is allowable in using the chi-square approximation

to the distribution of the χ^2 statistic.

In the more recent past several authors have shown that the above restrictions are quite conservative in estimating the degree to which the χ^2 statistic approximates its limiting chi-square distribution for small sample sizes. Maxwell (1961) stated that for fifteen or more degrees of freedom, the Pearson χ^2 statistic is well approximated by its limiting chi-square distribution even when most of the expected cell frequencies are as low as 1 or 2. Wise (1963), using a theoretical approach, showed that the χ^2 statistic is well approximated by its limiting distribution when the expected cell sizes are small but nearly equal.

Slakter (1966) obtained the empirical null distributions of the χ^2 statistic when the P_i each equal 1/k, for n's of 10, 15, and 50, and k's providing expected frequencies from 5 (n = 50, k = 10) to .05 (n = 10, k = 200). His study provided further evidence that the chi-square goodness of fit test is robust with respect to small but equal expected cell frequencies. Roscoe and Byars (1971) obtained results similar to those of Slakter in a Monte Carlo study which examined the small sample behavior of χ^2 in tests of fit to a uniform distribution. In addition these authors examined the χ^2 statistic in the context of tests of fit to nonuniform distributions. Roscoe and Byars found that for nonuniform distributions far removed from the uniform, average expected cell frequencies greater than five were needed in order to achieve a good approximation to the limiting distribution.

Good, Gover and Mitchell (1970) theoretically derived the exact distributions for the Pearson χ^2 and the $-2 \ln \lambda$ statistics

in the context of a goodness of fit test for an equiprobable k-category multinomial distribution. The authors found that the chi-square approximation was better for the $-2 \ln \lambda$ statistic than the χ^2 statistic when $n > 3k/2$ or $n > k + 9$ for tests of size less than .005 where n is the total sample size. The Pearson χ^2 statistic was better than the $-2 \ln \lambda$ statistic in terms of approaching its limiting chi-square distribution when $n/k < 1$ for tests of any size.

Yarnold (1970), working in the context of the goodness of fit test, showed that if the number of classes k is three or more, and if r denotes the number of expectations less than five, then the minimum expectation may be as small as $5r/k$ and still achieve a good approximation to the chi-square distribution. Yarnold found that by using this rule, the empirical lower and upper bounds for the true probability of a type I error would be .006 and .0162 for the nominal .01 level and .0375 and .060 for the nominal .05 level.

The cited results of Wise, Slakter, Roscoe, Good and Yarnold are all restricted to the goodness of fit test in which the parameters are completely specified a priori. The results do not reflect the possible effects of estimating the parameters from the data, as in tests of independence or homogeneity. Because the χ^2 procedure does estimate parameters from the sample data, some research findings in this context follow.

The small sample properties of the χ^2 test of homogeneity of independent samples were studied by Roscoe and Byars (1970).

Cases of 2 to 5 samples from multinomial distributions of 2 to 5 categories were investigated. Both uniform and skewed multinomial distributions were examined. For the uniform cases average expected cell sizes of 3 to 5 frequently resulted in empirical estimates of alpha that were within two standard errors of the nominal values of .01 and .05. For the skewed distributions average expected cell sizes of 10 or more were typically needed to meet the same criterion. Use of the Pearson χ^2 statistic resulted in conservative tests when chi-square was used as the limiting reference distribution.

Lewontin and Felsenstein (1965) looked at the Pearson χ^2 test for homogeneity for 2 X k tables and found the test to be conservative for most of the cases considered. They found the chi-square approximation to the Pearson χ^2 test to be good for 5 or more degrees of freedom and all expectations at least of size 1.

Margolin and Light (1974) as part of their study compared the Pearson χ^2 statistic to the likelihood ratio statistic for testing for homogeneity in the context of 3 X 2 contingency tables (three response categories and two independent groups). The authors found that the Pearson χ^2 statistic is considerably better approximated by its limiting chi-square distribution under the null hypothesis than is the likelihood ratio statistic. The authors also proved that when the two independent groups are equal in size the likelihood ratio statistic is numerically larger than the Pearson χ^2 statistic. The authors recommended the use

of the more conservative Pearson χ^2 procedure for small sample sizes.

March's (1970) study, referenced by Tate and Hyer (1973), examined the χ^2 test for independence in 2×3 tables with random margins, sample sizes from 8 to 42 and no expectations less than one. March found that when the average expected cell sizes were 1.5, 2.0, 3.0, 4.0, 5.0, 6.0, and 7.0, the mean absolute percentage errors over the .01 - .10 region of test sizes were 69, 36, 30, 23, 21, 20, and 19 respectively. March concluded that if close approximations to the exact probabilities are needed, the chi-square test may at times be poor.

A special case of the weighted least squares approach as described in Chapter II is the minimum logit χ^2 approach of Berkson (1955). The Berkson approach is used primarily in the testing of hypotheses involving J independent binomial distributions. Let P_j denote the probability of success in population j. The logit Y_j for population j is defined as

$$Y_j = \ell n \frac{P_j}{1-P_j} \quad , \quad \underset{1 \times J}{\underline{Y}'} = [Y_1, Y_2, \ldots, Y_J] \qquad (4.6)$$

Let a null hypothesis to be tested be defined in terms of the linear model

$$\underline{Y} = X \underline{B} \qquad (4.7)$$

where X is a matrix of constants and $\underline{B}$ is a vector of unknown parameters.

Let the observed logits be defined as $\hat{Y}_j = \ell n \frac{\hat{P}_j}{1-\hat{P}_j}$
where the $\{\hat{P}_j\}$ are the observed relative frequencies. Let n_j
denote the size of sample j. The χ^2 (logit) statistic to test
the hypothesis given in (4-7) can be written in the form

$$\chi^2(\text{logit}) = \sum_{j=1}^{J} [n_j \hat{P}_j (1 - \hat{P}_j)(\hat{Y}_j - \tilde{Y}_j)^2] \qquad (4-8)$$

where the $\{\tilde{Y}_j\}$ are those values of the $\{Y_j\}$ which minimize
the quantity

$$\chi^2(\text{logit}) = \sum_{j=1}^{J} [n_j \hat{P}_j (1 - \hat{P}_j)(\hat{Y}_j - Y_j)^2]$$

with respect to the parameters of the model given in (4-7).
Bishop et al. (1974) showed that the statistic in (4-7) could be
written in the form

$$\chi^2(\text{logit}) = (\underline{\hat{Y}} - \underline{\tilde{Y}})' S_{\underline{\hat{Y}}}^{-1} (\underline{\hat{Y}} - \underline{\tilde{Y}}) \qquad (4-9)$$

where $S_{\underline{\hat{Y}}}$ is the estimated variance-covariance matrix of $\underline{\hat{Y}}$. The
χ^2(logit) statistic is thus seen to be a special case of the weighted
least squares approach discussed in Chapter II.

Taylor (1953) has shown that $\{\tilde{Y}_j\}$ belong to the class
of BAN estimators of Neyman. Because the χ^2 statistic is
algebraically equal to a statistic based upon a weighted least
square approach, as has already been noted in Chapter II, the
small sample properties of χ^2(logit) statistics may provide
some information about the behavior of the χ^2 statistic for finite
sample sizes.

Berkson (1968) examined the relationships between the Pearson χ^2 and χ^2 (logit) statistics using maximum likelihood and minimum logit χ^2 estimators respectively for the two statistics. This study was done in the context of tests for no interaction in higher order contingency tables. Berkson found no appreciable difference between the two methods for the sample sizes examined, and recommended the χ^2 (logit) because of its computational simplicity.

In another comparative study Odoroff (1970) looked at the samll sample properties of 12 goodness of fit tests for interaction in $2 \times 2 \times 2$ and $3 \times 2 \times 2$ contingency tables. The 12 tests were constructed by combining three tests (the minimum logit chi-square test, the Pearson chi-square test and the likelihood ratio test) with four methods of estimation (iterative maximum likelihood estimation and three variations of the minimum logit chi-square estimation). Odoroff found that the χ^2 (logit) and the Pearson χ^2 statistic approximated their chi-square limiting distribution better than did the likelihood ratio statistic. In addition Odoroff found that the approximation to the limiting distribution was better when minimum logit chi-square estimation was used in place of estimation by maximum likelihood. Odoroff's was a small sample study which employed contingency tables whose minimum expectations ranged from one to ten observations in a cell.

There still is little agreement as to the requisite sample sizes for use of the chi-square large sample procedures. Some authors were very optimistic as to the applicability of the

techniques to small samples while others were more pessimistic.
The works cited in this literature review aided the author in de-
signing a Monte Carlo study of the small sample properties of
the χ^2 test. The articles suggested possible sample sizes to
examine and distributions, both uniform and nonuniform, to con-
sider. Bhapkar (1966) noted that the χ^2 technique might be in-
accurate in the presence of many empty cells, but he was unsure
of the extent of this inaccuracy. This issue led to the inclusion
of distributions which would give rise to many empty cells, since
such a phenomenon is not rare in social science data analysis.
The design of the Monte Carlo study of the small sample behavior
of χ^2 follows.

DESIGN PARAMETERS USED IN THE INVESTIGATION

Contingency Tables Examined

A design consisting of d matched-samples or d repeated
measures on the same sample can be represented by an
$r \times r \times ... \times r$ contingency table of d dimensions. A multinomial
distribution defined by r^d parameters and a sample of size n
are assumed to characterize the frequencies of the r^d cells of
the contingency table. The number of cells, r^d, grows very rapidly
as either r or d increases. For many data analysis situations
higher order tables are impractical because of the large sample
sizes needed to obtain valid statistical tests. For example, a
4 x 4 x 4 x 4 table contains 256 cells. Even at the very low

figure of an average expected cell size of three, 768 observations would be needed and simulation would be prohibitively expensive.

In this study, 3 x 3, 4 x 4, 5 x 5, and 3 x 3 x 3 tables are considered. Their relatively modest sample size demands, their frequent use in social science data analysis and the reduced simulation costs all prompted this choice.

Distributions Considered Under the Null Hypothesis

The χ^2 procedure tests for homogeneity of correlated marginal distributions. Because the marginal distributions are correlated, the joint distribution of the variables is needed to specify the contingency table completely. The choice of which parameter sets to include in the study must be made on the basis of both the marginal distributions generated and the nature of the configuration specified by the joint distribution. There is no way to investigate the χ^2 procedure for the general case of a set of specified marginal distributions. The procedure can only be investigated for the marginal distributions in the context of a specific configuration or joint distribution. For example, the contingency table given in Figure 4-1 is incompletely specified until at least four joint probabilities are given although its homogeneous marginal distributions are specified.

Distribution 1

Distribution 2

.3
.2
.5

.3   .2   .5

FIGURE 4-1

Two Correlated Distributions

For each of the 3 x 3, 4 x 4, and 5 X 5 contingency

tables 5 different parameter sets were considered under the null

hypothesis of marginal homogeneity for a total of 15 discrete dis-

tributions. The distributions fall into 5 general categories.

The 5 general categories or classes of distributions are denoted

by the letters A through E and are given in Appendix C.

The five general classes of null distributions considered

for study are each characterized by the common property of symmetry.

The symmetric contingency table represents an important subset

of the class of all contingency tables which specify the null hypo-

thesis of marginal homogeneity. It was believed that systematic

investigation of this important subclass would be more enlightening

than attempting to study the entire class of contingency tables

which specify the null hypothesis, a task which is clearly beyond

the scope of any Monte Carlo study of limited finances.

The five general classes of null distributions considered

for the two dimensional contingency tables within the subclass of

symmetric tables provide a wide range of possible practical settings

for the χ^2 technique. The joint distributions which make up

classes A through C have certain common properties which provide
the basis for some interesting comparisons among these classes.
The joint distributions for each of the classes A through C give
rise to uniform marginal distributions. Secondly, cells off the
main diagonal of each contingency table have equal probabilities
associated with them and finally, the main diagonal of each con-
tingency table is characterized by cells of equal probability.

The classes A through C differ in one major respect and
that is the degree to which the probability is concentrated on the
main diagonal of the contingency table. The degree of concentration
of probability on the main diagonal for the two dimensional tables
corresponds to the degree to which subjects do not change their
response for the case of the repeated measures design, and corresponds
to the degree of agreement between pairs of dependent subjects in
the case of two dependent matched samples. It should be noted
that marginal homogeneity can occur regardless of the degree of
probability concentration on the main diagonal of the contingency
table.

The distributions of class-A are each characterized by a
uniform distribution of the probability about the contingency table.
The distributions of class-B are each characterized by having main
diagonal cell probabilities four times those of the off diagonal
cells. This class represents a moderate concentration of proba-
bility on the main diagonal with the concentration being approximately
67%, 57%, and 50% for the 3×3, 4×4, and 5×5 tables
respectively. The distributions of class-C are each characterized

by a very heavy concentration of the total probability on the main diagonal of the contingency table. For each of the distributions within this class between 85% and 88% of the total probability is concentrated on the main diagonal.

The last two classes of distributions, class-D and class-E, define contingency tables in which the marginal distributions are not uniform over the categories of the dependent variable and the off diagonal cell probabilities are not all equal. The class-D distributions were chosen to examine the effect of having a small marginal probability for at least one of the categories of the dependent variable. The low marginal probabilities were chosen to be .10 and .07 respectively for the 3×3 and 4×4 tables and .065 for each of two categories in the 5×5 table. The concentration of probability on the main diagonal was chosen to be relatively high for this class of distributions. The concentration for the 3×3, 4×4, and 5×5 tables were chosen to be approximately 80%, 70% and 68% respectively.

The class-E distributions represent a series of configurations which might arise when there is some meaningful ordering of the categories of the dependent variable. For each of the distributions in this class, cell probabilities decrease in simplex fashion as one moves away from the main diagonal of the contingency table. The χ^2 procedure can be employed to test homogeneity of the marginal distributions, although such an hypothesis does not make use of the ordinality of the data.

Four different configurations specifying the null hypo-
thesis were considered for the 3 × 3 × 3 contingency table. The
distributions are depicted in Appendix D. For the 3 × 3 × 3
table the main diagonal is composed of three cells; (1,1,1), (2,2,2),
and (3,3,3). In the case of a repeated measures design, the degree
of concentration on the main diagonal corresponds to the degree to
which subjects do not change their response over the repeated
measures. For the non repeated measures design, the degree of
concentration on the main diagonal corresponds to the degree of
agreement among the triples of dependent subjects. Just as in the
case of the two dimensional tables, marginal homogeneity can occur
regardless of the level of the concentration of probability on
the main diagonal.

The configurations A and C given in Appendix D are extensions
of the corresponding classes in the two dimensional case to three
dimensions. The two distributions have the same similarities pre-
viously described for the two dimensional case. The joint distribu-
tion A in Appendix D uniformly distributes the total probability
among the 27 cells of the contingency table. Configuration C in
Appendix D indicates a very high concentration of the total probability
in the main diagonal cells with over 85% concentrated on the main
diagonal. Configuration D in Appendix D is an extension of the
class-D distributions to the three dimensional table. Configuration
B in Appendix D has no counterpart in the two dimensional case.
The distribution might arise in the context of a repeated measures
design in which the probability of making only one change in

response over the three repeated measures is greater than the probability of making two changes, and the probability of making two changes but responding with only two different categories is greater than the probability of using all three categories. The concentration of probability on the main diagonal is 46%.

The class-C distributions for both the two and three dimensional contingency tables have the property that for small samples, many of the cells of the corresponding contingency tables may be empty. This class of distributions was chosen in order to investigate the claim of Bhapkar (1966) that the χ^2 statistic may be considerably inaccurate in approaching its limiting distribution when many cells are empty in the contingency table.

Each of the configurations mentioned in this subsection was investigated for sample sizes which were chosen to yield average expected cell frequencies of 3, 5, 10, 20, 40, and 60. The average expected cell frequency, $\bar{n}$, is defined as the total sample size divided by the number of cells in the contingency table, $\bar{n} = \frac{n}{r^d}$. The sample sizes investigated cover a range which extends from the very small samples which might be found in some experimental research studies to the large samples which might be encountered in survey research.

Distributions Considered Under the Alternative Hypothesis

Three different sets of cell probabilities each specifying a form of the alternative hypothesis were investigated for each of the following size contingency tables: 3×3, 4×4, 5×5,

3 x 3 x 3. The configurations are depicted in Appendices E and F.

Estimates of exact power were computed by Monte Carlo sampling for

tests performed at the nominal one, five, and ten percent significance

levels for each of the parameter sets under consideration for sample

sizes yielding average expected cell frequencies of 5, 10, and 20.

These values were compared to their corresponding theoretical values

of power as computed under a noncentral chi-square distribution

with noncentrality parameter given by

$$\lambda = \underline{V}^{*}{}' \Sigma_{\hat{\underline{V}}^*}^{-1} \underline{V}^{*} \tag{4.10}$$

where $\underline{V}^{*}$ and $\Sigma_{\hat{\underline{V}}^*}$ are both functions of the distribution specified

by the alternative hypothesis and sample size under consideration.

The theoretical value of power was found in tables given in Haynam

et al. (1970) for a specified noncentrality parameter. Because

many of the values of λ considered in this study were not given

in the Haynam tables, a Lagrange interpolation formula suggested

by Haynam was used to compute the associated power values. If X

is the given value of λ which is not tabled, the interpolation

procedure consists of taking 6 tabled values of λ such that 3 of

them are above X and 3 are below X. Let these values be denoted

by X_1, X_2, X_3, X_4, X_5, X_6 such that $X_1 < X_2 < X_3 < X_4 < X_5 < X_6$.

Let Y_1, Y_2, Y_3, Y_4, Y_5, Y_6 be the corresponding values of power

for the respective X's. The value of power Y which corresponds

to the noncentrality parameter X is then given by the formula

$$Y = \sum_{i=1}^{6} Y_i \prod_{\substack{j=1 \\ j \neq i}}^{6} \frac{(X - X_j)}{(X_i - X_j)} \qquad (4\text{-}11)$$

The interpolation procedure typically yields values of power that are within .0001 of the theoretical power, which is far in excess of the degree of accuracy needed for this study.

The distributions chosen for the power investigation were selected to yield most theoretical power values in the range .40 to .80 for the case of an average expected cell size of 10 for tests performed at the nominal .01 and .05 significance levels. Moderate values of power were chosen as reflective of true data analytic situations. For average expected cell sizes of 5 and 20 more extreme values of theoretical power were achieved, thus providing the investigation with a very broad range of situations to be examined.

For the two dimensional tables three general classes of distributions specifying a form of the alternative hypothesis were considered. The configurations are given in Appendix E. The class-F distributions are each categorized by having the sole change in marginal probabilities occur between two categories. The change is concentrated in a single off diagonal cell. For the 3 X 3, 4 X 4, and 5 X 5 tables these off diagonal cells are (1,2), (2,1), and (2,1) respectively. The change in the marginal probabilities is relatively large for this distribution class. The class-G distributions are each categorized by having a large change in one of the marginal proportions accompanied by smaller changes

TABLE 4-1

Cases Investigated Under the Null Hypothesis

Table	Distribution	Average Expected Cell Size					$\bar{n}$
3 x 3	A	3	5	10	20	40	60
	B	3	5	10	20	40	60
	C	3	5	10	20	40	60
	D	3	5	10	20	40	60
	E	3	5	10	20	40	60
4 x 4	A	3	5	10	20	40	60
	B	3	5	10	20	40	60
	C	3	5	10	20	40	60
	D	3	5	10	20	40	60
	E	3	5	10	20	40	60
5 x 5	A	3	5	10	20	40	60
	B	3	5	10	20	40	60
	C	3	5	10	20	40	60
	D	3	5	10	20	40	60
	E	3	5	10	20	40	60
3 x 3 x 3	A	3	5	10	20	40	60
	B	3	5	10	20	40	60
	C	3	5	10	20	40	60
	D	3	5	10	20	40	60

TABLE 4-2

Cases Investigated Under the Alternative Hypothesis

Table	Distribution	$\bar{n}$			Noncentrality Parameter[a]
3 × 3	F	5	10	20	4.5283, 9.0566, 18.1132
	G	5	10	20	3.4450, 6.8900, 13.7799
	H	5	10	20	1.4876, 2.9752, 5.9504
4 × 4	F	5	10	20	4.7059, 9.4118, 18.8235
	G	5	10	20	4.4063, 8.8126, 17.6252
	H	5	10	20	5.5920, 11.1840, 22.3680
5 × 5	F	5	10	20	5.3215, 10.6429, 21.2858
	G	5	10	20	4.6661, 9.3322, 18.6643
	H	5	10	20	4.2989, 8.5979, 17.1957
3 × 3 × 3	I	5	10	20	3.4069, 6.8139, 13.6278
	J	5	10	20	4.7230, 9.4461, 18.8921
	K	5	10	20	9.9387, 19.8773, 39.7546

[a] Values of noncentrality parameters correspond to average expected cell sizes of 5, 10, and 20.

in the other categories. The class-H distributions are each categorized by having small to moderate changes in the marginal proportions for each of the categories.

