

The Preserve: Lehigh Library Digital Collections

Accuracy Of The Chi-square Approximation For 2x3 Contingency Tables With Small Expectations.

Citation

MARCH, DAVID L. *Accuracy Of The Chi-Square Approximation For 2x3 Contingency Tables With Small Expectations.* 1970, https://preserve.lehigh.edu/lehigh-scholarship/graduate-publications-theses-dissertations/theses-dissertations/accuracy-chi-0.

Find more at https://preserve.lehigh.edu/

This document is brought to you for free and open access by Lehigh Preserve. It has been accepted for inclusion by an authorized administrator of Lehigh Preserve. For more information, please contact preserve@lehigh.edu.

71-10,519

MARCH, David L., 1942-
ACCURACY OF THE CHI-SQUARE APPROXIMATION FOR 2X3 CONTINGENCY TABLES WITH SMALL EXPECTATIONS.

Lehigh University, Ed.D., 1970
Statistics

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

ACCURACY OF THE CHI-SQUARE APPROXIMATION FOR 2X3 CONTINGENCY TABLES WITH SMALL EXPECTATIONS

by

David L. March

A Dissertation

Presented to the Graduate Committee

of Lehigh University

in candidacy for the Degree of

Doctor of Education

in the

School of Education

Lehigh University

1970

The work presented or reported herein was performed pursuant to a grant from the U. S. Office of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the U. S. Office of Education, and no official endorsement by the U. S. Office of Education should be inferred.

Approved and recommended for acceptance as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Education.

9-8-70
(date)

Merle W Tate
Professor in Charge

Accepted, 9-23-70
(date)

Special committee directing the doctoral work of David March

Merle W Tate
Chairman

Roy C. Herrendeen

Paul Van R. Miller

ACKNOWLEDGMENTS

The author is indebted to Dr. Merle W. Tate for introducing this problem and for his valuable comments and suggestions throughout the study, to Lehigh University for its generous contribution of computer time, and to Mrs. Carole Whittlinger for her rapid and expert typing of the rough and final drafts of this dissertation.

TABLE OF CONTENTS

CHAPTER		PAGE
I	EXACT AND CHI-SQUARE TESTS FOR CONTINGENCY TABLES	3
	Introduction	3
	Exact Tests for Contingency Tables	4
	Fisher's exact test for the 2 X 2 table	4
	An exact test for an R X C contingency table	5
	The assumption of fixed marginal totals	7
	The Chi-square Test for Contingency Tables	8
	Historical Background	8
	Chi-square test for R X C contingency tables	9
	The assumption of fixed marginal totals	9
	Minimum expectations for the chi-square test	10
	Corrections and modifications for X^2	14
	Purpose of this Study	14
II	COMPUTATIONAL PROCEDURES	16
	Introduction	16
	Stage 1: Generation of the Marginal Totals	17
	Generation of the row totals and column totals	17
	Formation of the sets of marginal totals	18

CHAPTER		PAGE
	Stage 2: Computation of Chi-square and Cumulative Exact Probabilities	18
	Stage 3: Summarization of the Results	21
	Program Validity	22
	Machine Precision	23
III	RESULTS	24
	Introduction	24
	Accuracy of the Chi-square Approximation	24
	Patterns of Overestimates and Underestimates	25
	Extent of Agreement Between Chi-square and Exact Probabilities	29
IV	SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS	32
	Summary	32
	Exact tests for contingency tables	32
	The chi-square test for contingency tables	32
	Purpose of this study	34
	Computational procedures	34
	Presentation of results	35
	Conclusions	35
	A Suggestion for Further Research	36
APPENDIX A	FORTRAN IV PROGRAMS FOR COMPUTATIONS	38
APPENDIX B	SUMMARY REPORTS	67
APPENDIX C	SAMPLE PRINTOUTS OF STAGE 2 COMPUTATION	94
BIBLIOGRAPHY		97