For the 3 X 3 X 3 table three distributions specifying a form of the alternative hypothesis were considered. The configurations are given in Appendix F. Distribution I reflects small to moderate changes in the marginal proportions for all of the categories, with these changes occurring exclusively between the first and second measures. Differences of this type characterize teachers' choices of a disciplinary strategy when increasing increments of information were given about the child involved in the disciplinary incident, Yoshinaga (1974). Although the teacher's initial choice of a disciplinary strategy may not have been a behavior formation strategy, once such a strategy was chosen, it was typically resistant to change on subsequent trials. Distribution J indicates a steady, moderate decrease in the proportion concentrated in a single category accompanied by smaller increases for the other two categories. Such differences might typify the status of farm-related occupations in a generational study of of occupations. Distribution K reflects the same marginal pattern as distribution I; the changes in the marginal proportions for distribution K are each 1.5 times the corresponding changes in distribution I.

DATA GENERATION

For each case which appears in Tables 4-1 and 4-2, 2000 samples were generated. Estimates of the exact significance levels

which correspond to the theoretical values of .01, .05, and .10

for each case considered under the null hypothesis were calculated

by counting the number of rejections out of 2000 using the respective

cutoff points of the central chi-square distribution. Estimates

of exact power were found by counting the number of rejections

out of 2000 for tests performed at the nominal significance levels

of .01, .05, and .10 for each case considered under the alternative

hypothesis.

Data were generated from a population which has a discrete

distribution. For the general r X r X...X r contingency table of

d dimensions the discrete distribution can be characterized by

r^d cell probabilities. Two main steps comprise this generation

procedure:

    (i)   Generating independent random variables which are

            uniformly distributed between zero and one.

    (ii)   Converting the uniformly distributed random vari-

            ables into random variables from the discrete dis-

            tributions specified in the investigation.

Each step is discussed in the next two subsections.

Random Number Generator

A multiplicative congruential generator was used to obtain

the uniform random variates. The generator was described by

Naylor (1968) and adapted to the CDC 6500 computer by Sidney Sytsma.

Starting with an initial number n_0, pseudo-random numbers are

generated according to the recursive formula

$$n_{i+1} \equiv a\, n_i \pmod{P^e} .$$

The numbers a, P, e, and n_0 are chosen to maximize the period of the sequence of pseudo-random numbers generated and minimize the first-order serial correlation between the pseudo-random numbers. For a binary computer P is chosen to be 2 and e is chosen to be 48, the number of bits in a single word integer constant on the CDC 6500 computer. The maximum period which can then be achieved is of length 2^{46}. The size of this period is sufficiently large for almost any Monte Carlo study. Naylor (1968) showed that, to achieve the maximum period, a must be of the form a = 8t + 3, where t is any positive integer. Naylor (1968) recommended choosing a close to 2^{24} in order to minimize first-order serial correlation between the pseudo-random numbers. A further requirement suggests that n , the initial starting point must be chosen relatively prime to 2^{48}.

For Sytsma's adaptation of the generator to the CDC 6500, a was chosen to be a = 2^{24} + 3, and n was chosen as n_0 = 135791357. Such choices guarantee a maximum period of length 2^{46} and a low first-order serial correlation between the pseudo-random numbers. By dividing each random number produced by the generator by the quantity 2^{48}, a uniformly distributed variate defined on the unit interval is obtained.

Three different statistical tests were applied to the generator in order to ascertain if the generator has the desired statistical properties. A description of each of these tests follows.

Each of the tests is applied to a sequence, $\langle u_n \rangle = u_0, u_1, u_2, \ldots$, of real numbers produced by the generator, which purports to be uniformly distributed between zero and one in an independent manner. The three tests applied in this study are designed primarily for integer-valued sequences instead of the real-valued sequence $\langle u_n \rangle$.

The first test applied to the generator determines whether the numbers generated are uniformly distributed between zero and one. In the case of this first test, the auxiliary sequence

$$\langle Y_n \rangle = Y_0, Y_1, Y_2, \ldots,$$

which is defined by the rule

$$Y_n = [100 \, u_n],$$

where $[x]$ is the greatest integer less than or equal to x formed from the sequence $\langle u_n \rangle$, the numbers which are the product of the generator. The sequence $\langle Y_n \rangle$ is a sequence of integers which are uniformly distributed between 0 and 99 if and only if $\langle u_n \rangle$ is uniform between 0 and 1. A sequence $\langle u_n \rangle = u_0, u_1, \ldots, u_{9999}$ of numbers was generated and the corresponding $\langle Y_n \rangle$ sequence was examined. For each integer r, $0 \leq r < 100$, the number of times $Y_j = r$ for $0 \leq j < 10,000$ was counted. Under the null hypothesis that $\langle u_n \rangle$ is uniform on $[0,1]$ or equivalently that $\langle Y_n \rangle$ is a sequence of integers uniformly distributed between 0 and 99, the expected number of times $Y_j = r$ is 100 for each integer, r, $0 \leq r < 100$.

A chi-square statistic of the form

$$\chi^2 = \sum_{r=0}^{99} \frac{(\mathcal{O}_r - 100)^2}{100} , \qquad (4\text{-}12)$$

where $\mathcal{O}_r$ is the observed number of Y_j's which equal r, was used to test the null hypothesis that $\langle Y_n \rangle$ is uniform between 0 and 99. This process was repeated for 20 different sequences $\langle Y_n \rangle$, each of 10,000 observations, and the results combined into an overall chi-square test which retained the null hypothesis that $\langle Y_n \rangle$ and equivalently that $\langle u_n \rangle$ is uniformly distributed. The test was performed at the .10 significance level.

The second test applied to the generator determined whether pairs of successive numbers are uniformly distributed in an independent manner. For this test an auxiliary sequence

$$\langle Y_n \rangle = Y_0, Y_1, Y_2, \ldots$$

which is defined by the rule

$$Y_n = [10 \, u_n]$$

is formed from the sequence

$$\langle u_n \rangle = u_0, u_1, u_2, \ldots$$

This test counts the number of times the pair $(y_{2j}, Y_{2j+1}) = (q, r)$ occurs for $0 \leq j < 10{,}000$; these counts are made for each pair of integers (q, r) with $0 \leq q, r < 10$. Each observation (Y_{2j}, Y_{2j+1}) can fall into any of 100 categories with equal probability of $\frac{1}{100}$. Under the null hypothesis that pairs of successive numbers are uniformly distributed in an independent manner, 100 pairs are expected

to fall into any given category.

A chi-square statistic of the form

$$\chi^2 = \sum_{q=0}^{9} \sum_{r=0}^{9} \frac{(O_{qr} - 100)^2}{100} , \qquad (4-13)$$

where O_{qr} is the number of observed pairs of the form (q,r), is used to test the null hypothesis that $\langle Y_n \rangle$ and equivalently $\langle u_n \rangle$ has pairs of successive elements uniformly distributed in an independent manner. This process was repeated for 20 different sequences $\langle Y_n \rangle$, each consisting of 20,000 numbers or 10,000 pairs of numbers. The results were combined into an overall chi-square test which retained the null hypothesis at a .10 significance level.

The third test was applied to the random number generator to determine whether triples of successive numbers are uniformly distributed in an independent manner. For this test an auxiliary sequence

$$\langle Y_n \rangle = Y_0, Y_1, Y_2, \cdots$$

which is defined by the rule

$$Y_n = \left[\frac{10\ u_n}{2} \right]$$

is formed from the sequence

$$\langle u_n \rangle = u_0, u_1, u_2, \cdots .$$

This test counts the number of times the triple $(Y_{3j}, Y_{3j+1}, Y_{3j+2}) = (p,q,r)$ occurs for $0 \leq j < 10,000$. These counts are made for each triple of integers (p,q,r) with $0 \leq p,q,r < 5$. Each observation $(Y_{3j}, Y_{3j+1}, Y_{3j+2})$ can fall into any of 125 categories with equal probability under the null hypothesis that triples of successive

numbers are uniformly distributed in an independent manner. For
10,000 triples a chi-square statistic of the form

$$\chi^2 = \sum_{p=0}^{4} \sum_{q=0}^{4} \sum_{r=0}^{4} \frac{(O_{pqr} - 80)^2}{80} , \qquad (4-14)$$

where O_{pqr} is the number of observed triples of the form (p,q,r),
is used to test the null hypothesis that $\langle Y_n \rangle$ and equivalently
$\langle U_n \rangle$ has triples of successive elements uniformly distributed in
an independent manner. This process was repeated for 20 different
sequences $\langle Y_n \rangle$, each consisting of 30,000 numbers or 10,000 triples.
The results were combined into an overall chi-square test which re-
tained the null hypotehsis at a .10 significance level.

The results of the three statistical tests performed on the
random number generator indicate that the generator does what it
purports to do; that is, it generates independent uniform random
variates on the unit interval. The second step in the generation
process converts the uniformly distributed random variates produced
by the generator into random variates from a population which has
a specified discrete distribution. The next subsection describes
the conversion process used.

Generation of Discrete Valued Random Variables

Let X be a discrete random variable with $P[X = V_i] = p_i$.
A method for generating X in a computer is to generate a unit
uniform random variable U and put $X = V_i$ if

$$p_1 + \ldots + p_{i-1} < U \leq p_1 + \ldots + p_i \qquad (p_o = 0) \qquad (4-15)$$

Although this method is theoretically sound the method proves to be

too time consuming for discrete random variables which can take on a considerable number of values. Norman and Cannon (1972) developed a computer program for the generation of random variables from any discrete distribution which is far less time consuming than the method given in (4-15). The algorithm which these authors developed generates numbers rapidly (many thousands per second of computer central processing unit time) while requiring a modest amount of computer core storage (no more than a few hundred words). The algorithm is based upon previous work by Marsaglia (1963).

The algorithm is now described with the aid of a particular example. Let X be a discrete random variable of the form given in Table 4-3. Suppose that the computer used for the data generation has storage blocks which may be called for by number. The fastest method of generating the discrete random variable X is probably the following:

In memory locations 0 - 999, store 23 a's, 38 b's, 74 c's, 103 d's,..., 11 m's. Then if $U = .d_1d_2d_3...$ is a unit uniform random variable generated in the computer, look up the number in location $d_1d_2d_3$ and designate it as X.

While the method just described may be the fastest it is not necessarily the most economical. The method requires 1000 storage locations, and had the probabilities been given to four decimal places, 10,000 storage locations would have been required. The method of Marsaglia and the more general method of Norman and Cannon is an improvement on the fastest method in that it uses much less memory space and takes only slightly longer to execute. This makes the Norman and Cannon procedure less costly than the fastest method just described.

TABLE 4-3

Illustrative Discrete Random Variable

Value of X	Probability
a	.023
b	.038
c	.074
d	.103
e	.148
f	.206
g	.140
h	.101
i	.093
j	.037
k	.026
m	.011

The fastest method has been described by Marsaglia as analogous to choosing at random a ball from an urn composed of different balls which correspond to the values of the discrete random variable, with the number of balls of each type relative to the total the same as the corresponding probabilities given by the discrete distribution.

The method of Marsaglia can also be compared to choosing a ball at random but the sampling scheme is slightly different. Instead of one urn there are three urns. The urns are filled according to the digits of the probabilities of X given in Table 4-3. The composition of urn 1 reflects the digits in the tenths position, that of urn 2, the digits in hundreths position, and urn 3 reflects the digits in thousandths position. Thus the contents are:

urn 1: 1 d, 1 e, 2 f's, 1 g, 1 h

urn 2: 2 a's, 3 b's, 7 c's, 4 e's, 4 g's, 9 i's, 3 j's, 2 k's, 1 m

urn 3: 3 a's, 8 b's, 4 c's, 3 d's, 8 e's, 6 f's, 1 h, 3 i's, 7 j's,

$$6 \text{ k's}, 1 \text{ m} \tag{4-16}$$

Let M_i denote the sum of the digits in the ith decimal position of all 12 probabilities given in Table 4-3. Let urn 1 be chosen with probability $10^{-1} \times M_1$, urn 2 be chosen with probability $10^{-2} \times M_2$, and urn 3 be chosen with probability $10^{-3} \times M_3$. For this example these probabilities are .600, .350, and .050 respectively. Urn 1 consists of 6 balls, urn 2 consists of 35 balls, and urn 3 consists of 50 balls, each of which is marked as indicated in (4-16). In all there are 91 balls available for selection. The Marsalgia technique is analogous to a two step procedure in ball selection. First an

urn is selected with probability .600, .350, and .050 for urns 1, 2 and 3 respectively. Once the urn is selected, a ball is randomly chosen. The principle underlying the Marsaglia technique is given in the equality (4-17)

$$P(X = x) = \sum_{i=1}^{3} [P(\text{urn } i \text{ is chosen})P(X = x|\text{urn } i)]$$

(4-17)

for x = a,b,...,m .

The ball and urn approach is now translated into computer terminology and a corresponding generation procedure is given. In a computer the contents of the urns can be stacked in the following manner. Urn 1 occupies storage locations 0 - 5, urn 2, locations 6 - 40, and urn 3, locations 41 - 90. This arrangement is shown in Table 4-4. The source of Table 4-4 is Marsaglia's (1963) article.

TABLE 4-4

Computer Memory Scheme for Discrete Generation

Location	Contents																	
0	d	10	b	20	e	30	i	40	m	50	b	60	e	70	f	80	j	90	m
1	e	11	c	21	e	31	i	41	a	51	b	61	e	71	f	81	j		
2	f	12	c	22	g	32	i	42	a	52	c	62	e	72	f	82	j		
3	f	13	c	23	g	33	i	43	a	53	c	63	e	73	h	83	j		
4	g	14	c	24	g	34	i	44	b	54	c	64	e	74	i	84	k		
5	h	15	c	25	g	35	j	45	b	55	c	65	e	75	i	85	k		
6	a	16	c	26	i	36	j	45	b	56	d	66	e	76	i	86	k		
7	a	17	c	27	i	37	j	47	b	57	d	67	f	77	j	87	k		
8	b	18	e	28	i	38	k	48	b	58	d	68	f	78	j	88	k		
9	b	19	e	29	i	39	k	49	b	59	e	69	f	79	j	89	k		

The two step selection process of choosing an urn and then a ball from the selected urn is translated to the following procedure. Let $u = .d_1 d_2 ...$ be a unit uniform random variable. Let $V(n)$ be the contents of memory location n. Then

	(a)		(b)	
1)	if $d_1 < 6$	put	$X = V(d_1)$	
2)	if $60 \leq d_1 d_2 < 95$	put	$X = V(d_1 d_2 - 54)$	(4-18)
3)	if $950 \leq d_1 d_2 d_3$	put	$X = V(d_1 d_2 d_3 - 919)$.	

In this rule (a) is analogous to determining the urn and (b) to choosing a ball randomly from the selected urn in part (a). Using rule (4-18) together with the equality in (4-17) results in the generation of random variates with the discrete distribution given in Table 4-3. In all only 91 memory locations are needed.

The Marsaglia algorithm was described for the particular distribution given in Table 4-3. Norman and Cannon developed a computer program for the generation of random variables from any discrete distribution. Their procedure follows the specific illustration given with the exception that the probabilities of the discrete distributions are expressed to four decimal places. The program consists of first setting up an array similar to that given in Table 4-4 to describe the desired discrete probability distribution. Once the array is set up, a rule similar to (4-18) is determined. If M_1, M_2, M_3, and M_4 are the sums of the digits in the first, second, third, and fourth positions of the probabilities of the discrete distribution, the number of storage or memory locations needed is given by $M_1 + M_2 + M_3 + M_4$ which is considerably less than the 10,000 locations needed for the computationally faster procedure.

The generation program GEN given in Appendix G is a Fortran
IV version of the general program given by Norman and Cannon. The
program reads in the probabilities which comprise the desired dis-
crete distribution, creates an array similar to the one given in
Table 4-4, generates the unit uniform random numbers using the
multiplicative congruential generator and then converts the uniform
random variates to random variates of the required discrete distribu-
tion using the previously created array and a general assignment
rule similar to that given in (4-18).

For each of the discrete distributions considered in Tables
4-1 and 4-2, 4,000 observations were generated using the Norman and
Cannon algorithm. A chi-square goodness of fit test was performed
in order to determine the adequacy of the fit of the generated data
to their respective theoretical distributions. For a contingency
table characterized by the set of cell probabilities $\{P_1, P_2, \ldots, P_k\}$
the chi-square goodness of fit test has the form

$$\chi^2 = \sum_{i=1}^{k} \frac{(\mathcal{O}_i - 4000\ P_i)^2}{4000\ P_i} \tag{4-19}$$

where $\mathcal{O}_i$ is the number of observations out of 4000 which fall into
cell i. A test of the form given in (4-19) was performed for each
configuration considered in the Monte Carlo study. The results
of these tests are given in Tables 4-5 and 4-6. From these tables
it is seen that out of 31 goodness of fit tests only 2 show signifi-
cant lack of fit for tests performed at the .10 significance level.
These results provide strong evidence that the algorithm effectively
generates the desired discrete distributions.

TABLE 4-5

Goodness of Fit Tests for Null Distributions

Table	Distribution	χ^2 Value	Df	p > [a]
3 x 3	A	9.11	8	.30
	B	11.48	8	.10
	C	5.53	8	.50
	D	4.48	8	.80
	E	9.01	8	.30
4 x 4	A	12.22	15	.50
	B	12.68	15	.50
	C	17.27	15	.30
	D	7.45	15	.90
	E	19.90	15	.10
5 x 5	A	22.92	24	.50
	B	24.61	24	.30
	C	27.46	24	.20
	D	14.07	24	.90
	E	21.97	24	.50
3 x 3 x 3	A	22.27	26	.50
	B	15.73	26	.90
	C	16.05	26	.90
	D	36.06	26	.05[*]

[a] $p = P[\chi^2 \geq \text{critical value}]$

[*] lack of fit at $\alpha = .10$

TABLE 4-6

Goodness of Fit Tests for Non-Null Distributions

Table	Distribution	χ^2 Value	Df	$p >$ [a]
3 X 3	F	12.89	8	.10
	G	6.65	8	.50
	H	11.13	8	.10
4 X 4	F	10.60	15	.70
	G	17.05	15	.30
	H	7.54	15	.90
5 X 5	F	20.99	24	.50
	G	19.09	24	.70
	H	33.90	24	.05*
3 X 3 X 3	I	21.60	26	.70
	J	20.91	26	.70
	K	31.55	26	.20

[a] $p = P[\chi^2 \geq \text{critical value}]$

* lack of fit at $\alpha = .10$

Analysis Routine

Output from the generation program GEN consists of a vector of cell frequencies for each sample. For each case given in Tables 4-1 and 4-2, 2000 such samples were generated. The vector of cell frequencies serves as the input for the analysis routine XSTAT. XSTAT is a program written in Fortran IV which is used to calculate the value of the χ^2 statistic. As input the program uses the vector of cell frequencies, the number of matched samples, d, the number of categories of the dependent variable, r, and a matrix of constants A^* defined in Chapter III. The program is given in general form for any values of d and r in Appendix H. Special programs were written to adapt the more general XSTAT program to the specific designs used in the study.

The calculation of the χ^2 statistic,

$$\chi^2 = \hat{\underline{V}}^{*\prime} \hat{\Sigma}_{\hat{\underline{V}}^*}^{-1} \hat{\underline{V}}^* \quad , \tag{4-20}$$

assumes the nonsingularity of $\hat{\Sigma}_{\hat{\underline{V}}^*}$ which is an estimator of the variance-covariance matrix of $\hat{\underline{V}}^*$. For certain types of contingency tables this matrix may be singular when the sample size is small. This may occur when many of the cells in the contingency table are empty. Berkson (1955) suggested replacing the zero cell frequencies with some small number. Grizzle et al. (1969) and Koch et al. (1974) suggested replacing the zero cell frequencies with a number such as $\frac{1}{r}$ in the event that $\hat{\Sigma}_{\hat{\underline{V}}^*}$ is singular. Such a procedure was employed in this study. This procedure was effective in producing nonsingular estimators of $\Sigma_{\hat{\underline{V}}^*}$. The simulation results, including a discussion of the occurrence of singularities, follow.

SIMULATION RESULTS

This section is divided into five subsections. The first

subsection reports the extent to which singularities occur in the

$\hat{\Sigma}_{\hat{V}^*}$ matrix. The second subsection reports the estimates of the

actual significance level of the χ^2 procedure and discusses

these results. The third subsection reports the estimates of the

actual significance level of the χ^{2*} statistic, which is a slight

variant of χ^2. The fourth subsection reports the estimates of

the actual power of the χ^2 procedure and compares these values

with the nominal values obtained from a noncentral chi-square dis-

tribution. The fifth subsection deals with the power of the χ^{2*}

statistic.

Occurrence of Singularities

In Chapter III it was stated that $\Sigma_{\hat{V}^*}$, the variance-co-

variance matrix of $\hat{\underline{V}}^*$ is nonsingular and that the corresponding

consistent estimator $\hat{\Sigma}_{\hat{V}^*}$ is asymptotically nonsingular. For small

sample sizes and for certain types of data the matrix $\hat{\Sigma}_{\hat{V}^*}$ may be

singular. The presence of singularities was encountered for dis-

tributions of classes C and D for small sample sizes. These results

are reported in Table 4-7.