LIST OF TABLES

TABLE		PAGE
I	A generalized R X C contingency table	3
II	Means of absolute percentage errors and numbers of underestimates and overestimates in selected regions for the chi-square approximation to the exact cumulative probabilities for 2 X 3 contingency tables.	26
III	Ratios of underestimates to overestimates in selected regions for the chi-square approximation to the exact cumulative probabilities for 2 X 3 contingency tables with N=12, 15, 18, 21, 24, 27, 30, 36, 42.	27
IV	Patterns of underestimates (-), overestimates (+), and equality (=) for the chi-square approximation to exact probabilities in 14 distributions for 2 X 3 contingency tables with N=12.	28
V	Percentages and numbers of times that the chi-square probability fell in a region below, the same as, and above that for the corresponding exact probability in the regions $P \leq .01$, $.01 < P \leq .05$, $.05 < P \leq .10$ and $.10 < P$ for 2 X 3 contingency tables	29
VI	Percentages and numbers of times that the chi-square probability fell in a region below, the same as, and above that for the corresponding exact probability in the regions $P \leq .05$ and $.05 < P$ for 2 X 3 contingency tables.	30

ACCURACY OF THE CHI-SQUARE APPROXIMATION FOR 2X3 CONTINGENCY TABLES WITH SMALL EXPECTATIONS

by

David L. March

An Abstract

ABSTRACT

The familiar X^2 test of independence for contingency data has wide appeal because it is simpler to use than the available exact tests. However, when the expected frequencies are "small", theory indicates that the X^2 test might yield poor approximations to the exact probabilities. The purpose of this study was to investigate the accuracy of the chi-square probabilities of X^2 when compared with exact cumulative probabilities for 2 X 3 tables with fixed marginal totals and expectations ranging from 1 to not less than 5.

Under the restriction that none of the expected frequencies should be less than one, 2,845 different sets of marginal totals were generated for sample sizes from 6 through 30. In addition, 878 sets for a sample size of 36 and 1,516 sets for a sample size of 42 were generated. Exact cumulative probabilities and the corresponding chi-square probabilities of X^2 were calculated for all possible 2 X 3 contingency tables that could be generated subject to the conditions imposed by each set of marginal totals. These probabilities were compared in terms of their absolute percentage differences and on their extent of agreement in the conventional regions of significance. The numbers of times that chi-square underestimated and overestimated the exact probabilities were also obtained.

These comparisons show that the X^2 approximation steadily improves as the sample size, and therefore, the average expected frequency, increases. However, if close approximations to the exact probabilities

are desired, the chi-square test is at times poor. On the other hand, the test seems to be quite satisfactory if one is only interested in accepting or rejecting some hypothesis at a specified level of significance. In fact, the X^2 test leads to the same decision as the exact probability at the .050 level better than 95 percent of the time if the average expectation is 3 or larger and none of the expectations are less than 1.

Numbers of underestimates and overestimates indicate that the X^2 test is biased toward overestimation in the .005-.010 probability region and toward underestimation in the .010-.200 region. The relative amount of underestimation decreases as the sample size increases.

CHAPTER I

EXACT AND CHI-SQUARE TESTS FOR CONTINGENCY TABLES

Introduction

If each member of a sample of size N is classified by one characteristic into R classes, say $A_1 \ldots A_i \ldots A_R$, and by a second characteristic into C classes, say $B_1 \ldots B_j \ldots B_C$, the data may be arranged in a two way table with R rows and C columns. The entry in each cell of the table, say x_{ij}, is then the number of members of the sample that fall under characteristic A_i and characteristic B_j. Such a table, illustrated in Table I, is known as an R X C "contingency table".

TABLE I

A generalized R X C contingency table.

	B_1 ... B_j ... B_C	Totals
A_1	x_{11} ... x_{1j} ... x_{1C}	r_1
..	... ...	..
A_i	x_{i1} ... x_{ij} ... x_{iC}	r_i
..	... ...	..
A_R	x_{R1} ... x_{Rj} ... x_{RC}	r_R
Totals	c_1 ... c_j ... c_C	N

Cochran (1952) points out that there are at least three distinct experimental situations in which data for a contingency table can occur.

(a) A random sample of size N is drawn from some population and each member is classified into one of the R X C cells. The null hypothesis is that the two classifications, A and B, are "independent." None of the marginal totals is considered fixed in repeated sampling.

(b) R independent random samples of sizes $r_1..r_i..r_R$ are drawn from populations $A_1..A_i..A_R$. The null hypothesis is that with respect to characteristic B all of the samples are drawn from the same population.[1] In repeated sampling, the marginal row totals are considered to be fixed.

(c) The third situation arises if the design of the experiment is such that both sets of marginal totals are regarded as fixed in repeated sampling; that is, the sample is one of an infinite number of possible samples with the same marginal totals. The null hypothesis is that the two classifications, A and B, are independent.

The application of the significance tests presented below to these experimental situations will be discussed later.

Exact Tests for Contingency Tables

Fisher's exact test for the 2 X 2 table. The best known exact test of significance on contingency data is that developed by R. A. Fisher

[1] This is commonly referred to in statistical text books as a test of homogeneity.

for the 2 X 2 table. Since this test is a special case of the general method discussed in the next section, it will not be described in detail here. Examples and explanations of its use can be found in many of the statistical texts such as Ferguson (1966), Tate and Clelland (1957), and Maxwell (1961). Leslie (1955) illustrates a simple method for calculating the exact probability using binomial coefficients.

Finney (1948) used Fisher's method to tabulate significance values for 2 X 2 tables with marginal totals less than or equal to 15. This table has been extended by Latscha (1953) to marginal totals less than or equal to 20. Armsen (1955) gives tables which are applicable for sample sizes less than or equal to 50.

An exact test for an R X C contingency table. Freeman and Halton (1951) derive a general method for computing exact probabilities for contingency tables that result if a sample is subjected to k different and independent classifications. The discussion here will be limited to the case where k = 2.

If a sample of size N is subjected to two different and independent classifications, A and B, with R and C classes respectively, the probability P_X of obtaining the observed array of cell frequencies $X(x_{ij})$, under the conditions imposed by the arrays of marginal totals $A(r_i)$ and $B(c_j)$ is given by

$$P_X = \frac{\prod_{i=1}^{R} (r_i!) \prod_{j=1}^{C} (c_j!)}{N! \prod_{i=1}^{R} \prod_{j=1}^{C} (x_{ij}!)} \qquad (1)$$

Expression (1) is exact and holds if (a) the parent population is infinite or the sampling is done with replacement of the sampled items, (b) the sampling is random, (c) the population is homogeneous, and (d) the marginal totals are considered fixed in repeated sampling. Simply stated, P_X is the exact probability of obtaining the array of cell frequencies in a given R X C contingency table with fixed marginal totals.[1]

For tests of significance, the probability of obtaining an array as probable as, or less probable than, the observed array is needed. The probability is found as follows:

(a) All possible arrays of cell frequencies, subject to the conditions imposed by the fixed marginal totals, are written out.

(b) The corresponding probability for each of these arrays is calculated by means of expression (1).[2]

(c) The probability for the observed array is noted.

(d) The desired probability is then found by summing all of the probability values found in (b) that are less than or equal to the probability value for the observed array noted in (c).

The calculations can be simplified by the use of logarithms or tables of binomial coefficients and, unless an exact probability value is desired, it is often the case that the null hypothesis can be accepted or rejected without fully completing the computations.

---

[1] If R = C = 2, expression (1) becomes the familiar result for Fisher's exact test of the 2 X 2 table.

[2] The sum of all of these probabilities will, except for round-off error, equal 1.0.

However, it should not be difficult to see that this method can still require a great deal of tedious computational work, especially for the larger tables.

Numerical examples which illustrate the method and the various computational shortcuts are given by Freeman and Halton. An additional numerical example can be found in Maxwell (1961, p.46).

Leyton (1968) gives a shortcut method using binomial coefficients for finding the probabilities of all possible arrays of cell frequencies for a 2 X 3 table with given marginal totals. After completion of all of the computations, the exact test is easily performed since arrays more improbable than the observed are easily recognized. The method, an extension of that given by Leslie (1955) for the 2 X 2 table, can be extended to 2 X n tables and, with the use of trinomial coefficients, to 3 X n tables.

The assumption of fixed marginal totals. Cochran (1952) derives the exact probability distributions for 2 X 2 tables in each of the three experimental situations previously described. Differences in these distributions occur because of assumptions based on the number of sets of fixed marginal totals.

The exact test of significance described here requires the assumption that all of the marginal totals are fixed in repeated sampling. Thus, the test is technically not applicable to the first two types of contingency data. The question of whether or not it is logically valid to use the same test of significance in all three situations is discussed by Cochran and also by Barnard (1947) and Pearson (1947).