TABLE 4-7

Number of Singularities in 2000 Samples

Distribution	$\bar{n} =$	3	5	10
3 x 3 C		485	119	4
D		217	0	0
4 x 4 C		340	47	1
D		19	3	0
5 x 5 C		91	4	0
D		38	5	0
3 x 3 x 3 C		23	0	0
D		47	1	0

For sample sizes in which $\bar{n} > 10$ no singularities were encountered. When $\bar{n} = 10$ the singularity problem is nearly non-existent. At $\bar{n} = 5$ the most noticeable problem occurs for the 3 X 3 - C table in which approximately one out of every twenty of the $\hat{\Sigma}_{\hat{V}}*$ generated was singular. The problem of singularities is also noticeable for the 4 X 4 - C table at $\bar{n} = 5$ in which approximately one out of every forty of the $\hat{\Sigma}_{\hat{V}}*$ generated was singular. The class-C distributions are the most troublesome because of the very small probabilities attached to the off-diagonal cells. When $\bar{n}$ is 5 the expected cell frequencies for the off-diagonal cells are between .8 and .9 for the 3 X 3 - C, 4 X 4 - C, 5 X 5 - C, and 3 X 3 X 3 - C tables. In the generation process this may cause some of the off-diagonal cells to be empty. This does not appear to cause inversion problems of the $\hat{\Sigma}_{\hat{V}}*$ matrix in

the 5 x 5 and 3 x 3 x 3 tables. The presence of zeros in the 3 x 3 and 4 x 4 contingency tables will more frequently cause the matrix to be singular.

When the sample size is such that $\bar{n} = 3$ a problem of singularities in the estimated variance-covariance matrix becomes decidedly more pronounced. When $\bar{n} = 3$ for the class-C distributions the off-diagonal cells each have a cell size expectation of approximately .5. This will cause many of the off-diagonal cells to be empty in the generation process. The problem of singularity is quite apparent in the 3 x 3 - C and 4 x 4 - C tables and, to a lesser extent, in the 5 x 5 - C table. The presence of singularities is also quite noticeable for the 3 x 3 - D distribution at $\bar{n} = 3$. At this sample size, 4 of the 6 off-diagonal cells have expected frequencies of .54. This will cause some of the off-diagonal cells to be empty in the generation process.

From the results given in Table 4-7 it appears that the presence of empty cells in the 3 x 3 and 4 x 4 tables tends to be more serious in terms of causing the $\hat{\Sigma}_{\hat{V}}*$ matrix to be singular than a comparable proportion of empty cells in the 5 x 5 and 3 x 3 x 3 table.

It was found that by applying the Berkson procedure of replacing the empty cells with the frequency $\frac{1}{r}$, all singularities were successfully eliminated. The Berkson correction was used only when $\hat{\Sigma}_{\hat{V}}*$ was found to be singular. The results in Table 4-7 correspond to the number of times the Berkson procedure had to be employed out of 2000 samples.

Estimates of Actual Alpha for the χ^2 Statistic

The estimates of the actual alpha are presented in Tables
4-8 through 4-10. The standard errors for the empirical estimates
of alpha are estimated by

$\hat{\sigma}_{\hat{\alpha}} = \sqrt{\text{theoretical alpha} (1 - \text{theoretical alpha})/N}$ where N, the

number of samples for this investigation, is 2000. For the nominal
alpha values of .01, .05, and .10 the corresponding estimates of the
standard errors are .00222, .00487, and .00671 respectively. For the
purposes of this investigation, an empirical estimate which falls
within 1.96 standard errors of its corresponding nominal value will
be considered an adequate approximation to the theoretical value.
Such estimates are within 95% confidence intervals of the nominal
values, and there is insufficient evidence to state that the actual
alpha level and the nominal level are different. The 95% confidence
intervals for the nominal values of alpha of .01, .05, and .10 are
(.0056, .0143), (.0404, .0595), and (.0869, .1131) respectively.

The choice of a 95% level statement is completely arbitrary
although many researchers use it as a rule of thumb. Empirical
estimates which fall outside the 95% confidence bounds may still be
considered adequate approximations of the theoretical alpha values,
depending upon the degree of accuracy required by the data analyst.
Roscoe and Byars (1970) regarded empirical values which fall within
$\pm$ 20% of their corresponding theoretical values as excellent and
believed that $\pm$ 50% would be quite acceptable in most behavioral
applications. Roscoe and Byars claimed that these figures are con-
sistent with the notion of robustness as applied to parametric tests
of hypotheses about means. The empirical estimates of α are given

TABLE 4-8

Monte Carlo Estimates of Exact Level of χ^2
Test Associated with Nominal 1% Level

Distribution		Average Expected Cell Size					
	n =	3	5	10	20	40	60
3 X 3	A	.0265	.0145	.0180	.0130	.0115	.0105
	B	.0135	.0115	.0125	.0115	.0100	.0105
	C	.0010	.0005	.0045	.0090	.0090	.0115
	D	.0045	.0090	.0100	.0060	.0115	.0110
	E	.0150	.0150	.0120	.0135	.0120	.0080
4 x 4	A	.0220	.0205	.0160	.0125	.0090	.0070
	B	.0195	.0145	.0115	.0120	.0120	.0095
	C	.0000	.0035	.0050	.0120	.0090	.0090
	D	.0075	.0115	.0125	.0100	.0105	.0075
	E	.0165	.0060	.0105	.0070	.0080	.0100
5 x 5	A	.0215	.0155	.0130	.0130	.0140	.0105
	B	.0185	.0155	.0115	.0110	.0095	.0065
	C	.0015	.0025	.0085	.0110	.0080	.0140
	D	.0185	.0155	.0115	.0110	.0095	.0065
	E	.0165	.0125	.0100	.0100	.0075	.0110
3 x 3 x 3	A	.0150	.0215	.0135	.0125	.0155	.0130
	B	.0225	.0100	.0150	.0085	.0075	.0070
	C	.0010	.0055	.0095	.0120	.0100	.0100
	D	.0210	.0085	.0095	.0170	.0075	.0110

TABLE 4-9

Monte Carlo Estimates of Exact Level of χ^2
Test Associated with Nominal 5% Level

Distribution	$\bar{n} =$	Average Expected Cell Size					
		3	5	10	20	40	60
3 x 3 A		.0755	.0685	.0650	.0520	.0620	.0470
B		.0725	.0735	.0575	.0595	.0495	.0465
C		.0160	.0300	.0465	.0475	.0525	.0510
D		.0460	.0540	.0480	.0475	.0510	.0530
E		.0750	.0650	.0505	.0560	.0595	.0450
4 x 4 A		.0750	.0645	.0585	.0550	.0560	.0525
B		.0765	.0690	.0540	.0540	.0535	.0560
C		.0105	.0355	.0490	.0535	.0540	.0515
D		.0615	.0590	.0645	.0500	.0530	.0490
E		.0710	.0630	.0590	.0460	.0405	.0510
5 x 5 A		.0760	.0660	.0565	.0545	.0575	.0555
B		.0685	.0640	.0570	.0640	.0500	.0450
C		.0265	.0460	.0515	.0515	.0445	.0605
D		.0620	.0565	.0505	.0585	.0530	.0585
E		.0700	.0685	.0535	.0555	.0475	.0515
3 x 3 x 3 A		.0650	.0685	.0570	.0535	.0535	.0540
B		.0655	.0605	.0570	.0510	.0480	.0390
C		.0330	.0475	.0500	.0500	.0545	.0490
D		.0685	.0510	.0555	.0615	.0505	.0515

TABLE 4-10

Monte Carlo Estimates of Exact Level of χ^2
Test Associated with Nominal 10% Level

Distribution	$\bar{n}$ =	Average Expected Cell Size					
		3	5	10	20	40	60
3 x 3 A		.1380	.1210	.1120	.0980	.1130	.0980
B		.1415	.1260	.1105	.1065	.1060	.0905
C		.0560	.0750	.1220	.0985	.1055	.1025
D		.1140	.1055	.1050	.1000	.1100	.0990
E		.1415	.1220	.1090	.1115	.1090	.0985
4 x 4 A		.1405	.1145	.1135	.1025	.1095	.0965
B		.1435	.1250	.1100	.1060	.1050	.1080
C		.0515	.1015	.1030	.1045	.1050	.0985
D		.1275	.1140	.1220	.0995	.0970	.1065
E		.1260	.1190	.1070	.0990	.0980	.1090
5 x 5 A		.1300	.1225	.1155	.0985	.1135	.1160
B		.1225	.1165	.1105	.1150	.1095	.0975
C		.0800	.1085	.1065	.1055	.0940	.1150
D		.1160	.1140	.1055	.1085	.1000	.1105
E		.1325	.1190	.1010	.1075	.1055	.0970
3 x 3 x 3 A		.1260	.1260	.1105	.1010	.1040	.1130
B		.1225	.1110	.1105	.0985	.0910	.0970
C		.0925	.1150	.1055	.1070	.1015	.1065
D		.1335	.1115	.1005	.1145	.1000	.1045

In Tables 4-8 through 4-10. It is left to the reader to judge

which criterion to use although the author feels that any empirical

estimate falling within the appropriate 95% confidence interval

should be considered a good approximation to the theoretical value.

Table 4-11 reports the number of estimated actual alphas

which are within the limits of a 95% confidence interval for the

corresponding nominal alpha levels. These data exhibit the rela-

tionship between average expected cell size and degree of approxima-

tion to the limiting distribution under the null hypothesis.

TABLE 4-11

Number and Percentage (Out of 19 Different

Distributions) of $\hat{\alpha}$ Within 95% Confidence

Limits of the Corresponding Nominal α

Average Expected Cell size $\bar{n}$	α =	Nominal Alpha .01		.05		.10	
3		3	(15%)	1	(5%)	1	(5%)
5		8	(42%)	6	(32%)	5	(26%)
10		14	(74%)	18	(95%)	15	(79%)
20		18	(95%)	17	(89%)	17	(89%)
40		18	(95%)	17	(89%)	18	(95%)
60		19	(100%)	17	(89%)	17	(89%)

The results given in Table 4-11 are computed across 19

different combinations of distributions and table sizes. The most

striking improvement in the chi-square approximation to the distribu-

tion of χ^2 occurs when $\bar{n}$ is increased from 5 to 10. At $\bar{n}$ = 10

the approximation to the limiting distribution generally appears to

be quite good. There is some further improvement in the approxima-
tion when $\bar{n}$ is increased to 20. Beyond $\bar{n} = 20$, there is no
noticeable improvement in the chi-square approximation to the dis-
tribution of the χ^2 statistic.

One of the more striking characteristics of the results in
Tables 4-8 through 4-10 is the almost uniform liberalness of the
χ^2 procedure for very small sample sizes $(\bar{n} = 3,5)$. The one major
exception to this rule is found in the behavior of the χ^2 statistic
for the distributions of class-C. Each of these distributions is
characterized by a very heavy concentration of the total probability
on the main diagonal of the contingency table. When the average
expected cell size is 10 the expected number of observations within
each off-diagonal cell is between 1.60 and 1.75 for the 3 x 3,
4 x 4, 5 x 5, and 3 x 3 x 3 contingency table. When $\bar{n} = 5$ these
figures drop to .48 and .52 respectively. In general the
procedure appears to be conservative for the class-C distributions,
especially for $\bar{n} = 3$. The chi-square approximation is noticeably
better for the 5 x 5 and 3 x 3 x 3 contingency tables than it is
for the 3 x 3 and 4 x 4 tables. The approximation to the limiting
distribution becomes quite good when $\bar{n} = 5$ for the 5 x 5 - C
and 3 x 3 x 3 - C distributions. For contingency tables of any of
the sizes considered in this investigation the approximation to the
limiting distribution is quite good when $\bar{n} \geq 10$ for the class-C
distributions. These results are quite encouraging because the
distributions of this type give rise to contingency tables in which
a considerable number of off-diagonal cells may be empty or contain
less than two observations when $\bar{n} = 10$, and yet the procedure re-
mains quite valid.

In general for the distributions of classes A, B, C and E for the two dimensional tables and for the distributions of classes A, B, and C for the 3 X 3 X 3 table, the approximation to the chi-square limiting distribution typically does not become good until $\bar{n}$ is at least 10. The χ^2 statistic behaves quite well, however, for the distributions of class-D at $\bar{n} = 5$. At $\bar{n} = 5$, 4 out of 9 cells in the 3 X 3 - D table, 6 out of 16 cells in the 4 X 4 - D table, 14 of 25 cells in the 5 X 5 - D table, and 18 of 27 cells in the 3 X 3 X 3 - D table have expectations of size less than one. Yet despite the presence of a large number of tiny cell size expectations the empirical estimates of alpha are quite close to their nominal values.

At $\bar{n} = 5$, for the distributions of classes B and E for the two dimensional tables and B distribution for the three dimensional case, the majority of the cells have expectations of sizes between 2 and 3 with none below 1. For the distributions of class-A the cell size expectations are 5. It seems unusual that the behavior of the χ^2 statistic for the distributions of class-D should be better than the behavior under these other classes of distributions. An explanation for these results is offered in the next subsection.

Estimates of Actual Alpha for the χ^{2*} Statistic

It was mentioned in Chapter III that the χ^2 statistic,

$$\chi^2 = \hat{\underline{V}}^{*'} \hat{\Sigma}_{\hat{\underline{V}}*}^{-1} \hat{\underline{V}}^{*} \; ,$$

uses $\hat{\Sigma}_{\hat{\underline{V}}*}$, a consistent estimator of $\Sigma_{\hat{\underline{V}}*}$ both under the null and alternative hypothesis. A slightly different statistic, which is

a generalization of the Stuart (1955) statistic discussed in Chapter II, can be used to test for marginal homogeneity. The statistic, denoted as χ^{2*}, has the form

$$\chi^{2*} = \underline{\hat{V}}^{*\prime}\, {}_{0}\hat{\Sigma}_{\underline{\hat{V}}^{*}}^{-1}\, \underline{\hat{V}}^{*}$$

with ${}_{0}\hat{\Sigma}_{\underline{\hat{V}}^{*}}$ having the property that   (4-21)

$$\hat{\Sigma}_{\underline{\hat{V}}^{*}} = {}_{0}\hat{\Sigma}_{\underline{\hat{V}}^{*}} - \frac{1}{n}\, \underline{\hat{V}}^{*}\underline{\hat{V}}^{*\prime} \quad .$$

The matrix ${}_{0}\hat{\Sigma}_{\underline{\hat{V}}^{*}}$ is a consistent estimator of the variance-covariance matrix of $\underline{\hat{V}}^{*}$ under the assumption that the null hypothesis of marginal homogeneity is true. The use of smaller terms in variance-covariance matrix of the χ^{2} statistic, as seen from (4-21), makes the procedure more liberal than the χ^{2*} procedure. An explicit relationship between χ^{2} and χ^{2*} is given by

$$\chi^{2*} = \frac{\chi^{2}}{1 + \frac{1}{n}\chi^{2}} \quad .$$ (4-22)

This relationship is a generalization of the relationship (2-67) given in Chapter II. When n is small the disparity between the χ^{2} and χ^{2*} statistics may be large enough to cause considerable differences in the exact significance levels of the two procedures when used to test the null hypothesis.

Empirical studies cited earlier in this chapter indicated that the fit of the small sample chi-square statistics to the limiting chi-square distribution was quite good for uniform data. When the data were markedly nonuniform with certain cell probabilities very small, the tests tended to be quite conservative for small sample sizes.

The results in the last subsection showed the χ^2 statistic to be quite liberal for small sample sizes ($\bar{n} = 3,5$) for distributions which were uniform (class-A) and other distributions which did not have extremely small off-diagonal cell probabilities. The approximation to the limiting distribution was quite good for the class-D distributions for small sample sizes, especially for $\bar{n} = 5$. From the description of the class-D distributions given in the previous subsection one would expect the χ^2 statistic to be somewhat conservative for this class of distributions. The class-C distributions did produce conservative results but the results were surprisingly better than would be expected for such an extreme distribution in which the off-diagonal cells have such small probabilities.

The author speculated that a statistic based upon the use of $_0\hat{\Sigma}_{\hat{V}}*$ as an estimator of $\Sigma_{\hat{V}}*$ would produce results consistent with some of the findings in the literature review. Exact levels of significance for the χ^{2*} procedure were estimated by Monte Carlo sampling for tests performed at the nominal .01, .05, and .10 levels. These estimates are based upon the same data that were used in estimating the corresponding exact levels for the χ^2 procedure given in Tables 4-8 through 4-10. The results for the χ^{2*} statistic are given in Tables 4-12 through 4-14.

The χ^{2*} statistic behaves quite well for the class A, B, and E distributions for an $\bar{n}$ as low as 5. At the nominal .05 level 9 out of 11 empirical estimates of alpha are within the 95% confidence limits for a nominal alpha of .05 while at the .10

TABLE 4-12

Monte Carlo Estimates of Exact Level of χ^{2*}
Test Associated with Nominal 1% Level

Distribution	$\bar{n} =$	Average Expected Cell Size			
		3	5	10	20
3 x 3 A		.0060	.0065	.0110	.0100
B		.0020	.0035	.0075	.0100
C		.0000	.0000	.0025	.0065
D		.0000	.0020	.0060	.0035
E		.0010	.0060	.0065	.0095
4 x 4 A		.0080	.0095	.0110	.0135
B		.0065	.0040	.0085	.0085
C		.0000	.0010	.0030	.0095
D		.0010	.0050	.0085	.0065
E		.0030	.0015	.0060	.0065
5 x 5 A		.0085	.0075	.0120	.0100
B		.0060	.0090	.0080	.0100
C		.0000	.0015	.0050	.0095
D		.0050	.0050	.0070	.0090
E		.0065	.0070	.0075	.0080
3 x 3 x 3 A		.0070	.0155	.0105	.0095
B		.0075	.0045	.0090	.0080
C		.0005	.0030	.0060	.0095
D		.0060	.0030	.0075	.0155

TABLE 4-13

Monte Carlo Estimates of Exact Level of χ^{2*}
Test Associated with Nominal 5% Level

Distribution		$\bar{n} =$	Average Expected Cell Size			
			3	5	10	20
3 x 3	A		.0440	.0430	.0520	.0470
	B		.0350	.0470	.0440	.0515
	C		.0040	.0140	.0365	.0450
	D		.0155	.0325	.0410	.0420
	E		.0370	.0410	.0435	.0565
4 x 4	A		.0370	.0510	.0485	.0535
	B		.0435	.0510	.0435	.0485
	C		.0025	.0155	.0380	.0490
	D		.0310	.0455	.0545	.0445
	E		.0345	.0475	.0465	.0410
5 x 5	A		.0465	.0495	.0525	.0510
	B		.0465	.0480	.0440	.0580
	C		.0090	.0270	.0435	.0490
	D		.0350	.0430	.0385	.0570
	E		.0430	.0440	.0455	.0525
3 x 3 x 3	A		.0415	.0550	.0495	.0510
	B		.0415	.0485	.0485	.0475
	C		.0130	.0315	.0455	.0540
	D		.0415	.0365	.0495	.0560

TABLE 4-14

Monte Carlo Estimates of Exact Level of χ^{2*}
Test Associated with Nominal 10% Level

Distribution	n =	Average Expected Cell Size			
		3	5	10	20
3 x 3 A		.0995	.0990	.0995	.0940
B		.0845	.1025	.1020	.1055
C		.0160	.0450	.0955	.0880
D		.0505	.0745	.0905	.0910
E		.0870	.0935	.0970	.1030
4 x 4 A		.0970	.0960	.1020	.1090
B		.1030	.1025	.0985	.1005
C		.0245	.0780	.0910	.1015
D		.0805	.0895	.1120	.0950
E		.0950	.0980	.0970	.0920
5 x 5 A		.1030	.1030	.1065	.0955
B		.0880	.0960	.0970	.1120
C		.0420	.0855	.0980	.0975
D		.0805	.0885	.0925	.1035
E		.0935	.0990	.0905	.1040
3 x 3 x 3 A		.0935	.1040	.1050	.0985
B		.0900	.0955	.1035	.0940
C		.0570	.0925	.0965	.1020
D		.0990	.0935	.0825	.1085

nominal level every estimate is within the 95% confidence limits. The behavior of the χ^{2*} statistic is quite good for the uniform distributions of class-A even when $\bar{n}$ is as small as 3. These values are somewhat conservative in contrast to the corresponding χ^2 values which were decidedly liberal especially at $\bar{n} = 3$. The empirical estimates of alpha for the cases using the distributions of class-D are quite conservative at $\bar{n} = 3$ especially for the 3 x 3 and 4 x 4 tables. The fit of χ^{2*} to its limiting chi-square distribution is extremely poor for the distributions of class-C with the procedure being extremely conservative at $\bar{n}$ equalling 3 or 5 and somewhat conservative for the 3 x 3 and 4 x 4 tables at $\bar{n} = 10$.

Table 4-15 reports the number of estimated actual alphas which are within the limits of a 95% confidence interval for the corresponding nominal alpha values when the χ^{2*} statistic is used.

TABLE 4-15

Number and Percentage (Out of 19 Different Distributions) of $\hat{\alpha}$

Within 95% Confidence Limits of the Corresponding

Nominal α for the χ^{2*} Statistic

Average Expected Cell Size $\bar{n}$	Nominal Alpha					
	$\alpha = .01$		.05		.10	
3	9	(47%)	8	(42%)	11	(58%)
5	6	(32%)	13	(68%)	15	(79%)
10	16	(84%)	16	(84%)	19	(100%)
20	17	(89%)	19	(100%)	19	(100%)

A comparison of Tables 4-11 and 4-15 reveals that the χ^{2*} statistic better approximates its limiting distribution than does the χ^2 statistic when $\bar{n} = 3, 5$ and the null distributions are considered as a whole. However, χ^{2*} is extremely conservative when computed for class C and D distributions when $\bar{n} = 3,5$ and at least half of the off-diagonal expected values are less than one. χ^2 should be preferred in these cases. The use of a variance-covariance matrix with smaller terms inflates the χ^2 statistic relative to χ^{2*} and cancels out some of the conservativeness of the χ^{2*} procedure. For the other classes of distributions considered under the null hypothesis (classes A, B, and E) for $\bar{n} = 3$ or 5 none of the off-diagonal cells have cell expectations of size less than one. For such distributions the approximation of the χ^{2*} statistic to its limiting distribution is quite good especially when $\bar{n} = 5$. The smaller terms in the variance-covariance matrix of the χ^2 statistic inflate χ^2 relative to χ^{2*}, making the χ^2 procedure liberal when a good approximation to the limiting distribution might have been expected. The power of both the χ^2 and χ^{2*} procedures is examined in the next two subsections.

Estimates of Actual Power for the χ^2 Procedure

This subsection reports the degree to which the actual power of the χ^2 procedure approximates the power computed using a non-central chi-square distribution with the appropriate noncentrality parameter. The value of power found from the noncentral chi-square distribution is referred to as the asymptotic power. Tables 4-16 through 4-18 each report estimates of the actual power of the χ^2

TABLE 4-16

Monte Carlo Estimates of Actual Power and
Tabled Asymptotic Power for the χ^2 Test
Associated with the Nominal 1% Level

		Average Expected Cell Size					
		5		10		20	
Distribution	$\bar{n}$ =	Est. Actual Power	Asymptotic Power	Est. Actual Power	Asymptotic Power	Est. Actual Power	Asymptotic Power
3 × 3							
	F	.2695	.2402	.5512	.5569	.9275	.9133
	G	.1960	.1676	.4385	.4095	.8005	.7962
	H	.0725	.0603	.1615	.1385	.3500	.3417
4 × 4							
	F	.2310	.2050	.5445	.5062	.9105	.8925
	G	.2265	.1872	.4875	.4688	.8615	.8641
	H	.2995	.2598	.6315	.6097	.9525	.9467
5 × 5							
	F	.2215	.2073	.5535	.5237	.9390	.9097
	G	.1870	.1720	.4645	.4469	.8670	.8550
	H	.1640	.1532	.4075	.4025	.8290	.8141
3 × 3 × 3							
	I	.1335	.1111	.3100	.2939	.6930	.6780
	J	.2065	.1750	.4685	.4537	.8590	.8607
	K	.5220	.4829	.9035	.8830	.9990	.9985

TABLE 4-17

Monte Carlo Estimates of Actual Power and
Tabled Asymptotic Power for the χ^2 Test
Associated with the Nominal 5% Level

		Average Expected Cell Size					
Distribution	$\bar{n} =$	5		10		20	
		Est. Actual Power	Asymptotic Power	Est. Actual Power	Asymptotic Power	Est. Actual Power	Asymptotic Power
3 × 3							
	F	.4995	.4629	.7880	.7734	.9820	.9753
	G	.3990	.3638	.6511	.6480	.9270	.9240
	H	.2020	.1772	.3445	.3191	.5990	.5802
4 × 4							
	F	.4450	.4167	.7575	.7327	.9710	.9666
	G	.4240	.3922	.7080	.7011	.9560	.9552
	H	.5295	.4871	.8325	.8110	.9880	.9865
5 × 5							
	F	.4700	.4203	.7780	.7464	.9860	.9735
	G	.4005	.3709	.7045	.6816	.9700	.9511
	H	.3565	.3430	.6545	.6406	.9345	.9320
3 × 3 × 3							
	I	.3140	.2748	.5445	.5274	.8630	.8348
	J	.4090	.3753	.6925	.6877	.9515	.9536
	K	.7725	.7130	.9765	.9630	.9995	.9998

TABLE 4-18

Monte Carlo Estimates of Actual Power and
Tabled Asymptotic Power for the χ^2 Test
Associated with the Nominal 10% Level

		Average Expected Cell Size					
Distribution		$\bar{n}$ = 5		10		20	
		Est. Actual Power	Asymptotic Power	Est. Actual Power	Asymptotic Power	Est. Actual Power	Asymptotic Power
3 × 3							
	F	.6175	.5893	.8745	.8566	.9905	.9886
	G	.5040	.4891	.7560	.7569	.9513	.9596
	H	.2885	.2752	.4570	.4412	.7330	.6983
4 × 4							
	F	.5705	.5444	.8500	.8255	.9875	.9839
	G	.5420	.5195	.8045	.8006	.9790	.9777
	H	.6520	.6132	.8955	.8842	.9960	.9941
5 × 5							
	F	.5985	.5486	.8670	.8364	.9940	.9876
	G	.5205	.4981	.7990	.7852	.9885	.9754
	H	.4750	.4685	.7580	.7512	.9650	.9643
3 × 3 × 3							
	I	.4305	.3936	.6545	.6515	.9125	.9157
	J	.5345	.5026	.7920	.7901	.9710	.9768
	K	.8500	.8103	.9860	.9820	1.0000	.9999

procedure together with the corresponding values of the asymptotic power assumed under the noncentral chi-square distribution for tests performed at the nominal significance levels of .01, .05, and .10.

The asymptotic power understates the actual power of the χ^2 procedure. Of 109 empirical estimates of the actual power, only 9 are below their corresponding asymptotic values. Of these 9, 7 are within one standard error of their corresponding asymptotic values while the other 2 are within the 95% confidence limits of their respective theoretical values. Table 4-19 reports the number of cases in which the empirical value falls beyond the upper limits of the corresponding 95% interval about the asymptotic power value.

The degree to which the asymptotic power understates the actual power decreases as the average expected cell size increases. This trend is expected since the distribution of the non-null χ^2 statistic should better approximate its limiting noncentral chi-square distribution as the sample size increases. The percentages of empirical power estimates falling within the 95% confidence bounds of their respective asymptotic values are approximately 10%, 60%, and 75% for values of $\bar{n}$ of 5, 10, and 20 respectively.

TABLE 4-19

Number and Percentage (out of 12 Different Distributions) of Empirical Power Values Beyond the Upper Limits of the 95% Confidence Intervals for their Respective Asymptotic Values for Nominal α of .01, .05, .10

Average Expected Cell Size $\bar{n}$	α = .01		Nominal Alpha .05		.10	
5	9	(75%)	11	(92%)	9	(75%)
10	5	(42%)	6	(50%)	3	(25%)
20	3	(25%)	3	(25%)	3	(25%)

Although the results for power are not as theoretically pleasing as the results for actual alpha, the deviation from the theoretical power is in the desirable direction. For those distributions considered under the alternative hypothesis, the asymptotic power generally serves as a lower bound for the actual power of the procedure. The power appears to be at least as good as would be predicted using asymptotic theory, a situation which is highly desirable for practical applications of the statistical procedure.

Although many of the empirical estimates of the actual power of the χ^2 procedure are not within the 95% confidence limits of their respective theoretical values, implying that the actual values of power probably do differ from their corresponding asymptotic values, the actual magnitudes of the differences are quite small for most cases considered. At the .01 nominal significance level the average absolute percent error of the empirical estimate relative to the asymptotic value of power was 13.9% when $\bar{n}$ = 5, 5.1% when $\bar{n}$ = 10, and 1.3% when $\bar{n}$ = 20. When the test was performed at the nominal .05 level, the comparable measures were 9.2%, 2.7% and 1.0% respectively, and when the test was performed at the .10 nominal level the corresponding averages were 5.3%, 1.5% and .8% respectively.

Estimates of Actual Power for the χ^{2*} Procedure

The χ^{2*} statistic is always less powerful than χ^2 as a consequence of the respective definitions of the variance-covariance matrices. In this subsection the degree to which the actual power of the χ^{2*} procedure approximates the asymptotic power computed with noncentrality parameter λ^*, defined as

TABLE 4-20

Monte Carlo Estimates of Actual Power and
Tabled Asymptotic Power for the χ^{2*} Test
Associated with the Nominal 1% Level

Distribution $\bar{n} =$		Average Expected Cell Size					
		5		10		20	
		Est. Actual Power	Asymptotic Power	Est. Actual Power	Asymptotic Power	Est. Actual Power	Asymptotic Power
3 × 3							
	F	.1675	.2116	.4905	.5026	.9170	.8786
	G	.1155	.1522	.3750	.3742	.7785	.7575
	H	.0400	.0582	.1235	.1328	.3235	.3281
4 × 4							
	F	.1555	.1894	.4905	.4736	.9040	.8705
	G	.1555	.1739	.3630	.3732	.8485	.8421
	H	.2095	.2368	.4665	.4586	.9470	.9301
5 × 5							
	F	.1595	.1953	.5060	.4987	.9325	.8940
	G	.1285	.1633	.4260	.4267	.8530	.8373
	H	.1140	.1461	.3685	.3851	.8130	.7959
3 × 3 × 3							
	I	.0955	.1074	.2810	.2837	.6795	.6624
	J	.1515	.1667	.4240	.4346	.8440	.8444
	K	.4390	.4424	.8820	.8512	.9990	.9971

TABLE 4-21

Monte Carlo Estimates of Actual Power and
Tabled Asymptotic Power for the χ^{2*} Test
Associated with the Nominal 5% Level

		Average Expected Cell Size					
		$n =$ 5		10		20	
Distribution		Est. Actual Power	Asymptotic Power	Est. Actual Power	Asymptotic Power	Est. Actual Power	Asymptotic Power
3 × 3							
	F	.4270	.4254	.7605	.7305	.9795	.9617
	G	.3340	.3404	.6330	.6137	.9205	.9037
	H	.1540	.1728	.3165	.3101	.5780	.5657
4 × 4							
	F	.3830	.3953	.7390	.7053	.9675	.9580
	G	.3630	.3732	.6855	.6751	.9535	.9454
	H	.4665	.4586	.8095	.7813	.9865	.9810
5 × 5							
	F	.4110	.4040	.7535	.7261	.9840	.9674
	G	.3495	.3582	.6860	.6633	.9505	.9430
	H	.3110	.3320	.6305	.6237	.9315	.9230
3 × 3 × 3							
	I	.2695	.2685	.5195	.5158	.8560	.8461
	J	.3645	.3631	.6750	.6705	.9485	.9463
	K	.7200	.6776	.9760	.9494	.9995	.9996

TABLE 4-22

Monte Carlo Estimates of Actual Power and
Tabled Asymptotic Power for the χ^{2*} Test
Associated with the Nominal 10% Level

		Average Expected Cell Size					
Distribution	$\bar{n} = 5$		10		20	
		Est. Actual Power	Asymptotic Power	Est. Actual Power	Asymptotic Power	Est. Actual Power	Asymptotic Power
3 × 3							
	F	.5675	.5528	.8615	.8237	.9895	.9814
	G	.4610	.4644	.7370	.7276	.9610	.9470
	H	.2540	.2696	.4320	.4312	.7210	.6853
4 × 4							
	F	.5320	.5227	.8385	.8040	.9855	.9793
	G	.4990	.4998	.7895	.7796	.9780	.9721
	H	.6145	.5858	.8840	.8625	.9955	.9914
5 × 5							
	F	.5590	.5322	.8575	.8207	.9935	.9844
	G	.4750	.4847	.7880	.7701	.9690	.9707
	H	.4360	.4568	.7415	.7369	.9630	.9589
3 × 3 × 3							
	I	.3950	.3863	.6385	.6408	.9100	.9089
	J	.4935	.4899	.7805	.7761	.9700	.9726
	K	.8350	.7819	.9855	.9744	1.0000	.9999

$$\lambda^* = \underline{V}^{*\prime} \,_{o}\Sigma_{\hat{\underline{V}}}^{-1}\!* \, \underline{V}^* \,, \qquad\qquad (4\text{-}23)$$

is examined. Values of the asymptotic power were computed from the Haynam tables for each of the distributions and sample sizes given in Table 4-2. Empirical estimates of the corresponding actual power values were calculated using the same data used in calculating the empirical power values for the χ^2 procedure. The results of these calculations are given in Tables 4-20, 4-21, and 4-22.

Table 4-23 provides a summary of the relationship between sample size and the degree to which the actual power of the χ^{2*} procedure approximates the asymptotic power. All of the estimates indicated in Table 4-23 with the exception of the case of $\alpha = .01$ at $\overline{n} = 5$, are beyond the upper limits of their corresponding 95% confidence intervals. This indicates that the asymptotic power computed by (4-23) tends to underestimate the actual power of the

TABLE 4-23

Number and Percentage (out of 12 Different Distributions) of Empirical Power Values of χ^{2*} Beyond the 95% Confidence Limits of their Respective Asymptotic Values for Nominal α of .01, .05, .10

Average Expected Cell Size n	Nominal Alpha					
	$\alpha = .01$		.05		.10	
5	9	(75%)	3	(25%)	3	(25%)
10	1	(8%)	6	(50%)	5	(42%)
20	5	(42%)	4	(33%)	5	(42%)

χ^{2*} procedure with some exceptions in the case of tests performed at the nominal .01 significance level. These results indicate that, in general, the asymptotic power serves as a lower bound for the

TABLE 4-24

Estimates of Actual Power of the χ^2 and χ^{2*}
Procedures for Tests Performed at the Nominal 5% Level

Distribution	$\bar{n}$ =	Average Expected Cell Size					
		5		10		20	
3 × 3		χ^2	χ^{2*}	χ^2	χ^{2*}	χ^2	χ^{2*}
	F	.4995	.4270	.7880	.7605	.9820	.9795
	G	.3990	.3340	.6511	.6330	.9270	.9205
	H	.2020	.1540	.3445	.3165	.5990	.5780
4 × 4							
	F	.4450	.3830	.7575	.7390	.9710	.9675
	G	.4240	.3630	.7080	.6855	.9560	.9535
	H	.5295	.4665	.8325	.8095	.9880	.9865
5 × 5							
	F	.4700	.4110	.7780	.7535	.9860	.9840
	G	.4005	.3495	.7045	.6860	.9700	.9505
	H	.3565	.3110	.6545	.6305	.9345	.9315
3 × 3 × 3							
	I	.3140	.2695	.5445	.5195	.8630	.8560
	J	.4090	.3645	.6925	.6750	.9515	.9485
	K	.7725	.7200	.9765	.9760	.9995	.9995

actual power of the χ^{2*} procedure.

Table 4-24 indicates that at $\bar{n} = 5$ the differences of the actual power values of the χ^2 and χ^{2*} statistics are rather sizable relative to the power values involved. However, for the larger sample sizes the differences are quite small.

The following conclusions result from this Monte Carlo study.

CONCLUSIONS AND IMPLICATIONS OF THE SIMULATION RESULTS

The main purpose of this simulation study was to examine the behavior of the χ^2 statistic for finite sample sizes both under the null and alternative hypotheses. In addition the behavior of the χ^{2*} statistic, a slight variant of χ^2, was examined for finite sample sizes both under the null and alternative hypotheses. In this section recommendations regarding the use of each of the statistics are made based upon the findings of the study. In the strict sense generalizations from this study are limited to the 3 x 3, 4 x 4, 5 x 5, and 3 x 3 x 3 contingency tables for the con- figurations and sample sizes considered. Nevertheless, certain con- clusions can be drawn from the general patterns of observed statistical behavior of the χ^2 and χ^{2*} procedures which may be applicable to a larger class of cases than those actually studied.

In choosing between the χ^2 and χ^{2*} procedures an important question to consider is the relative importance which the experi- menter places upon guarding against committing a type I or type II error. If a type I error is considered by the experimenter to be the more serious, then a procedure which is conservative might be preferred if the sacrifice in power is not too great. On the other

hand, if a type II error is considered to be the more serious error, then the experimenter might be willing to tolerate a higher actual alpha level for the procedure if some increase in statistical power can be gained.

For the configurations considered in this investigation the distributions of both the χ^2 and χ^{2*} statistics are very well approximated by their limiting central chi-square distribution under the null hypothesis for samples of size $\bar{n} \geq 20$. The actual power values for both of the procedures are quite close in value and are well approximated by the respective asymptotic power values. In general the asymptotic power serves as a lower bound for the actual power of the two procedures. These results indicate that for contingency tables in which $\bar{n} \geq 20$ and all of the cells contain 3 or more observations, one can expect the asymptotic theory to hold quite well. The choice of which statistic to use for this case is a matter of personal choice. The χ^2 procedure is slightly more powerful but the χ^{2*} procedure may provide a slightly better approximation to the respective theoretical alpha values.

The distribution of the χ^2 statistic is well approximated by its limiting central chi-square distribution under the null hypothesis for the configurations considered for samples of size $\bar{n} = 10$. The procedure tends to be slightly liberal when $\bar{n} = 10$. The actual levels of significance tend to be between .01 and .015, .05 and .06, and .10 and .115 for tests performed at the nominal .01, .05, and .10 levels respectively. The actual power of the χ^2 procedure tends to be at least as great as the theoretical power for this size sample. The χ^{2*} statistic tends to be slightly conservative when $\bar{n} = 10$.

In general the actual levels of significance tend to be between .005 and .01, .04 and .05, and .095 and .105 for tests performed at the nominal .01, .05, and .10 levels respectively. For data in which 80 percent or more of the observations are on the main diagonal the χ^2 procedure tends to provide actual levels of significance which are closer to the respective theoretical values. The use of the χ^2 procedure also provides a slight increase in power over the χ^{2*} procedure. The χ^2 is therefore recommended for cases in which 80 percent or more of the observations are concentrated on the main diagonal when $\bar{n} = 10$. The experimenter should expect the asymptotic theory to hold fairly well for both the χ^2 and χ^{2*} procedures for sample sizes of $\bar{n} = 10$ for contingency tables in which each of the off-diagonal cells contains at least a single observation.

For sample sizes in which $\bar{n} \geq 10$ the behaviors of the χ^2 and χ^{2*} statistics were not radically different. When $\bar{n}$ is reduced to 5, however, noticeable differences in the two behaviors are evident. When $\bar{n} = 5$ the χ^2 procedure is quite liberal for configurations in which most of the off-diagonal cells contain two or more observations. For tests performed at the nominal .05 level the experimenter can expect actual alpha values between .06 and .07 for such configurations. For contingency tables in which most of the off-diagonal cells contain 2 or more observations the χ^{2*} statistic is recommended, especially if the experimenter is most concerned with guarding type I errors. For such data the χ^{2*} procedure will typically yield actual levels of significance between .045 and .05 for tests performed at the nominal .05 level. The actual power of the χ^{2*} procedure is somewhat less than the corresponding power of

the χ^2 procedure when $\bar{n} = 5$. The loss in power from using the

χ^{2*} procedure may be from 10 to 15 percent of the actual power of

the χ^2 procedure for values of the latter in the range .30 to .70.

When there is a heavy concentration of the observations on

the main diagonal (80% or more) for samples of size $\bar{n} = 5$, most of

the off-diagonal cells contain less than two observations. If only

20% of the off-diagonal cells are empty the .05 nominal alpha level

is well approximated by χ^2 for the 5 X 5 and 3 X 3 X 3 tables. The χ^2

procedure for such configurations at $\bar{n} = 5$ tends to be slightly

conservative for the larger contingency tables (5 X 5 and 3 X 3 X 3)

and somewhat more conservative for the smaller contingency tables

(3 X 3, 4 X 4). For these smaller tables actual alpha values of

.03 to .04 occur for tests performed at the .05 level. Use of the

χ^{2*} statistic at $\bar{n} = 5$ for tables in which 80% or more of the

observations are concentrated on the main diagonal is not recommended.

For such data the χ^{2*} procedure may have an actual significance

level as low as .015 to .02 for the smaller tables and as low as .025

to .03 for the larger tables for tests performed at the nominal .05

level. For such data singularities in the variance-covariance matrix

may be a problem in the 3 X 3 table and to a slightly lesser extent

in the 4 X 4 table.