The Chi-square Test for Contingency Tables

Historical Background. In 1900, Karl Pearson proposed that the statistic

$$X^2 = \sum_{i=1}^{k} \frac{(x_i - m_i)^2}{m_i} \qquad (2)$$

be used as a test criterion to determine if the set of observed frequencies x_i is consistent with the set of frequencies m_i expected if some theory or null hypothesis is true. The observed frequencies are obtained by classifying a random sample of size N into k mutually exclusive classes.

In his paper[1], Pearson shows that, as N becomes large, X^2 follows the chi-square distribution with (k-1) degrees of freedom. He makes two assumptions in arriving at this result. The first is that the expectations m_i are known numbers. That is, he assumes that the observed frequencies follow a multinomial distribution with known probabilities p_i so that $m_i = Np_i$. The second assumption is that the observed x_i are normally distributed. It is this assumption which requires that all of the expectations m_i be "large".

It is interesting to note that when Pearson applied the test to 2 X 2 contingency tables he attributed 3 degrees of freedom to X^2. As Cochran (1952) points out, this caused much confusion and controversy until the matter was cleared up by Fisher some 20 years later. It is

[1] Cochran (1952) gives a detailed review of Pearson's 1900 paper and subsequent historical developments.

now known that the correct limiting distribution of X^2 is obtained when the degrees of freedom are reduced to take into consideration the restrictions imposed by the marginal totals.

Chi-square test for R X C contingency tables. When X^2 is calculated for an R X C table, expression (2) becomes

$$X^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{(x_{ij}-m_{ij})^2}{m_{ij}} \tag{3}$$

where the x_{ij} and m_{ij} are the observed and expected cell frequencies respectively. The estimates for the expectations m_{ij} are obtained from the marginal totals as

$$m_{ij} = \frac{r_i c_j}{N} \tag{4}$$

If all the expectations m_{ij} are sufficiently "large", X^2 will follow the chi-square distribution with (R-1)(C-1) degrees of freedom. For this reason, the X^2 test is often referred to as a chi-square test.

The probability of obtaining an array of frequencies as probable as, or less probable than, the observed array is found as follows:

(a) The expectations m_{ij} are calculated using expression (4).

(b) X^2 is computed using expression (3).

(c) The desired probability is obtained by referring X^2 to the chi-square table with (R-1)(C-1) degrees of freedom.

Numerical examples of the chi-square test can be found in most statistical texts such as Tate (1965) and Ferguson (1966).

The assumption of fixed marginal totals. If the marginal totals are fixed in repeated sampling, the expectations m_{ij} obtained from

expression (4) can be considered to be known numbers. Pearson's original result is then applicable as stated. If the marginal totals are not fixed, expression (4) can only be considered to give estimates of the expectations m_{ij}. However, as a result of his discussion of a proof presented by Fisher in 1924, Cochran (1952, p.322) points out that

> ...Any method of estimation that is efficient gives estimates which become, in the limit, identical with the maximum likelihood estimate. Thus the chi-square distribution, with the appropriate reduction in the degrees of freedom, is valid for any efficient method of estimation.

This means that, if the expected frequencies are sufficiently large, the chi-square test will yield satisfactory probabilities even if all of the marginal totals are not fixed in repeated sampling.[1]

Minimum expectations for the chi-square test. Recommendations about the minimum expectation to be used in the chi-square test are varied. A widely published "rule of thumb" is that no cell should have an expected frequency less than 5. Cramer (1946) suggests a minimum of 10 per cell. Tate and Clelland (1957) indicate that the more-than-5 rule is "somewhat too high" when there are 2 or more degrees of freedom and the average expectation is more than 5 per cell.

The most complete set of recommendations is given by Cochran (1954). For the 2 X 2 table, he suggests using Fisher's exact test if N is less than 20 or if N is less than 40 and the smallest expectation is less than 5. When there are 2 or more degrees of freedom in the table, he

[1] Rigorous proofs of the general limiting distribution of X^2 are given by Cramer (1946).

suggests that if 80 percent of the expected frequencies are greater than 5, a minimum expectation of 1 is allowable. He also states (p.420) that "if X^2 has less than 30 degrees of freedom and the minimum expectation is 2 or more, use of the ordinary chi-square tables is usually adequate."