When the sample size is reduced to $\bar{n} = 3$ the χ^2 statistic

is quite liberal for configurations in which each off-diagonal cell

contains at least one observation. For tests performed at the nominal

.05 level actual significance levels of .065 to .075 are to be expected

using the χ^2 statistic. For the same type of data actual levels

of significance for χ^{2*} between .035 and .045 occur for tests

performed at the nominal .05 level. For such data the χ^{2*} statistic is recommended especially if the experimenter wants protection against making a type I error. However, use of the χ^{2*} procedure results in a considerable loss in power. When $\bar{n} = 3$ and the majority of the off-diagonal cells is empty, the approximation of the χ^2 statistic to its limiting distribution is quite poor for the 3 X 3 and 4 X 4 tables but considerably better for the 5 X 5 and 3 X 3 X 3 tables. For tests performed at either the nominal .01, .05, or .10 levels the χ^2 statistic is conservative for data of this type. For the 3 X 3 and 4 X 4 tables tests performed at the nominal .05 level may have an actual alpha level as low as .01 to .02. For the 5 x 5 and 3 X 3 X 3 tables the experimenter can expect actual alpha values between .03 and .04 for tests performed at the nominal .05 level. The corresponding values of actual alpha for the χ^{2*} statistic are .004 for the 3 X 3 and 4 X 4 tables and .0165 for the 5 x 5 and 3 X 3 X 3 tables for tests performed at the nominal .05 level. The χ^2 statistic should be preferred to χ^{2*} for data in which the majority of off-diagonal cells is empty. For the larger tables the experimenter should expect a test which is moderately conservative.

The presence of empty cells in the 3 X 3 and 4 X 4 tables results in a relatively frequent occurrence of singularities in the matrix $\hat{\Sigma}_{\hat{V}^*}$. Use of either χ^2 or χ^{2*} for the 3 X 3 or 4 X 4 tables is not recommended if the majority of the off-diagonal cells is empty.

By judiciously choosing between the χ^2 and χ^{2*} procedures the experimenter can expect to have a reasonably valid statistical

procedure for sample sizes as small as $\bar{n} = 5$ and in some cases when $\bar{n}$ is as small as 3. When the choice is between a liberal χ^2 procedure or a conservative χ^{2*} procedure the experimenter must evaluate the relative importance which is attached to guarding against committing a type-I error or type-II error. For those situations in which χ^2 is conservative the choice is made by virtue of the relationship between χ^2 and χ^{2*}.

If the χ^{2*} procedure leads to a statistically significant test, any post-hoc procedures should involve the matrix $\hat{\Sigma}_{\hat{V}*}$ rather than $_{0}\hat{\Sigma}_{\hat{V}*}$ since the $_{0}\hat{\Sigma}_{\hat{V}*}$ matrix is a consistent estimator of $\Sigma_{\hat{V}*}$ only under the assumption that the null hypothesis is true.

The results of this study emphasize the need to examine the small sample properties of statistics which are based upon asymptotic results. Theoretically one should expect the χ^2 and χ^{2*} statistics to have very similar behaviors, yet it was seen that when the sample sizes are sufficiently small the two behaviors can be quite divergent.

Chapter V provides a summary of the results which have been given in this and the previous three Chapters. In addition suggestions for further research are given.

CHAPTER V

SUMMARY AND SUGGESTIONS FOR FURTHER RESEARCH

SUMMARY

This dissertation has dealt exclusively with the problem
of testing for marginal homogeneity in the mixed categorical data
model of order-d. The design implied by the mixed categorical data
model precludes the use of such standard techniques as the chi-
square test of homogeneity to test for the homogeneity of the
marginal distributions. It was demonstrated in Chapter I that a
technique which takes into account the correlated nature of the
marginal distributions is necessary in such an analysis. The
practical utility of such a technique in behavioral science re-
search was also demonstrated in Chapter I.

In Chapter II the probability model associated with the
mixed categorical data model was stated and a rationale for its
structure was given. The necessity of using an $r \times r \times \ldots \times r$
contingency table of d dimensions rather than an $r \times d$ con-
tingency table for such a model was explained. Four different
statistical procedures for dealing with the problem of testing for
homogeneity in the mixed categorical data model were then presented
in some detail. The four procedures presented were the x_S^2
statistic of Stuart (1955); the x_M^2 statistic of Madansky (1963),
a procedure based on the likelihood ratio criterion; the x_I^2

statistic of Ireland et al. (1969), a procedure based on minimum dis-
crimination information estimation used in conjunction with the
minimum discrimination information statistic; and the X^2_{GSK} statistic
of Koch and Reinfurt (1971), a statistic based on weighted least
squares.

All four procedures were shown to be asymptotically equivalent.
The X^2_{GSK} statistic was shown to be algebraically equal to the
modified chi-square statistic χ^2_1 of Neyman (1949) used in conjunc-
tion with modified minimum chi-square estimators of the cell proba-
bilities computed under the null hypothesis. The X^2_S statistic was
shown to be approximately equal to the modified chi-square statistic
χ^2_1 of Neyman used in conjunction with approximate maximum likelihood
estimators computed under the null hypothesis. All four statistics
were shown to belong to the same general class of asymptotic chi-
square statistics employing BAN estimators of the cell probabilities
computed under the null hypothesis. Each of the statistics has a
limiting chi-square distribution with (d-1)(r-1) degrees of freedom
under the null hypothesis of marginal homogeneity.

In Chapter II it was also demonstrated that the X^2_{GSK}
statistic can be put in the form

$$X^2_{GSK} = \hat{\underline{f}}' G^{-1} \hat{\underline{f}} \ . \tag{5-1}$$

The linear model specifying the null hypothesis of marginal homo-
geneity using the X^2_{GSK} procedure gives rise to (d-1)(r-1)
linearly independent constraint equations of the form

$$f_i(\underline{P}) = 0 \qquad i = 1,2,\ldots,(d-1)(r-1) \tag{5-2}$$

The null hypothesis can then be stated in the form

$$H_0: \quad \frac{f(P)}{(d-1)(r-1)} = 0 \tag{5-3}$$

and $\underline{f}' = [f_1(\underline{P}), \ldots, f_{(d-1)(r-1)}(\underline{P})]$.

A statistic which can be used to test H_0 has the form

$$\hat{\underline{f}}' \; G^{-1} \; \hat{\underline{f}}$$

where $\hspace{6cm}$ (5-4)

$$\hat{\underline{f}}' = [f_1(\hat{\underline{P}}), \ldots, f_{(d-1)(r-1)}(\hat{\underline{P}})]$$

and G is a consistent estimator of the variance-covariance matrix of $\hat{\underline{f}}$.

It was also shown that the X_S^2 statistic can be written in the form

$$X_S^2 = \hat{\underline{f}}'\, {}_o G^{-1} \hat{\underline{f}} \tag{5-5}$$

with ${}_o G$ being a consistent estimator of the variance-covariance matrix of $\hat{\underline{f}}$ under the assumption that the null hypothesis is true. The ${}_o G$ matrix has the property that

$$G = {}_o G - \frac{1}{n} \hat{\underline{f}} \; \hat{\underline{f}}' \; . \tag{5-6}$$

From statements (5-1), (5-5), and (5-6) it was demonstrated that the X_{GSK}^2 and X_S^2 are related by the equation

$$X_S^2 = \frac{X_{GSK}^2}{1 + \frac{1}{n} X_{GSK}^2} \; . \tag{5-7}$$

For n large, the numerical difference between the two statistics is small and as n approaches infinity the two procedures are equivalent.

In Chapter III a statistic of the form given in (5-1) was developed. The statistic does not require knowledge of linear models for its understanding and has more intuitive appeal than the x^2_{GSK} statistic. The statistic designated as χ^2 was shown, using asymptotic theory, to have a limiting central chi-square distribution under the null hypothesis and a noncentral chi-square distribution under the alternative hypothesis. The degrees of freedom are $(d-1)(r-1)$. A detailed set of computational formulas for the χ^2 statistic were derived.

The development of two different techniques for generating confidence intervals for contrasts involving the marginal proportions was also given in Chapter III. One of the procedures, a simultaneous procedure, was developed along the lines of the results of Scheffé (1959) and Goodman (1964). A second technique was developed which is based on the Bonferroni inequality. The simultaneous Scheffé-like procedure was found to yield narrower confidence intervals for the usual values of α (.05 or .01) when the number of contrasts examined, k, is such that $k > \frac{1}{2}d(d-1)r$.

In Chapter IV the behavior of the χ^2 statistic in the finite sample situation was examined by the method of simulation. Estimates of both actual alpha and power were found and compared to their respective asymptotic values for a number of different cell probability configurations and sample sizes. In addition the small sample properties of the χ^{2*} statistic were examined. The χ^{2*} statistic is related to χ^2 by the equation

$$\chi^{2*} = \frac{\chi^2}{1 + \frac{1}{n}\chi^2} \quad . \tag{5-8}$$

The extent of the disparity between the χ^2 and χ^{2*} procedures for small n was also considered.

For those configurations considered it was found that the asymptotic theory held quite well for sample sizes of $\bar{n} \geq 20$ and in most cases when $\bar{n} \geq 10$. When $\bar{n}$ was reduced to 5, however, a noticeable decrease in the degree of approximation to the asymptotic results was found. The χ^2 procedure was found to be quite liberal for configurations in which most of the off-diagonal cells contained two or more observations when $\bar{n} = 5$. The χ^{2*} procedure for the same situation behaved quite closely to what asymptotic theory would predict under the null hypothesis. It was found, however, that a considerable loss in power relative to χ^2 took place when the χ^{2*} statistic was used. For the experimenter who is most concerned with controlling for the probability of making a type I error the χ^{2*} procedure was recommended.

It was also shown that when most of the observations were concentrated on the main diagonal of the contingency table (greater than 80%) for samples of size $\bar{n} = 5$, both the χ^2 and χ^{2*} were conservative. The χ^2 statistic did, however, provide a much better approximation to the limiting distribution and was recommended in such situations.

When the sample size was reduced to $\bar{n} = 3$ it was found that the χ^2 was quite liberal for configurations in which each of the off-diagonal cells contained at least one observation. The χ^{2*} procedure for the same situation was moderately conservative. For such data the χ^{2*} procedure was recommended especially if the experimenter wants protection against making a type I error. The

χ^{2*} procedure can result, however, in a considerable loss in power

relative to the χ^2 procedure.

When the sample size is $\bar{n} = 3$ and the data are such that

the majority of the off-diagonal cells is empty, it was found that

the fit of the χ^{2*} statistic to its limiting distribution is

extremely poor. It was also found that the χ^2 statistic is extremely

conservative in such situations for the 3 X 3 and 4 X 4 tables,

but only moderately conservative for the 5 X 5 and 3 X 3 X 3 tables.

The use of the χ^2 was recommended for such data if the contingency

table is either 5 X 5 or 3 X 3 X 3. The experimenter was recommended

not to use either procedure if the table is 3 X 3 or 4 X 4 for such

types of data.

The examination of asymptotic results in the context of finite

sample sizes is a necessity if the asymptotic results are to be of

any use to the applied statistician. The aim of this dissertation

has been to explore a problem along the lines of asymptotic theory

and then to test the theory in the context of data analytic circum-

stances. The financial and time limitations have left certain issues

unexplored. Suggestions for further research in the area under con-

sideration are now given.

SUGGESTIONS FOR FURTHER RESEARCH

In Chapter II four asymptotically equivalent procedures which

test the null hypothesis of marginal homogeneity were presented.

The small sample properties of two of these procedures have been

investigated by the method of simulation. Although asymptotically

equivalent, the behaviors of the χ^2 and χ^{2*} statistics were quite

different for small sample sizes. This suggests that a study of the small sample properties of x_M^2 (likelihood ratio statistic) and x_I^2 (minimum discrimination information statistic) would be both meaningful and extremely useful. A study in which all four statistics are compared would help establish additional guidelines for the applied statistician.

Another area for possible research concerns the manner in which singularities of the variance-covariance matrix $\hat{\Sigma}_{\hat{\underline{V}}^*}$ are handled. The problem of singularities results from the presence of a number of empty cells in the contingency table. Rather than re-placing zero cell frequencies by some arbitrary constant such as $\frac{1}{r}$, Bishop et al. (1974) suggested replacing the observed table of cell frequencies with a new table of what they called "pseudo-Bayes estimates". The method consists of selecting an a priori set of probabilities $\{P_i\}$, which may be based on external information or the data themselves. Based on this a priori distribution, cell estimates are computed using a set of formulas which the authors supply. The advantage of this technique is that one can distinguish among zero frequency cells which have varying probabilities of occurrence. The Berkson rule which was used in this simulation study treats all empty cells in the same way. The rule assumes each of the empty cells is equally likely. The pseudo-Bayes estimates approach can make use of prior knowledge in determining how to deal with the empty cells. Comparisons between the Berkson rule and that suggested by Bishop et al. can be made and their effect on the sampling distributions of the large sample statistics such as χ^2, χ^{2*}, x_M^2 and x_I^2 can be determined.

In the recent past authors such as Koch et al. (1974), and Koch and Reinfurt (1971), have adapted the weighted least squares approach discussed in Chapter II to a series of complex designs in which the dependent variable is categorical. The analysis of categorical data from such designs as the split-plot is quite easily handled using the general weighted least squares methodology. The small sample properties of the statistics generated by these authors have not as yet been investigated. Such an investigation would prove of extreme value to the potential users of the techniques in the behavioral sciences.

Much of the data in the behavior sciences are categorical, but the development of categorical analogues to ANOVA and MANOVA for complex designs is a relatively recent phenomenon. The availability of such techniques and the ability to assess their statistical validity relative to an a priori set of standards will greatly enhance the behavioral scientist's repertoire of categorical data analysis techniques. The findings of this dissertation together with further research along the lines suggested should help greatly in the pursuit of this goal.

APPENDICES

APPENDIX A

PROOF OF THEOREM 2.1

APPENDIX A

PROOF OF THEOREM 2.1

THEOREM 2.1

$$\min \chi_1^2 = \hat{\underline{f}}' G^{-1} \hat{\underline{f}}$$
$$\text{Subject to } H_0$$

The proof consists of finding algebraic expressions for the modified minimum chi-square estimators by minimizing the quantity

$$\sum_i \frac{(n_i - nP_i)^2}{n_i} = n \sum_i \frac{(P_i - \hat{P}_i)^2}{P_i} \qquad (A-1)$$

with respect to the $\{P_i\}$ subject to the constraints specifying the null hypothesis (2-45) with the additional constraint that $\sum_i P_i = 1$. The estimators $\{\tilde{P}_i\}$ are then substituted into (A-1) with the resulting expression shown to be the required quadratic form.

Minimizing the expression in (A-1) subject to the required constraints can be accomplished by minimizing

$$Q = n \sum_i \frac{(P_i - \hat{P}_i)^2}{P_i} - 2\lambda_0 \left[\sum_i P_i - 1 \right] - 2 \sum_m u_m f_m(\underline{P}) \qquad (A-2)$$

with respect to the $\{P_i\}$, where λ_0, $u_1, \ldots, u_t$ are the required Lagrangian multipliers which correspond to the respective constraints under which the minimization takes place.

Differentiating Q with respect to P_i and equating to zero results in r^d equations of the form

$$\frac{n(P_i - \hat{P}_i)}{\hat{P}_i} - \lambda_0 - \sum_m u_m f_{mi} = 0 \qquad (A-3)$$

169

The r^d equations in (A-3) are a consequence of the fact that r^d parameters define the $r \times r \times \ldots \times r$ ccontingency table which is the model under consideration. Multiplying equation (A-3) by $\hat{P}_i$, summing over i, and using the fact that $\sum_i P_i = \sum_i \hat{P}_i = 1$ leads to the result

$$\lambda_0 = \sum_m u_m \hat{f}_m \quad . \tag{A-4}$$

Putting the result in (A-4) back into (A-3) one obtains

$$\frac{n(P_i - \hat{P}_i)}{\hat{P}_i} - \sum_m u_m \hat{f}_m - \sum_m u_m f_{mi} = 0 \quad . \tag{A-5}$$

Solving for P_i one obtains

$$\tilde{P}_i = \hat{P}_i \left[1 + \frac{\sum_m u_m (f_{mi} - \hat{f}_m)}{n} \right] \quad . \tag{A-6}$$

The $\{\tilde{P}_i\}$ are the modified minimum chi-square estimators. Replacing the P_i in (A-1) by the corresponding estimates one obtains after some simplification the resulting expression

$$\sum_i \frac{1}{n} \hat{P}_i [\sum_m \sum_{m'} u_m u_{m'} (f_{mi} - \hat{f}_m)(f_{m'i} - \hat{f}_{m'})]$$

$$= \sum_m u_m \sum_{m'} u_{m'} \sum_i [\frac{1}{n} (f_{mi} - \hat{f}_m)(f_{m'i} - \hat{f}_{m'})\hat{P}_i] \tag{A-7}$$

Lemma 2.1

$$\sum_i [\frac{1}{n} (f_{mi} - \hat{f}_m)(f_{m'i} - \hat{f}_{m'})\hat{P}_i] = g_{mm'} \tag{A-8}$$

Proof of Lemma 2.1

The expression in (A-8) can be written as

$$\sum_{i} \frac{1}{n} [f_{mi} f_{m'i} \hat{P}_{i}] - \frac{1}{n} \hat{f}_{m} \sum_{i} [f_{m'i} \hat{P}_{i}] - \frac{1}{n} \hat{f}_{m'} \sum_{i} [f_{mi} \hat{P}_{i}] + \frac{1}{n} \hat{f}_{m} \hat{f}_{m'}$$

$$= \sum_{i} \frac{1}{n} f_{mi} f_{m'i} \hat{P}_{i} - \frac{1}{n} \hat{f}_{m} \hat{f}_{m'} - \frac{1}{n} \hat{f}_{m'} \hat{f}_{m} + \frac{1}{n} \hat{f}_{m'} \hat{f}_{m}$$

$$= \sum_{i} \frac{1}{n} [f_{mi} f_{m'i} \hat{P}_{i}] - \frac{1}{n} \hat{f}_{m} \hat{f}_{m'}$$

$$= g_{mm'} \quad \text{as defined in (2-47) and (2-49).}$$

Using the result of Lemma 2.1 the second expression in (A-7) can

be written as

$$\sum_{m'} \sum_{m} u_{m} u_{m'} g_{mm'} = \underline{U}' G \underline{U} \quad .$$

Thus far it has been shown that

$$\begin{array}{c} \text{Min} \\ \text{Subject to } H_0 \end{array} \chi_1^2 = \underline{U}' G \underline{U} \qquad \text{(A-9)}$$

where $\quad \underset{1 \times t}{\underline{U}'} = [u_1, u_2, \ldots, u_t]$

Lemma 2.2

$$G \underline{U} + \hat{\underline{f}} = \underline{0} \qquad \text{(A-10)}$$

Proof of Lemma 2.2

The $\{P_i\}$ have been determined in such a way as to satisfy

the constraints specified by the null hypothesis (2-45). By sub-

stituting those values for the estimates found in (A-6) into the

constraint equations (2-45) one obtains the equations

$$\sum_{i} f_{mi} \hat{P}_i \left[1 + \frac{\sum_{m'} u_{m'} (f_{m'i} - \hat{f}_{m'})}{n} \right] = 0 \quad m = 1, 2, \ldots, t .\qquad \text{(A-11)}$$

The t equations in (A-11) after some simplification can be written as

$$\sum_{i} f_{mi} \hat{P}_i + \sum_{m'} u_{m'} [\frac{1}{n} \sum_{i} f_{mi} f_{m'i} \hat{P}_i - \frac{1}{n} \hat{f}_{m'} \hat{f}_m] = 0 \quad m = 1, 2, \ldots, t$$

which further simplifies to

$$\hat{f}_m + \sum_{m'} [u_{m'} g_{mm'}] = 0 \quad m = 1, 2, \ldots, t .\qquad \text{(A-12)}$$

From (A-12) the Lemma 2.2 follows

$$G \underline{U} + \underline{\hat{f}} = \underline{0} .\qquad \text{(A-13)}$$

Because the constraints (2-45) are linearly independent it follows that G is of full rank. Solving for $\underline{U}$ in (A-13) one obtains

$$\underline{U} = -G^{-1} \underline{\hat{f}} .\qquad \text{(A-14)}$$

When the expression for $\underline{U}$ is substituted into the left hand side of the equation (A-9) the required result that

$$\underset{\text{Subject to } H_0}{\text{Min}} \quad \chi_1^2 = \underline{\hat{f}}' \, G^{-1} \, \underline{\hat{f}} \quad \text{follows} .$$

APPENDIX B

PROOF OF THEOREM 2.2

APPENDIX B

PROOF OF THEOREM 2.2

THEOREM 2.2

$$\chi_1^{2*} \doteq \hat{\underline{f}}'_o G^{-1} \hat{\underline{f}}$$

The proof of Theorem 2.2 consists of first obtaining expressions

for the approximate maximum likelihood estimators of the $\{P_j\}$ correct

to terms of order n^{-2}. The obtained expressions are substituted

into the χ_1^2 statistic and the result in (2-60) then follows.

Assuming the multinomial probability model as defined in

(2-38), the likelihood function can be written as

$$\phi = \frac{n!}{\prod_i n_i!} \prod_i P_i^{n_i} . \tag{B-1}$$

Maximizing ϕ subject to the constraints specified by the null

hypothesis (2-45) and the further constraint that $\sum_i P_i = 1$ can be

accomplished by maximizing

$$\log \phi = (\text{constant}) + \sum_i n_i \log P_i - \lambda_0^*(\sum_i P_i - 1) - \sum_m u_m^* f_m(\underline{P}) \tag{B-2}$$

where λ_0^*, $u_1^*, \ldots, u_t^*$ are the appropriate Lagrangian multipliers

which correspond to the constraint conditions.