The recommendations given above are, as Cochran notes, somewhat arbitrary due to the lack of comprehensive research in this area. This is particularly true with respect to contingency tables.

Lewontin and Felsenstein (1965) studied the robustness of the chi-square test for 2 X 5 and 2 X 10 tables. They used Monte Carlo methods to obtain 5,000 random samples of cell frequencies for each of 54 sets of fixed marginal totals. The sample sizes studied were 20, 40, and 60 for the 2 X 5 tables and 40, 80 and 120 for the 2 X 10 tables. Three sets of symmetrically distributed column totals were generated for each of these N's. The variances of these sets were 0, 1.6, 3.6. The smallest row totals were obtained as .05N, .25N, and .50N. The means and variances of the observed distributions of X^2 and the values of X^2 corresponding to the .10, .05, and .01 levels for one-sided tests are reported for each of the sets of parameters. As a result of their study, Lewontin and Felsenstein conclude that (p.31) "the 2 X n table can be tested by the conventional chi-square criterion if all of the expectations are 1 or greater." Two things should be emphasized about this study. First, the 27 sets of marginal totals for the 2 X 10 tables gave the same expected cell frequencies as the corresponding sets for the 2 X 5 tables. Second, the symmetrical distribution of the marginal column totals restricted the range of expectations within the cells.

As part of a larger study, Roscoe and Byars (1970) studied the robustness of the chi-square test of independence for 2 X 2 through 5 X 5 tables with fixed and equal row totals of 10, 15, 20, 30, 50 and 100. Their method consisted of taking 10,000 random samples from a uniform and a skew population for each of the 60 table-row total combinations. X^2 was calculated for each sample and compared with the tabled values of chi-square at the .05 and .01 levels. They report the number of rejections in each set of 10,000 samples. Roscoe and Byars conclude (p.11):

> Generally, the chi-square test of independence proved to be strikingly robust. With data drawn from a uniform population, an excellent approximation was achieved with average expected frequencies as low as two. If the data depart markedly from the uniform, there is a tendency for the test to become conservative with small samples.

It should be noted that the generality of their statement is limited by the fact that the row totals were equal for each set of samples.

Both of the studies described so far were sampling experiments. They seem to indicate that the distribution of X^2 is reasonably close to that of chi-square, even if there are expectations as low as 1 or 2. However, they do not provide evidence on the accuracy of the probabilities which result when X^2 is referred to the chi-square table.

Reports of comparisons between chi-square and exact probabilities seem to be limited to the related use of chi-square for testing goodness-of-fit to the multinomial distribution. The most comprehensive study of this kind is reported by Tate and Hyer (1969).

In the Tate and Hyer study, 126 multinomial distributions with equal expectations were constructed. The number of classes varied

from 3 to 7 while the total sample size varied such that the expectations ranged from 1 to not less than 5. The exact cumulative probability was calculated for each outcome. For these distributions, they report the mean absolute percentage errors of the chi-square probabilities and the number of times they underestimated and overestimated the exact probabilities in selected regions of significance. The median of the mean absolute percentage errors over the .005-.205 region is given as 49. In addition to these comparisons, chi-square and exact probabilities for 18 "representative" distributions were cross tabulated to determine to what extent they lead to the same decision in conventional regions of significance. They state (p.2) that:

> ...on the average, the probabilities fell in the same region about two-thirds of the time. Considering only the two regions, .050 or less and .051 or more, the percentage agreement was not less than 76 in any of the 18 distributions. In all but three of the distributions, the percentage was 85 or more.

Tate and Hyer also constructed 36 distributions with 3 classes and unequal expectations. These included distributions originally studied by Neyman and Pearson (1931) and El Shanawany (1936). The means of absolute percentage errors and the number of underestimates and overestimates in the .005-.205 region are reported for these distributions.

In their summary, Tate and Hyer state (p. 2):

> Judging from the multinomial distributions examined in this study, if close approximations to exact probabilities are needed, the X^2 test is not satisfactory when expectations are fewer than 10, and even when they are more than 10 the approximations may at times be poor. On the other hand, if one is interested only in whether the cumulative probability associated with an outcome in a multinomial distribution is less than or greater than .05, the X^2 test performs reasonably well with expectations as small as one.