Differentiating $\log \phi$ with respect to each P_i results

in r^d equations of the form

$$\frac{n_i}{P_i} - \lambda_0^* - \sum_m u_m^* f_{mi} = 0 . \tag{B-3}$$

173

Multiplying through by P_i in (B-3) and summing over i results in the equation

$$n - \lambda_0^* - \sum_m u_m^* \sum_i f_{mi} P_i = 0 .$$

(B-4)

But by the definition of H_0 (2-45) the equation in (B-4) reduces to

$$n - \lambda_0^* = 0 \quad \text{or} \quad \lambda_0^* = n .$$

(B-5)

Replacing λ_0^* in (B-3) by n and solving for P_i one obtains

$$\tilde{P}_i = \frac{n_i}{n} \left[\frac{1}{1 + \dfrac{\sum_m u_m^* f_{mi}}{n}} \right] .$$

(B-6)

The right side of the equation in (B-6) is then expanded in a Taylor series which can be written in the form

$$\frac{n_i}{n} \left[1 - \frac{\sum_m u_m^* f_{mi}}{n} + \frac{(\sum_m u_m^* f_{mi})^2}{n^2} - \frac{(\sum_m u_m^* f_{mi})^3}{n^3} + \ldots + \right.$$
$$\left. (-1)^k \frac{(\sum_m u_m^* f_{mi})^k}{n^k} + R_{k+1} \right] .$$

(B-7)

The maximum likelihood estimators can then be approximated as correct to terms of order n^{-2} by

$$\tilde{P}_i \doteq \frac{n_i}{n} [1 - \frac{1}{n} \sum_{m'} u_{m'}^* f_{m'i}] = \tilde{P}_i^* .$$

(B-8)

Because the $\{\tilde{P}_i\}$ are computed under the constraints specified by the null hypothesis (2-45), substituting the results in (B-8) back into (2-45) yields

$$\sum_{\underline{i}} \left[f_{m\underline{i}} \left[\frac{n_{\underline{i}}}{n} - \frac{1}{n} \cdot \frac{n_{\underline{i}}}{n} \sum_{m'} u^*_{m'} f_{m'\underline{i}} \right] \right] \doteq 0, \text{ for } m = 1,2,\ldots,t$$

which can be written as

$$\sum_{\underline{i}} f_{m\underline{i}} \hat{P}_{\underline{i}} - \sum_{m'} u^*_{m'} \sum_{\underline{i}} [f_{m\underline{i}} f_{m'\underline{i}} \hat{P}_{\underline{i}}] \doteq 0, \text{ for } m = 1,2,\ldots,t \qquad \text{(B-9)}$$

The results in (B-9) can be further simplified to

$$\hat{f}_m - \sum_{m'} [u^*_{m'} \, {_0g_{mm'}}] \doteq 0, \text{ for } m = 1,2,\ldots,t . \qquad \text{(B-10)}$$

where $\hat{f}_m$ and $_0g_{mm'}$ are defined by (2-46) and (2-58) respectively. From (B-10) the result

$$\underline{\hat{f}} \doteq {_0G} \, \underline{U}^*, \text{ with } \underline{U}^{*'} = [u^*_1, u^*_2, \ldots, u^*_t] \qquad \text{(B-11)}$$

follows, with $\underline{\hat{f}}$ and $_0G$ being defined by (2-46) and (2-58) respectively.

The χ_1^{2*} statistic can be written in the form

$$\chi_1^{2*} = n \sum_{\underline{i}} \frac{(\tilde{P}^*_{\underline{i}} - \hat{P}_{\underline{i}})^2}{\hat{P}_{\underline{i}}} . \qquad \text{(B-12)}$$

Substituting the approximations to the $\{\tilde{P}_{\underline{i}}\}$ found in (B-8) into (B-12) gives the result

$$\chi_1^{2*} = n \sum_{\underline{i}} \left[\frac{1}{n^2} \hat{P}_{\underline{i}} (\sum_{m'} u^*_{m'} f_{m'\underline{i}})^2 \right]$$

$$= \sum_m \sum_{m'} [u^*_m u^*_{m'} \sum_{\underline{i}} \frac{1}{n} f_{m\underline{i}} f_{m'\underline{i}} \hat{P}_{\underline{i}}]$$

$$= \sum_m \sum_{m'} [u^*_m u^*_{m'} \, {_0g_{mm'}}] = \underline{U}^{*'} \, {_0G} \, \underline{U}^* . \qquad \text{(B-13)}$$

From (B-11) and using the fact that $_oG$ is nonsingular because of the linear independence of the constraints defining the null hypothesis one obtains

$$\underline{U}^* \doteq {}_oG^{-1}\underline{\hat{f}} \ . \tag{B-14}$$

Substituting the expression in (B-14) for $\underline{U}^*$ into (B-13) yields the final result that

$$\chi_1^{2*} \doteq \underline{\hat{f}}' \ {}_oG^{-1}\underline{\hat{f}} \ . \tag{B-15}$$

APPENDIX C

NULL DISTRIBUTIONS FOR
TWO DIMENSIONAL TABLES

NULL DISTRIBUTIONS FOR TWO DIMENSIONAL TABLES

3 x 3

	(1)	(2)	(3)		
(1)	.1111	.1111	.1111		.3333
(2)	.1111	.1111	.1111		.3333
(3)	.1111	.1111	.1111		.3333
	.3333	.3333	.3333		

4 x 4

	(1)	(2)	(3)	(4)		
(1)	.0625	.0625	.0625	.0625		.2500
(2)	.0625	.0625	.0625	.0625		.2500
(3)	.0625	.0625	.0625	.0625		.2500
(4)	.0625	.0625	.0625	.0625		.2500
	.2500	.2500	.2500	.2500		

5 x 5

	(1)	(2)	(3)	(4)	(5)		
(1)	.0400	.0400	.0400	.0400	.0400		.2000
(2)	.0400	.0400	.0400	.0400	.0400		.2000
(3)	.0400	.0400	.0400	.0400	.0400		.2000
(4)	.0400	.0400	.0400	.0400	.0400		.2000
(5)	.0400	.0400	.0400	.0400	.0400		.2000
	.2000	.2000	.2000	.2000	.2000		

FIGURE C-1

Class-A Null Distributions

3 x 3

	(1)	(2)	(3)	
(1)	.2223	.0555	.0555	.3333
(2)	.0555	.2223	.0555	.3333
(3)	.0555	.0555	.2223	.3333
	.3333	.3333	.3333	

4 x 4

	(1)	(2)	(3)	(4)	
(1)	.1429	.0357	.0357	.0357	.2500
(2)	.0357	.1429	.0357	.0357	.2500
(3)	.0357	.0357	.1429	.0357	.2500
(4)	.0357	.0357	.0357	.1429	.2500
	.2500	.2500	.2500	.2500	

5 x 5

	(1)	(2)	(3)	(4)	(5)	
(1)	.1000	.0250	.0250	.0250	.0250	.2000
(2)	.0250	.1000	.0250	.0250	.0250	.2000
(3)	.0250	.0250	.1000	.0250	.0250	.2000
(4)	.0250	.0250	.0250	.1000	.0250	.2000
(5)	.0250	.0250	.0250	.0250	.1000	.2000
	.2000	.2000	.2000	.2000	.2000	

FIGURE C-2

Class-B Null Distributions

3 x 3

	(1)	(2)	(3)	
(1)	.2933	.0200	.0200	.3333
(2)	.0200	.2933	.0200	.3333
(3)	.0200	.0200	.2933	.3333
	.3333	.3333	.3333	

4 x 4

	(1)	(2)	(3)	(4)	
(1)	.2200	.0100	.0100	.0100	.2500
(2)	.0100	.2200	.0100	.0100	.2500
(3)	.0100	.0100	.2200	.0100	.2500
(4)	.0100	.0100	.0100	.2200	.2500
	.2500	.2500	.2500	.2500	

5 x 5

	(1)	(2)	(3)	(4)	(5)	
(1)	.1720	.0070	.0070	.0070	.0070	.2000
(2)	.0070	.1720	.0070	.0070	.0070	.2000
(3)	.0070	.0070	.1720	.0070	.0070	.2000
(4)	.0070	.0070	.0070	.1720	.0070	.2000
(5)	.0070	.0070	.0070	.0070	.1720	.2000
	.2000	.2000	.2000	.2000	.2000	

FIGURE C-3

Class-C Null Distributions

3 x 3

	(1)	(2)	(3)	
(1)	.3700	.0600	.0200	.4500
(2)	.0600	.3700	.0200	.4500
(3)	.0200	.0200	.0600	.1000
	.4500	.4500	.1000	

4 x 4

	(1)	(2)	(3)	(4)	
(1)	.2200	.0400	.0400	.0100	.3100
(2)	.0400	.2200	.0400	.0100	.3100
(3)	.0400	.0400	.2200	.0100	.3100
(4)	.0100	.0100	.0100	.0400	.0700
	.3100	.3100	.3100	.0700	

5 x 5

	(1)	(2)	(3)	(4)	(5)	
(1)	.2000	.0390	.0390	.0060	.0060	.2900
(2)	.0390	.2000	.0390	.0060	.0060	.2900
(3)	.0390	.0390	.2000	.0060	.0060	.2900
(4)	.0060	.0060	.0060	.0410	.0060	.0650
(5)	.0060	.0060	.0060	.0060	.0410	.0650
	.2900	.2900	.2900	.0650	.0650	

FIGURE C-4

Class-D Null Distributions

3 x 3

	(1)	(2)	(3)	
(1)	.2962	.0741	.0556	.4259
(2)	.0741	.1852	.0370	.2963
(3)	.0556	.0370	.1852	.2778
	.4259	.2963	.2778	

4 x 4

	(1)	(2)	(3)	(4)	
(1)	.2500	.0625	.0313	.0156	.3594
(2)	.0625	.1562	.0313	.0156	.2656
(3)	.0313	.0313	.1250	.0313	.2189
(4)	.0156	.0156	.0313	.0936	.1561
	.3594	.2656	.2189	.1561	

5 x 5

	(1)	(2)	(3)	(4)	(5)	
(1)	.1800	.0400	.0200	.0100	.0100	.2600
(2)	.0400	.1200	.0400	.0200	.0200	.2400
(3)	.0200	.0400	.0800	.0400	.0200	.2000
(4)	.0100	.0200	.0400	.0800	.0200	.1700
(5)	.0100	.0200	.0200	.0200	.0600	.1300
	.2600	.2400	.2000	.1700	.1300	

FIGURE C-5

Class-E Null Distributions

APPENDIX D

NULL DISTRIBUTIONS FOR
THREE DIMENSIONAL TABLES

APPENDIX D

NULL DISTRIBUTIONS FOR THREE DIMENSIONAL TABLES

1st Measure

(1)

2nd Measure

		(1)	(2)	(3)
	(1)	.0370	.0370	.0370
3rd	(2)	.0370	.0370	.0370
Measure	(3)	.0370	.0370	.0370

(2)

2nd Measure

		(1)	(2)	(3)
	(1)	.0370	.0370	.0370
3rd	(2)	.0370	.0370	.0370
Measure	(3)	.0370	.0370	.0370

(3)

2nd Measure

		(1)	(2)	(3)
	(1)	.0370	.0370	.0370
3rd	(2)	.0370	.0370	.0370
Measure	(3)	.0370	.0370	.0370

marginal proportions

$$M_{11} = M_{21} = M_{31} = .3333$$

$$M_{12} = M_{22} = M_{32} = .3333$$

$$M_{13} = M_{23} = M_{33} = .3333$$

FIGURE D-1

The 3 x 3 x 3-A Distribution

1st Measure

(1)

2nd Measure

		(1)	(2)	(3)
	(1)	.1533	.0200	.0200
3rd	(2)	.0300	.0300	.0100
Measure	(3)	.0300	.0100	.0300

(2)

2nd Measure

		(1)	(2)	(3)
	(1)	.0300	.0300	.0100
3rd	(2)	.0200	.1533	.0200
Measure	(3)	.0100	.0300	.0300

(3)

2nd Measure

		(1)	(2)	(3)
	(1)	.0300	.0100	.0300
3rd	(2)	.0100	.0300	.0300
Measure	(3)	.0200	.0200	.1533

marginal proportions

$$M_{11} = M_{21} = M_{31} = .3333$$

$$M_{12} = M_{22} = M_{32} = .3333$$

$$M_{13} = M_{23} = M_{33} = .3333$$

FIGURE D-2

The 3 x 3 x 3-B Distribution

1st Measure

(1)

2nd Measure

		(1)	(2)	(3)
	(1)	.2853	.0060	.0060
3rd	(2)	.0060	.0060	.0060
Measure	(3)	.0060	.0060	.0060

(2)

2nd Measure

		(1)	(2)	(3)
	(1)	.0060	.0060	.0060
3rd	(2)	.0060	.2853	.0060
Measure	(3)	.0060	.0060	.0060

(3)

2nd Measure

		(1)	(2)	(3)
	(1)	.0060	.0060	.0060
3rd	(2)	.0060	.0060	.0060
Measure	(3)	.0060	.0060	.2853

marginal proportions

$M_{11} = M_{21} = M_{31} = .3333$

$M_{12} = M_{22} = M_{32} = .3333$

$M_{13} = M_{23} = M_{33} = .3333$

FIGURE D-3

The 3 x 3 x 3-C Distribution

1st Measure

1st Measure

(1)

2nd Measure

		(1)	(2)	(3)
	(1)	.0600	.0050	.0050
3rd	(2)	.0050	.0050	.0050
Measure	(3)	.0050	.0050	.0050

(2)

2nd Measure

		(1)	(2)	(3)
	(1)	.0050	.0050	.0050
3rd	(2)	.0050	.3050	.0400
Measure	(3)	.0050	.0400	.0400

(3)

2nd Measure

		(1)	(2)	(3)
	(1)	.0050	.0050	.0050
3rd	(2)	.0050	.0400	.0400
Measure	(3)	.0050	.0400	.3050

marginal proportions

$$M_{11} = M_{21} = M_{31} = .1000$$

$$M_{12} = M_{22} = M_{32} = .4500$$

$$M_{13} = M_{23} = M_{33} = .4500$$

FIGURE D-4

The 3 x 3 x 3-D Distribution

APPENDIX E

NON-NULL DISTRIBUTIONS FOR
TWO DIMENSIONAL TABLES

APPENDIX E

NON-NULL DISTRIBUTIONS FOR TWO DIMENSIONAL TABLES

3 x 3

	(1)	(2)	(3)	
(1)	.1600	.2000	.0400	.4000
(2)	.0400	.2200	.0400	.3000
(3)	.0400	.0400	.0400	.3000
	.2400	.4600	.3000	

4 x 4

	(1)	(2)	(3)	(4)	
(1)	.1800	.0100	.0300	.0300	.2500
(2)	.1100	.0800	.0300	.0300	.2500
(3)	.0300	.0300	.1600	.0300	.2500
(4)	.0300	.0300	.0300	.1600	.2500
	.3500	.1500	.2500	.2500	

5 x 5

	(1)	(2)	(3)	(4)	(5)	
(1)	.1600	.0100	.0100	.0100	.0100	.2000
(2)	.0800	.0900	.0100	.0100	.0100	.2000
(3)	.0100	.0100	.1600	.0100	.0100	.2000
(4)	.0100	.0100	.0100	.1600	.0100	.2000
(5)	.0100	.0100	.0100	.0100	.1600	.2000
	.2700	.1300	.2000	.2000	.2000	

FIGURE E-1

Class-F Distributions

186

3 × 3

	(1)	(2)	(3)	
(1)	.3000	.0500	.0500	.4000
(2)	.1300	.1400	.0300	.3000
(3)	.1300	.0300	.1400	.3000
	.5600	.2200	.2200	

4 × 4

	(1)	(2)	(3)	(4)	
(1)	.1200	.0400	.0400	.0500	.2500
(2)	.0500	.1200	.0200	.0600	.2500
(3)	.0100	.0300	.1200	.0900	.2500
(4)	.0300	.0200	.0300	.1700	.2500
	.2100	.2100	.2100	.3700	

5 × 5

	(1)	(2)	(3)	(4)	(5)	
(1)	.1500	.0200	.0100	.0100	.0100	.2000
(2)	.0400	.1000	.0100	.0200	.0300	.2000
(3)	.0400	.0200	.1000	.0200	.0200	.2000
(4)	.0300	.0200	.0300	.1000	.0200	.2000
(5)	.0200	.0200	.0300	.0300	.1000	.2000
	.2800	.1800	.1800	.1800	.1800	

FIGURE E-2

Class-G Distributions

3 x 3

	(1)	(2)	(3)	
(1)	.2400	.1200	.0400	.4000
(2)	.0400	.2200	.0400	.3000
(3)	.0400	.0400	.2200	.3000
	.3200	.3800	.3000	

4 x 4

	(1)	(2)	(3)	(4)	
(1)	.1200	.0200	.0500	.0600	.2500
(2)	.0200	.1200	.0600	.0500	.2500
(3)	.0200	.0200	.1800	.0300	.2500
(4)	.0200	.0200	.0300	.1800	.2500
	.1800	.1800	.3200	.3200	

5 x 5

	(1)	(2)	(3)	(4)	(5)	
(1)	.1400	.0100	.0200	.0100	.0200	.2000
(2)	.0300	.0700	.0200	.0300	.0500	.2000
(3)	.0100	.0200	.1400	.0200	.0100	.2000
(4)	.0400	.0100	.0400	.0700	.0400	.2000
(5)	.0200	.0400	.0200	.0200	.1000	.2000
	.2400	.1500	.2400	.1500	.2200	

FIGURE E-3

Class-H Distributions

APPENDIX F

NON-NULL DISTRIBUTIONS FOR
THREE DIMENSIONAL TABLES

NON-NULL DISTRIBUTIONS FOR THREE DIMENSIONAL TABLES

1st Measure

(1)

2nd Measure

		(1)	(2)	(3)
	(1)	.1400	.0200	.0200
3rd	(2)	.0200	.0400	.0400
Measure	(3)	.0200	.0400	.0400

(2)

2nd Measure

		(1)	(2)	(3)
	(1)	.0200	.0200	.0200
3rd	(2)	.0200	.1500	.0200
Measure	(3)	.0200	.0200	.0200

(3)

2nd Measure

		(1)	(2)	(3)
	(1)	.0200	.0200	.0200
3rd	(2)	.0200	.0200	.0200
Measure	(3)	.0200	.0200	.1500

marginal proportions

$M_{11} = .3800$ $M_{21} = .3000$ $M_{31} = .3000$

$M_{12} = .3100$ $M_{22} = .3500$ $M_{32} = .3500$

$M_{13} = .3100$ $M_{23} = .3500$ $M_{33} = .3500$

FIGURE F-1

The 3 x 3 x 3-I Distribution

(1)

	(1)	(2)	(3)
(1)	.1400	.0100	.0100
(2)	.0100	.0300	.0300
(3)	.0100	.0300	.0300

(2)

	(1)	(2)	(3)
(1)	.0100	.0100	.0100
(2)	.0200	.2200	.0200
(3)	.0200	.0200	.0200

(3)

	(1)	(2)	(3)
(1)	.0100	.0100	.0100
(2)	.0200	.0200	.0200
(3)	.0200	.0200	.2200

<u>marginal proportions</u>

$M_{11} = .3000$ $\quad M_{21} = .2600$ $\quad M_{31} = .2200$

$M_{12} = .3500$ $\quad M_{22} = .3700$ $\quad M_{32} = .3900$

$M_{13} = .3500$ $\quad M_{23} = .3700$ $\quad M_{33} = .3900$

FIGURE F-2

The $3 \times 3 \times 3-J$ Distribution

1st Measure

(1)

2nd Measure

		(1)	(2)	(3)
	(1)	.1800	.0100	.0100
3rd	(2)	.0100	.0400	.0400
Measure	(3)	.0100	.0400	.0400

(2)

2nd Measure

		(1)	(2)	(3)
	(1)	.0100	.0100	.0100
	(2)	.0100	.1400	.0400
	(3)	.0100	.0400	.0400

(3)

2nd Measure

		(1)	(2)	(3)
	(1)	.0100	.0100	.0100
3rd	(2)	.0100	.0100	.0400
Measure	(3)	.0100	.0400	.1400

marginal proportions

$M_{11} = .3800 \qquad M_{21} = .2600 \qquad M_{31} = .2600$

$M_{12} = .3100 \qquad M_{22} = .3700 \qquad M_{32} = .3700$

$M_{13} = .3100 \qquad M_{23} = .3700 \qquad M_{33} = .3700$

FIGURE F-3

The 3 X 3 X 3-K Distribution

APPENDIX G

PROGRAM GEN

APPENDIX G

PROGRAM GEN

```
PROGRAM      GEN

C         PROGRAM GEN(INPUT,OUTPUT,PUNCH,TAPE10)
C         THIS PROGRAM GENERATES 2000 SAMPLES EACH
C         OF SIZE NK FROM A GIVEN DISCRETE DISTRIBUTION
C         OF N CATEGORIES EACH OF WHOSE PROBABILITIES
C         IS GIVEN TO FOUR DECIMAL PLACES.   FOR EACH
C         SAMPLE GENERATED A VECTOR OF CELL FREQUENCIES
C         DENOTED AS OWN IS PUNCHED ONTO A TAPE SO THAT THE DATA
C         GENERATED CAN BE SAVED FOR FURTHER USE.
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C         THE GENERAL DIMENSIONS FOR ARRAYS ARE AS FOLLOWS
C
C         IFREQ(N),OWN(N),LA(K,N),LY(N),N=NUMBER OF CATEGORIES
C         LA(K),K=SUM OF DIGITS IN ALL OF THE PROBABILITIES
C         FOR THIS PARTICULAR PROGRAM N=9 AND K=150
C         DEFINITION OF ARRAYS
C
C         IFREQ AND OWN ARE VECTOR OF CELL FREQUENCIES
C         LA IS ADDRESS ARRAY WHICH IS COMPUTED ONCE
C         LP IS MATRIX WHOSE COLUMNS GIVE DIGITS OF PROBABILITIES
C         LY CONTAINS THE INTEGERS 1 TO N
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C         INPUT
C
C         IX (SEED OF GENERATOR WRITTEN IN BASE 8)
C         NK (SIZE OF SAMPLE)
C         N (NUMBER OF CATEGORIES)
C         LP(4,N) (DIGITS OF FIRST FOUR DECIMAL PLACES OF PROBABILITIES)
C         LY(N) (DIGITS 1 TO N)
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
          DIMENSION IFREQ(9),OWN(9)
          DIMENSION LA(150),LP(4,9),LY(9)
          REWIND 10
          READ(10,990)IX
          IF(EOF(10).NE.0)READ 990,IX
          PRINT 990,IX
C         TAKE LGT TO BE EQUAL TO K
          LGT=150
          READ 2,NK
    2     FORMAT(I4)
          AN=NK
          READ 5,N
    5     FORMAT(I2)
          DO 15 I=1,N
          READ 11,(LP(J,I),J=1,4)
   11     FORMAT(4I2)
   15     CONTINUE
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C         FORMING THE ADDRESS TABLE LOOK-UP FOR A GIVEN DISTRIBUTION
          B=10.0
          LB=10
          READ 20,(LY(M),M=1,N)
   20     FORMAT(9I2)
          I=0
          LT1=0
          LT2=0
          LT3=0
          LT4=0
   25     I=I+1
          LT1=LT1+LP(1,I)
          LT2=LT2+(LB**1)*LP(1,I)+LP(2,I)
          LT3=LT3+(LB**2)*LP(1,I)+(LB**1)*LP(2,I)+LP(3,I)
          LT4=LT4+(LB**3)*LP(1,I)+(LB**2)*LP(2,I)+(LB**1)*LP(3,I)+LP(4,I)
          IF(I.LT.N)GO TO 25
          I=0
          LP1=LB*LT1
          LP2=LB*LT2
          LP3=LB*LT3
   30     I=I+1
          LP1=LP1-LP(1,I)
          LP2=LP2-LP(2,I)
          LP3=LP3-LP(3,I)
          LP2=LP2-LP(1,I)
          LP3=LP3-LP(2,I)
          LP3=LP3-LP(1,I)
          IF(I.LT.N)GO TO 30
          L=1
          J=0
          K=0
```

192

```
PROGRAM     GEN

                 KEY=0
 80          33  I=0
                 J=J+1
             37  I=I+1
                 IF(LP(J,I).EQ.0)GO TO 40
                 K=K+LP(J,I)
 85              IF(K.GT.LGT)GO TO 32
                 CONTINUE
                 GO TO 33
             32  K=LGT+1
                 KEY=1
 90          33  LA(L)=LY(I)
                 L=L+1
                 MLL=L-1
                 IF(MLL.LT.K)GO TO 37
                 IF(KEY.EQ.1)GO TO 35
 95          40  CONTINUE
                 IF(I.LT.N)GO TO 37
                 IF(J.LT.4) GO TO 39
                 GO TO 39
             36  LR3=LT4-L
100          39  CONTINUE
                 PRINT 115,LR1,LR2,LR3
             115 FORMAT(1X,3I10)
                 PUNCH 122,N
             122 FORMAT(I5,2)
105     CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
                 DO 500 KLM=1,2000
                 NKK=0
                 DO 51 K=1,N
                 IFREQ(K)=0
110          51  CONTINUE
        CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
        C        RANDOM NUMBER GENERATOR
              6  IY=IX*65777219
                 IX=IY
115              IF(IY)7,8,8
              7  IY=IY+2314749757103535+1
              8  YFL=IY
                 YFL=YFL*.7552755-14
        CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
120     C        CORRESPONDING THE RANDOM NUMBER GENERATED TO THE
        C        CORRECT CATEGORY OF THE DISCRETE RANDOM VARIABLE
                 D1=8*YFL
                 ID1=D1
                 IF(ID1.LT.LT4)GO TO 50
125              D12=(8**2)*YFL
                 ID12=D12
                 IF(ID12.LT.LT2)GO TO 51
                 D123=(8**3)*YFL
                 ID123=D123
130              IF(ID123.LT.LT3)GO TO 52
                 D1234=(8**4)*YFL
                 ID1234=D1234
                 IF(ID1234.GE.LT4)GO TO 6
                 CONTINUE
135              GO TO 53
             50  IYY=LA(ID1+1)
                 GO TO 71
             51  IYY=LA(ID12-LR1+1)
                 GO TO 71
140          52  IYY=LA(ID123-LR2+1)
                 GO TO 71
             53  IYY=LA(ID1234-LR3+1)
             71  IFREQ(IYY)=IFREQ(IYY)+1
                 NKK=NKK+1
145              IF(NKK.LT.MK)GO TO 6
                 DO 82 LK=1,N
                 CNN(LK)=IFREQ(LK)
             82  CONTINUE
                 PUNCH 123,(CNN(LK),LK=1,N)
150          123 FORMAT(8F4.0)
             500 CONTINUE
                 REWIND 10
                 WRITE(10,990)IX
             990 FORMAT(1X,020)
155              PRINT 990,IX
                 END
```

APPENDIX H

PROGRAM XSTAT

APPENDIX H

PROGRAM XSTAT

```
PROGRAM      XSTAT

          PROGRAM XSTAT(INPUT,OUTPUT,PUNCH)
    C     WHAT FOLLOWS IS A PROGRAM TO CALCULATE THE STATISTIC XSTAT,A
    C     STATISTIC USED TO TEST FOR HOMOGENEITY OF CORRELATED MARGINAL
    C     PROPORTIONS IN A REPEATED MEASURES DESIGN
    C     THE UPPER LIMITS FOR THIS PROGRAM ARE A 5 CATEGORY VARIABLE AND
    C     5 REPEATED MEASURES ON THIS VARIABLE.
    C     THE LIMITS CAN EASILY BE CHANGED
    C
    C     INPUT
    C
    C     IR IS THE NUMBER OF CATEGORIES OF THE VARIABLE
    C     ID IS THE NUMBER OF REPEATED MEASURES
    C     AN IS THE NUMBER OF SUBJECTS
    C     A IS THE A MATRIX
    C     AS IS MATRIX A*
    C     CNN IS VECTOR OF CELL FREQUENCIES WITH
    C     NATURAL CELL ORDERING, RIGHT MOST
    C     SUBSCRIPT VARYING FIRST
    C     USE 1/IR AS THE FREQUENCY FOR AN EMPTY CELL IN
    C     THE EVENT OF THE DETERMINANT (DET) OF THE
    C     VAR-COV MATRIX BEING ZERO
    CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
          DIMENSION CNN(3125),V(625,1),AIP(5,5),SIGMA(25,25)
          DIMENSION AS(16,25),ASV(16,1),F(25,16)
          DIMENSION D(16,16),W1(16),W2(16),VV(16,1),X(1,1)
          DIMENSION A(25,25),VEC(25,1)
          READ 10,IR,ID,AN
       10 FORMAT(I2,1X,I2,1X,F14.4)
          IP=IR**ID
          READ 15,(CNN(I),I=1,IP)
       15 FORMAT(16F4.0)
          IT=(IR-1)*(ID-1)
          M=IR*ID
          DO 17 I=1,M
          READ 16,(A(I,J),J=1,M)
       16 FORMAT(12F5.0)
       17 CONTINUE
          DO 18 I=1,IT
          READ 20,(AS(I,J),J=1,M)
       20 FORMAT(12F5.0)
       18 CONTINUE
    C     WHAT FOLLOWS IS AN ALGORITHM FOR COMPUTING THE MARGINALS FROM THE
    C     CELL FREQUENCIES.
          DO 110 J=1,IR
          DO 110 ID=1,ID
          SUM=0.
          IF(ID.EQ.IQ)GO TO 601
          IRDMQ=IR**(ID-IQ)
      602 CONTINUE
          ILT=(J-1)*IRDMQ
          IUT=ILT+IRDMQ
          GO TO 603
      601 IRDMQ=1
          GO TO 602
      603 CONTINUE
          DO 100 IS=1,IP
          IL=IS/(IR*IRDMQ)
          ITE=IS-(IL*IR)*IRDMQ
          IF(J.EQ.IR)GO TO 201
          IF(ITE.LE.ILT)GO TO 100
          IF (ITE.GT.IUT)GO TO 100
          SUM=SUM+CNN(IS)
          AIP(IQ,J)=SUM
          GO TO 100
      201 IF(ITE.NE.0)GO TO 202
      210 SUM=SUM+CNN(IS)
          AIP(IQ,J)=SUM
          GO TO 100
      202 IF(ITE.LE.ILT)GO TO 100
          IF(ITE.GT.IUT)GO TO 100
          SUM=SUM+CNN(IS)
          AIP(IQ,J)=SUM
      100 CONTINUE
      110 CONTINUE
    C     FORM VECTOR OF MARGINALS FROM MATRIX AIP
          I=1
          DO 200 J=1,IR
          DO 200 IQ=1,ID
```

194

```
PROGRAM     XSTAT

                V(I,1)=AIP(IO,J)
80              I=I+1
            209 CONTINUE
          C     SIGMA IS AN CUBED TIMES VAR-COV MATRIX OF M-HAT
                DC 2 K=1,M
                DO 2 L=1,M
85              SIGMA(K,L)=COV(AIP,CNN,V,IF,IO,K,L,AN,IRO)
              2 CONTINUE
          C     PACKING TO USE MATRIX SUBROUTINES
                CALL PAK(AIP,AIP,5,IO,IP)
                CALL PAK(AS,AS,16,IT,4)
90              CALL PAK(V,V,625,4,1)
                CALL PAK(SIGMA,SIGMA,25,M,M)
                CALL PAK(A,A,25,4,4)
                CALL MPY(AS,V,ASV,IT,4,0,0,1)
                CALL MPYT(SIGMA,A,C,M,4,4,0,IT)
95              CALL MPY(AS,C,U,IT,4,0,0,IT)
                CALL INV(D,IT,DET,W1,W2)
                IF(DET.NE.0.)GO TO 40
                PRINT 35
            35  FORMAT(1X,*DET WAS ZERO-ABORT*)
100             GO TO 150
            40  CONTINUE
                CALL MPY(D,ASV,VV,IT,IT,0,0,1)
                CALL MPYTP(ASV,VV,X,IT,1,0,0,1)
                CALL MPY(A,V,VEC,4,4,0,0,1)
105             XSTAT=X(1,1)
                XSTAT=AN*XSTAT
          C         WE HAVE COMPUTED THE TEST STATISTIC
                CALL UNPAK(VEC,VEC,25,4,1)
                CALL UNPAK(D,D,16,IT,IT)
110             CALL UNPAK(AIP,AIP,5,IO,IP)
                CALL UNPAK(SIGMA,SIGMA,25,M,M)
                CALL UNPAK(VEC,VEC,25,4,1)
                CALL UNPAK(ASV,ASV,16,IT,1)
                PRINT 50
115         50  FORMAT(1H1,T38,*TEST FOR HOMOGENEITY OF CORRELATED MARGINALS*,//)
            65  FORMAT(1H0)
                DO 77 I=1,IO
                DO 77 J=1,IR
                AIP(I,J)=1./AN*AIP(I,J)
120         77  CONTINUE
                PRINT 55
            55  FORMAT(//,1X,*MARGINAL PROBABILITIES-COLUMNS ARE CATEGORIES AND
               1ROWS ARE DIMENSION OF RESPONSE*,//)
                DO 70 I=1,IO
125             PRINT 60,(AIP(I,J),J=1,IR)
            60  FORMAT(1X,5(F3.6,5X))
            70  CONTINUE
                DO 78 I=1,IO
                PUNCH 73,(AIP(I,J),J=1,IR)
130         73  FORMAT(F(F8.6,5X))
            78  CONTINUE
                H=AN**3
                DO 26 I=1,M
                PRINT 90,(SIGMA(I,J),J=1,M)
135         90  FORMAT(1X,5(F12.0,5X))
            26  CONTINUE
                DO 97 I=1,M
                PUNCH 96,(SIGMA(I,J),J=1,M)
140         96  FORMAT(5(F12.0,3X))
            97  CONTINUE
                G=IO
                DO 101 I=1,M
                VEC(I,1)=(1./G*1./AN)*VEC(I,1)
145        101  CONTINUE
                PRINT 102
           102  FORMAT(//,1X,*THE VECTOR V-HAT*,//)
                PRINT 104,(VEC(I,1),I=1,4)
           104  FORMAT(1X,F4.6)
150             PRINT 105
           105  FORMAT(//,1X,*THE VECTOR VSTAR-HAT*,//)
                DO 106 I=1,IT
                ASV(I,1)=(1./G*1./AN)*ASV(I,1)
           106  CONTINUE
155             PRINT 107,(ASV(I,1),I=1,IT)
           107  FORMAT(1X,F8.6)
```

```
PROGRAM    XSTAT
              DO 93 I=1,IT
              DO 93 J=1,IT
              E(I,J)=(AN**3)*(G**2)*D(I,J)
160        93 CONTINUE
              PRINT 117
          117 FORMAT(//,1X,*THE INVERSE OF VAR-COV MATRIX OF VSTAR-HAT
             1(NOTE SPACE BETWEEN ROWS)*)
              DO 118 I=1,IT
165           PRINT 119,(D(I,J),J=1,IT)
          119 FORMAT(1X,5(F20.6,3X))
              PRINT 65
          118 CONTINUE
              PRINT 140,RSTAT
170       140 FORMAT(//,1X,*THE VALUE OF TEST STATISTIC ESTAT IS*,/,1X,F26.6)
          150 CONTINUE
              END
```

```
FUNCTION    COV

            FUNCTION COV(AIP,CNN,V,IP,ID,K,L,AN,IRD)
            DIMENSION AIP(5,5),CNN(3125),V(625,1)
            IF(K.EQ.L) GO TO 11
            LL=(K-1)/ID+1
 5          NNN=K-(K/ID*ID)
            LLP=(L-1)/ID+1
            NNP=L-(L/ID*ID)
            IF(NNN.EQ.0)GO TO 13
        29  CONTINUE
10          IF(NNP.EQ.0)GO TO 14
        30  IF(NNN.EQ.NNP) GO TO 20
            SUM=0.
            IF(ID.EQ.NNN)GO TO 701
            IPDMQ=IP**(ID-NNN)
15      702 CONTINUE
            ILT=(LL-1)*IPDMQ
            IUT=ILT+IPDMQ
            GO TO 703
       701  IPDMQ=1
20          GO TO 702
       703  CONTINUE
            IF(ID.EQ.NNP)GO TO 801
            IPDMQP=IP**(ID-NNP)
       802  CONTINUE
25          ILTP=(LLP-1)*(IPDMQP)
            IUTP=ILTP+IPDMQP
            GO TO 803
       801  IPDMQP=1
            GO TO 802
30     803  CONTINUE
            IPD=IP**ID
            DO 1 IS=1,IPD
            IL=IS/(IP*IPDMQ)
            ITE=IS-IL*IP*IPDMQ
35          ILP=IS/(IP*IPDMQP)
            ITP=IS-ILP*IP*IPDMQP
            IF(LL.NE.IP.AND.LLP.NE.IP)GO TO 401
            IF(LL.EQ.IP.AND.LLP.EQ.IP)GO TO 402
            IF(LL.EQ.IP)GO TO 412
40          IF(ITE.LE.ILT.OR.ITE.GT.IUT)GO TO 1
            IF(ITP.EQ.0)GO TO 499
            IF(ITP.GT.ILTP.AND.ITP.LE.IUTP)GO TO 499
            CONTINUE
            GO TO 1
45     412  IF(ITP.LE.ILTP.OR.ITP.GT.IUTP)GO TO 1
            IF(ITE.EQ.0)GO TO 499
            IF(ITE.GT.ILT.AND.ITE.LE.IUT)GO TO 499
            CONTINUE
            GO TO 1
50     402  IF(ITE.EQ.0.AND.ITP.EQ.0)GO TO 499
            IF(ITE.NE.0.AND.ITE.NE.0)GO TO 401
            IF(ITE.NE.0.AND.ITP.EQ.0)GO TO 411
            IF(ITP.LE.ILTP.OR.ITP.GT.IUTP)GO TO 1
            CONTINUE
55          GO TO 499
       411  IF(ITE.LE.ILT.OR.ITE.GT.IUT)GO TO 1
            CONTINUE
            GO TO 499
       401  IF(ITE.LE.ILT.OR.ITE.GT.IUT)GO TO 1
60          IF(ITP.LE.ILTP.OR.ITP.GT.IUTP)GO TO 1
       499  SUM=SUM+CNN(IS)
         1  CONTINUE
            COV=SUM
            COV=AN*COV
65          COV=COV-AIP(NNN,LL)*AIP(NNP,LLP)
            GO TO 50
        10  COV=V(K,1)*(AN-V(K,1))
            GO TO 50
        20  COV=-1.*AIP(NNN,LL)*AIP(NNP,LLP)
70          GO TO 50
        13  NNN=ID
            GO TO 29
        14  NNP=ID
            GO TO 30
75      50  CONTINUE
            RETURN
            END
```

SUBROUTINE INV

```
                SUBROUTINE INV(A,N,C,W1,W2)
      C
      C         CALL INV(A,N,C,W1,W2)
      C
      C                 A        MATRIX TO BE INVERTED
      C                 N        NUMBER OF COLUMNS(ROWS) OF A
      C                 C        DETERMINANT OF A STORED HERE
      C                 W1       N LOCATIONS OF WORKING STORAGE
      C                 W2       N LOCATIONS OF WORKING STORAGE
      C
      C         INVERT AN N BY N NON-SINGULAR MATRIX A IN PLACE.   LEAVE
      C             THE DETERMINANT OF A IN C.
      C
      C         INVERSE REPLACES A
      C
                DIMENSION W1(N), W2(N), A(N,N)
                INTEGER W1,W2
                INTEGER FLIPS
                EQUIVALENCE (M,T,ISAV),(L,U,JSAV)
                FLIPS=0
      C
                C=1.0D0
                D=1.0D0
                DO 1 K=1,N
                ISAV=K
                JSAV=K
                X=0
                DO 3 I=K,N
                DO 3 J=K,N
                IF (X-ABS(A(I,J))) 4,3,3
           4    X=ABS(A(I,J))
                ISAV=I
                JSAV=J
           3    CONTINUE
          71    IF(ISAV-K)   7,6,7
           7    DO 8 J=1,N
                X=A(K,J)
                A(K,J)=A(ISAV,J)
           8    A(ISAV,J)=X
           6    IF(JSAV-K)   11,10,11
          11    DO 12 I=1,N
                X=A(I,K)
                A(I,K)=A(I,JSAV)
          12    A(I,JSAV)=X
          10    Y=A(K,K)
                W1(K)=ISAV
                W2(K)=JSAV
                FLIPS=FLIPS+ISAV-K+JSAV-K
                C=Y*D
                IF(Y) 70,777,70
          70    A(K,K)=1.0D0/Y
                DO 50 I=1,N
                IF(I-K)14,50,14
          14    Z=A(I,K)
                Z=-Z/Y
                A(I,K)=Z
                DO 18 J=1,N
                IF(J-K)17,18,17
          17    U=A(K,J)
                T=A(I,J)
                A(I,J)=T+U*Z
          18    CONTINUE
          50    CONTINUE
                DO 19 J=1,N
                IF(J-K)20,19,20
          20    Z=A(K,J)
                A(K,J)=Z/Y
          19    CONTINUE
           1    CONTINUE
                DO 26 L=1,N
                K=N-L+1
                IF(W2(K)-K)   23,22,23
          23    M=W2(K)
                DO 24 J=1,N
                X=A(K,J)
                A(K,J)=A(M,J)
          24    A(M,J)=X
          22    IF(W1(K)-K)   27,26,27
          27    M=W1(K)
```

```
SUBROUTINE INV

80              DO 28 I=1,N
                X=A(I,K)
                A(I,K)=A(I,M)
           28   A(I,M)=X
           25   CONTINUE
85              FLIPS=FLIPS-2*(FLIPS/2)
                IF(FLIPS)    500,777,500
          500   C=-D
          777   C=D
                RETURN
                END
```

SU3FCUTINE MPYRT

```
                    SUBROUTINE MPYPT(A,3,C,P,Q,MSA,MSB,P)
          C
          C         CALL MPYRT(A,3,C,MA,NA,MSA,MSB,MB)
          C
 5        C         A      AN MA BY NA ARRAY
          C         B      AN MB BY NA ARRAY
          C         C      AN ARRAY TO CONTAIN THE MATRIX PRODUCT OF A TIMES B TRANS
          C         MA     NUMBER OF ROWS IN A
          C         NA     NUMBER OF COLUMNS IN A
10        C         MSA    MODE OF STORAGE OF A
          C         MSB    MODE OF STORAGE OF B
          C         MB     NUMBER OF ROWS IN B
          C
          C         CALCULATE THE MATRIX PRODUCT OF A TIMES B TRANSPOSE AND STORE
15
                    INTEGER P,Q,R
                    DIMENSION A(P,Q),B(R,Q),C(P,R)
                    IF (MSA .EQ. 1) GO TO 200
                    IF (MSB .EQ. 3) GO TO 300
20                  DO 100 I=1,R
                    DO 100 J=1,P
                    X=0.
                    DO 101 K=1,Q
           101      X=X+A(J,K)*B(I,K)
25         100      C(J,I)=X
                    RETURN
           200      II=0
                    DO 205 I=1,P
                    DO 201 J=1,R
30                  X=0.
                    IC=II
                    DO 202 M=1,I
                    IC=IC+1
           202      X=X+B(J,M)*A(IC,1)
35                  IF (I .EQ. Q) GO TO 299
                    L=I+1
                    IC=IC+I
                    DO 203 M=L,Q
                    X=X+B(J,M)*A(IC,1)
40         203      IC=IC+M
           299      C(I,J)=X
           201      CONTINUE
           205      II=II+I
                    RETURN
45         300      II=0
                    DO 301 I=1,R
                    DO 302 J=1,P
                    IC=II
                    X=0.
50                  DO 303 K=1,I
                    IC=IC+1
           303      X=X+A(J,K)*B(IC,1)
           302      C(J,I)=X
           301      II=II+I
55                  RETURN
                    END
```

```
SUBROUTINE MPY

                     SUBROUTINE MPY(A,B,C,P,Q,MSA,MSB,R)
              C
              C         CALL MPY(A,B,C,MA,NA,MSA,MSB,NB)
              C
 5            C         A      AN MA BY NA MATRIX
              C         B      AN NA BY NB MATRIX
              C         C      AN ARRAY TO CONTAIN THE MATRIX PRODUCT OF A TIMES B
              C         MA     NUMBER OF ROWS IN A
              C         NA     NUMBER OF COLUMNS IN A
10            C         MSA    MODE OF STORAGE OF A
              C         MSB    MODE OF STORAGE OF B
              C         NB     NUMBER OF COLUMNS OF B
              C
              C         CALCULATE THE MATRIX PRODUCT OF A TIMES B AND STORE IN C
15            C
                     INTEGER P,Q,R,P1
                     DIMENSION A(P,Q),B(Q,R),C(P,R)
                     I=1+MSA+4*MSB
                     GO TO (100,200,300,400,500,600,700,800,900,1000,1100,1200,1300,
20                  11400,1500,1600),I
              100    DO 101 I=1,R
                     DO 101 J=1,P
                     X=0.
                     DO 102 K=1,Q
25            102    X=X+A(J,K)*B(K,I)
              101    C(J,I)=X
                     RETURN
              200    DO 205 I=1,R
                     JJ=0
30                   DO 201 J=1,C
                     X=0.
                     JC=JJ
                     DO 203 M=1,J
                     JC=JC+1
35            203    X=X+A(JC,1)*B(M,I)
                     IF (J .EQ. Q) GO TO 299
                     L=J+1
                     JC=JC+J
                     DO 204 M=L,Q
40                   X=X+A(JC,1)*B(M,I)
                     JC=JC+M
              299    C(J,I)=X
              201    JJ=JJ+J
              205    CONTINUE
45                   RETURN
              300    DO 301 I=1,R
                     DO 301 J=1,P
              301    C(J,I)=B(J,I)*A(J,1)
                     RETURN
50            400    DO 401 I=1,R
                     JJ=0
                     DO 402 J=1,Q
                     X=0.
                     JC=JJ
55                   DO 403 M=1,J
                     JC=JC+1
              403    X=X+A(JC,1)*B(M,I)
                     C(J,I)=X
              402    JJ=JJ+J
60                   RETURN
              401    CONTINUE
              500    II=0
                     DO 505 I=1,R
                     DO 501 J=1,P
65                   X=0.
                     IC=II
                     DO 502 M=1,I
                     IC=IC+1
              502    X=X+A(J,M)*B(IC,1)
70                   IF (I .EQ. Q) GO TO 599
                     L=I+1
                     IC=IC+I
                     DO 503 M=L,Q
                     X=X+A(J,M)*B(IC,1)
75            503    IC=IC+M
              599    C(J,I)=X
              501    CONTINUE
              505    II=II+I
```

```
SUBROUTINE MPY

                    RETURN
80          600     II=0
                    K1=P-1
                    DO 601 I=1,P
                    JJ=0
                    DO 602 J=1,P
85                  X=0.
                    IC=II
                    JC=JJ
                    IF (I-J) 625,651,675
            625     DO 626 M=1,I
90                  IC=IC+1
                    JC=JC+1
            626     X=X+A(IC,1)*B(JC,1)
                    L=J-1
                    DO 627 M=I,L
95                  JC=JC+1
                    IC=IC+M
            627     X=X+A(IC,1)*B(JC,1)
                    IF (J .EQ. P) GO TO 699
                    DO 629 M=J,P1
100                 IC=IC+M
                    JC=JC+M
            629     X=X+A(IC,1)*B(JC,1)
                    GO TO 699
            675     DO 651 M=1,J
105
                    IC=IC+1
                    JC=JC+1
            651     X=X+A(IC,1)*B(JC,1)
                    L=I-1
110                 DO 652 M=J,L
                    IC=IC+1
                    JC=JC+M
            652     X=X+A(IC,1)*B(JC,1)
                    IF (I .EQ. P) GO TO 699
115                 DO 654 M=I,P1
                    IC=IC+M
                    JC=JC+M
            654     X=X+A(IC,1)*B(JC,1)
                    GO TO 699
120         650     DO 676 M=1,I
                    IC=IC+1
            676     X=X+A(IC,1)*B(IC,1)
                    IF (I .EQ. P) GO TO 699
                    DO 678 M=I,P1
125                 IC=IC+M
            678     X=X+A(IC,1)*B(IC,1)
            699     C(I,J)=X
            602     JJ=JJ+J
            601     II=II+I
130                 RETURN
            900     DO 901 I=1,Q
                    DO 901 J=1,Q
            901     C(J,I)=A(J,I)*B(I,1)
                    RETURN
135         1100    DO 1101 I=1,P
            1101    C(I,1)=A(I,1)*B(I,1)
                    RETURN
            1300    II=0
                    DO 1301 I=1,P
140                 II=II+I
                    DO 1302 J=1,P
                    X=0.
                    IC=II
                    DO 1303 M=I,Q
145                 X=X+A(J,M)*B(IC,1)
            1303    IC=IC+M
            1302    C(I,I)=X
            1301    CONTINUE
                    RETURN
150         700     CONTINUE
            800     CONTINUE
            1000    CONTINUE
            1200    CONTINUE
            1400    CONTINUE
155         1500    CONTINUE
            1600    CONTINUE
                    PAUSE
                    RETURN
                    END
```

SUBROUTINE MPYTR

```
              SUBROUTINE MPYTR(A,B,C,P,Q,MSA,MSB,R)
      C
      C         CALL MPYTR(A,B,C,MA,NA,MSA,MSB,NB)
      C
      C         A      AN MA BY NA ARRAY
      C         B      AN MA BY NB ARRAY
      C         C      AN ARRAY TO CONTAIN THE MATRIX PRODUCT OF A TRANSPOSE TIM
      C         MA     NUMBER OF ROWS IN A
      C         NA     NUMBER OF COLUMNS IN A
      C         MSA    MODE OF STORAGE OF A
      C         MSB    MODE OF STORAGE OF B
      C         NB     NUMBER OF COLUMNS OF B
      C
      C         CALCULATE THE MATRIX PRODUCT OF A TRANSPOSE TIMES B AND STORE
      C
              INTEGER P,Q,R
              DIMENSION A(P,Q),B(P,R),C(Q,R)
              IF (MSB .EQ. 1) GO TO 200
              IF (MSA .EQ. 3) GO TO 300
              DO 100 I=1,Q
              DO 100 J=1,Q
              X=0.
              DO 101 K=1,P
  101         X=X+A(K,J)*B(K,I)
  100         C(J,I)=X
              RETURN
  200         II=0
              DO 205 I=1,R
              DO 201 J=1,Q
              X=0.
              IC=II
              DO 202 M=1,I
              IC=IC+1
  202         X=X+A(M,J)*B(IC,1)
              IF (I .EQ. R) GO TO 299
              L=I+1
              IC=IC+I
              DO 203 M=L,P
              X=X+A(M,J)*B(IC,1)
  203         IC=IC+M
  299         C(J,I)=X
  201         CONTINUE
  205         II=II+I
              RETURN
  300         II=0
              DO 301 I=1,P
              II=II+I
              DO 301 J=1,P
              IC=II
              X=0.
              DO 302 K=I,Q
              X=X+A(IC,1)*B(K,J)
  302         IC=IC+K
  301         C(I,J)=X
              RETURN
              END
```

```
SUBROUTINE  UNPAK

             SUBROUTINE UNPAK(A,B,MMA,M,N)
             DIMENSION B(MMA,N),A(M,N)
             LL = M + 1
             LT = N + 1
5            DO 1 J=1,N
             JL = LT - J
             DO 1 I=1,M
             IL = LL - I
          1  B(IL,JL) = A(IL,JL)
10           RETURN
             END
```
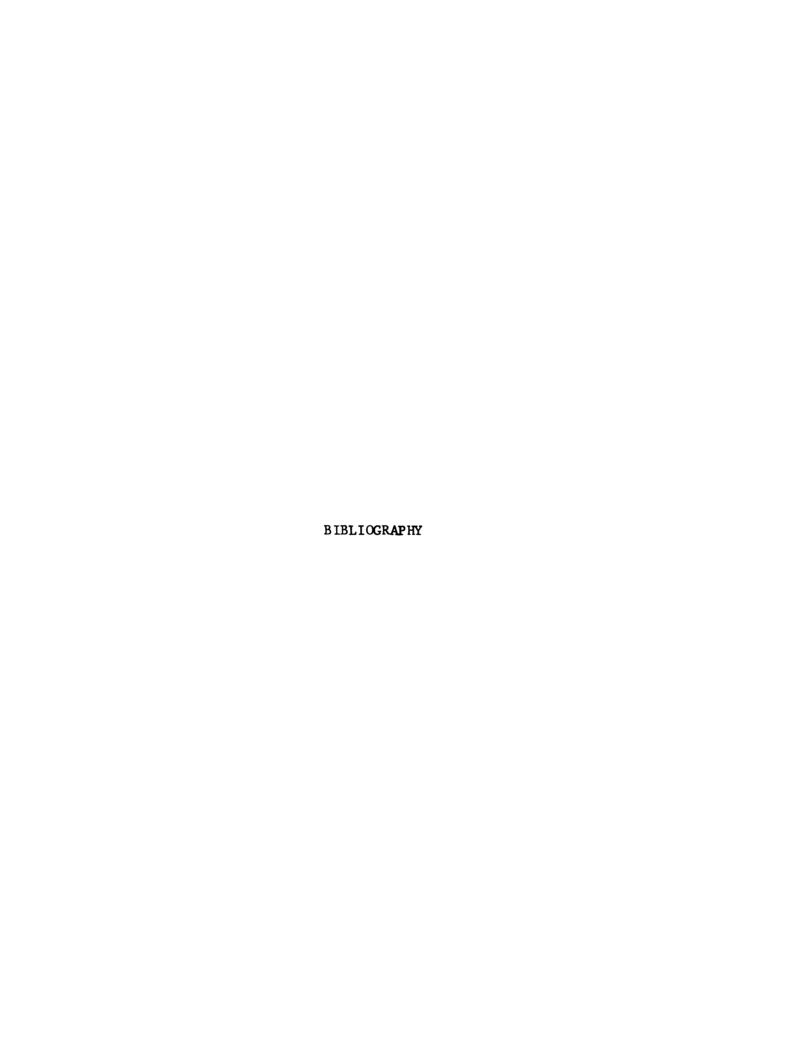
SUBROUTINE PAK

```
      SUBROUTINE PAK(A,B,MMA,M,N)
      DIMENSION A(MMA,N)
      DIMENSION B(M,N)
      DO 10 J = 1,N
      DO 10 I = 1,M
   10 B(I,J) = A(I,J)
      RETURN
      END
```

5

BIBLIOGRAPHY

BIBLIOGRAPHY

Bennett, B. Note on χ^2 tests for matched samples. Journal of the Royal Statistical Society (Series B), 1968, 30, 368-370.

Berkson, J. Maximum likelihood and minimum χ^2 estimates of the logistic function. Journal of the American Statistical Association, 1955, 50, 130-162.

_____. Application of minimum logit χ^2 estimates to a problem of Grizzle with a notation on the problem of no interaction. Biometrics, 1968, 24, 75-95.

_____. Minimum discrimination information, the 'no interaction' problem and the logistic function. Biometrics, 1972, 28, 443-468.

Bhapkar, V. Some tests for categorical data. Annals of Mathematical Statistics, 1961, 32, 72-83.

_____. Notes on analysis of categorical data. Institute of Statistics Mimeo Series No. 477 University of North Carolina, 1966.

Bishop, Y., Fienberg, S., and Holland, P. with the collaboration of Light, R. and Mosteller, G. Discrete Multivariate Analysis: Theory and Practice. Cambridge, Mass.: MIT Press, 1974.

Cochran, W. Some methods for strengthening the common χ^2 tests. Biometrics, 1954, 10, 417-451.

Cramér, H. Mathematical Methods of Statistics. Princeton, N.J.: Princeton University Press, 1946.

Fisher, R. Statistical Methods for Research Workers, 8th edition. Edinburgh: Oliver and Boyd, Ltd., 1941.

Good, I., Gover, T., and Mitchell, G. Exact distributions for X^2 and for the likelihood-ratio statistic for the equiprobable multinomial distribution. Journal of the American Statistical Association, 1970, 65, 267-283.

Goodman, L. Simultaneous confidence intervals for contrasts among multinomial populations. Annals of Mathematical Statistics, 1964, 35, 716-725.

Grizzle, J., Starmer, F., and Koch, G. Analysis of categorical data by linear models. *Biometrics*, 1969, 25, 486-503.

Haynam, G., Govindarajulu, Z., and Leone, F. Tables of the cumulative non-central chi-square distribution. *Selected Tables in Mathematical Statistics Vol. I.* Edited by Institute of Mathematical Statistics. Chicago: Markham Publishing Co., 1970.

Ireland, C., Ku , H., and Kullback, S. Symmetry and marginal homogeneity of an r X r contingency table. *Journal of the American Statistical Association*, 1969, 64, 1323-1331.

Koch, G. and Reinfurt, D. The analysis of categorical data from mixed models. *Biometrics*, 1971, 27, 157-173.

Koch, G., Freeman, J., Freeman, D. and Lehnen, Robert G. A general methodology for the analysis of experiments with repeated measurement of categorical data. *Institute of Statistics Mimeo Series No. 961.* University of North Carolina, 1974.

Kullback, S. *Information Theory and Statistics*. New York: Dover Publications, Inc., 1968.

Laumann, E. *Bonds of Pluralism: The Form and Substance of Urban Social Networks*. New York: John Wiley and Sons, Inc., 1973.

Lewontin, R. and Felsenstein, J. The robustness of homogeneity tests in 2 X N tables. *Biometrics*, 1965, 21, 19-33.

Madansky, A. Tests of homogeneity for correlated samples. *Journal of the American Statistical Association*, 1963, 58, 97-119.

Marascuilo, L. and McSweeney, M. The critics of school integration: opponents of specific programs or foes of integrated education. *Urban Education*, 1969, 4, 79-91.

Margolin, B. and Light, R. An analysis of variance for categorical data, II: Small sample comparisons with chi square and other competitors. *Journal of the American Statistical Association*, 1974, 69, 755-764.

Marsaglia, G. Generating discrete random variables in a computer. *Communications of the Association for Computing Machinary*, 1963, 6, 37-38.

Maxwell, A. Analysing Qualitative Data. London: Methuen, 1961.

Miller, D., Cohen, L., and Hill, K. A methodological investigation of Piaget's theory of object concept development in the sensory-motor period. *Journal of Experimental Child Psychology*, 1970, 9, 59-85.

Mitra, S. On the limiting power function of the frequency chi-square test. *Annals of Mathematical Statistics*, 1958, 29, 1221-1233.

Naylor, T. Computer Simulation Techniques. New York: John Wiley and Sons, Inc., 1968.

Neyman, J. Contributions to the theory of the χ^2 test. Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press, 1949.

Neyman J. and Pearson, E. On the use and interpretation of certain test criteria for the purposes of statistical inference. Biometrika 20A, 1928, 263-294.

Norman, J. and Cannon, L. A computer program for the generation of random variables from discrete distributions. Journal of Statistical Computation and Simulation, 1972, 1, 331-348.

Odoroff, C. A comparison of minimum logit chi-square estimation and maximum likelihood estimation in $2 \times 2 \times 2$ and $3 \times 2 \times 2$ contingency tables: tests for interaction. Journal of the American Statistical Association, 1970, 65, 1617-1631.

Rao, C. Linear Statistical Inference and Its Applications. New York: John Wiley and Sons, Inc., 1965.

Roscoe, J. and Byars, J. Some notes on the chi-square approximation of the multinomial and selected alternatives. Paper presented at the American Educational Research Association Convention, Minneapolis, 1970.

_____. An investigation of the restraints with respect to sample size commonly imposed on the use of the chi-square statistic. Journal of the American Statistical Association, 1971, 66, 755-759.

Scheffé, H. The Analysis of Variance. New York: John Wiley and Sons, Inc., 1959.

Slakter, M. Comparative validity of the chi-square and two modified chi-square goodness-of-fit tests for small but equal expected frequencies. Biometrika, 1966, 53, 619-623.

Stuart, A. A test for homogeneity of the marginal distributions in a two-way classification. Biometrika, 1955, 42, 412-416.

Taylor, W. Distance functions and regularly best asymptotically normal estimates. Annals of Mathematical Statistics, 1953, 24, 85-92.

Tate, M. and Hyer, L. Inaccuracy of the X^2 test of goodness of fit when expected cell frequencies are small. Journal of the American Statistical Association, 1973, 68, 836-841.

Wise, M. Multinomial probabilities and the χ^2 and X^2 distributions. _Biometrika_, 1963, 50, 145-154.

Yarnold, J. The minimum expectation in X^2 goodness of fit tests and the accuracy of approximations for the null distribution. _Journal of the American Statistical Association_, 1970, 65, 864-886.

Yoshinaga, M. Teachers' judgment in the area of classroom management. Unpublished doctoral dissertation, Michigan State University, 1974.

General References

Bennett, B. Tests of hypothesis concerning matched samples. Journal of the Royal Statistical Society (Series B), 1967, 29, 468-474.

Fleiss, J. and Everitt, B. Comparing the marginal totals of square contingency tables. British Journal of Mathematical and Statistical Psychology, 1971, 24, 117-123.

Gold, R. Tests auxiliary to χ^2 tests in a Markov chain. Annals of Mathematical Statistics, 1963, 34, 56-74.

Goodman, L. On simultaneous confidence intervals for multinomial proportions. Technometrics, 1965, 7, 247-254.

Knuth, D. Seminumerical Algorithms: The Art of Computer Programming, Reading, Mass.: Addison-Wesley Publishing Co., 1968.

Marascuilo, L. Large-sample multiple comparisons. Psychological Bulletin, 1966, 65, 280-290.

Maxwell, A. Comparing the classification of subjects by two independent judges. British Journal of Psychiatry, 1970, 116, 651-655.

Miller, R. Simultaneous Statistical Methods. New York: McGraw-Hill Book Co., 1966.

Morrison, D. Multivariate Statistical Methods. New York: McGraw-Hill Book Co., 1967.

Quesenberry, C. and Hurst, D. Large sample simultaneous confidence intervals for multinomial proportions. Technometrics, 1964, 6, 191-195.

Schmidt, W. Matrix operations subroutines for statistical computation -- an adaption of the matrix package of R.D. Bock and Allan Peterson for use on the Michigan State University CDC 6500. Occasional Paper No. 13, Office of Research Consultation, College of Education, Michigan State University, 1970.

Ury, H. and Wiggins, A. Large sample and other multiple comparisons among means. British Journal of Mathematical and Statistical Psychology, 1971, 24, 174-194.

Wald, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society, 1943, 54, 426-482.