Corrections and modifications for X^2. Various corrections and modifications have been suggested for the X^2 distribution in order to bring about closer agreement with the distribution of chi-square.[1] The most familiar of these is Yates's correction for continuity. However, as a result of their 1970 study, "the use of the continuity correction in two-by-two tables is not recommended" by Roscoe and Byars (p. 11).

For the chi-square test for goodness-of-fit to the multinomial, Tate and Hyer believe that (1969, P. 20)

> None of the corrections is generally effective; indeed, inspection of the numbers of underestimates and overestimates...suggests that no correction will succeed. There is little if any bias in the chi-square approximation to the multinomial for the values of k and N and the regions considered in this study.

Purpose of this Study

The chi-square test has wide appeal because it is simpler to use than the exact tests. However, when the expected frequencies are "small", theory indicates that the X^2 test might yield a poor approximation to the exact probability. The problem is further compounded by the fact that the various recommendations about minimum expectations are usually somewhat arbitrary; they are not based on extensive research.

The purpose of this study is to investigate the accuracy of chi-square probabilities when compared against exact cumulative probabilities for 2 X 3 tables with small expectations. Of particular interest are comparisons that determine to what extent chi-square probabilities lead to

---

[1] For example, see Wise (1963) and Nass (1959). Slakter (1966, p. 20) presents evidence "showing that Nass's modification does not result in better approximations, although a correction for continuity does appear to have some merit."

the same decision as exact probabilities in conventional regions of significance.

CHAPTER II

COMPUTATIONAL PROCEDURES

Introduction

To fulfill the purpose of this study, chi-square and exact probabilities were compared for 2 X 3 contingency tables with "small" expected frequencies. Expectations less than or equal to 5 were considered to be "small", primarily because of the published recommendations.

The average expected frequency for a contingency table is, of course, equal to the total sample size N divided by the number of cells in the table. This means that for the six celled 2 X 3 tables, an average expectation of 5 occurs when N=30. Consequently, probability comparisons for tables with N less than or equal to 30 were thought to be of greatest importance.

One restriction was imposed. Tables which had one or more expected frequencies less than one were eliminated. This was done for two reasons. First, there is almost universal agreement in the literature that the chi-square test gives poor approximations when an expected frequency is less than one and there are relatively few classes. Second, the restriction substantially reduced the number of tables to be considered.

Even with the above restriction, 2,845 different sets of marginal totals were generated for values of N from 6 through 30. In addition, 878 sets for N=36 and 1,516 sets for N=42 were generated. Exact and chi-square probabilities were calculated for all 2 X 3 contingency tables that could be generated subject to the conditions imposed by

each set of marginal totals. The probabilities were compared based on their absolute percentage differences and on their extent of agreement in the conventional regions of significance. Altogether, approximately 250,000 exact and chi-square probabilities were compared.

Due to the large number of probability values involved, the computations were performed in three stages involving four programs executed on the CDC 6400 digital computer at Lehigh University. These stages are described in the following sections. Program listings are given in Appendix A.

Stage 1: Generation of the Marginal Totals

The first step in the computations was to generate sets of marginal totals for all of the tables to be considered. This was accomplished in two phases.

Generation of the row totals and column totals. In the first phase, sets of row totals for two rows and column totals for three columns were obtained for each N and written to magnetic tape. Two factors limited the number of sets that were generated. First, the computations for the chi-square and exact tests do not depend on the arrangement of the marginal totals. Therefore, permutations of the sets were not considered. Second, if any marginal total is less than or equal to 1, at least one expected frequency will be less than 1. Since the expected frequencies were restricted to be greater than or equal to 1, sets of marginal totals with a 0 or a 1 were eliminated immediately.

The procedure for each N was as follows:

(a) Starting with (N,0), successive two digit sets were generated

to obtain (N,0), (N-1,1),...,(N-i,i),...($\frac{N}{3}$, N-$\frac{N}{3}$) where $\frac{N}{3}$ was truncated (not rounded) to an integer when necessary. As explained above, only combinations with $n-1 \geq i \geq 2$ were accepted as row totals. For example, for N=6 the sets (6,0), (5,1), (4,2), (3,3), and (2,4) were generated. However, only (4,2) and (3,3) were retained as row totals.

(b) Starting with each two digit set (m,n) obtained in (a), successive three digit sets were generated to obtain (m,n,0), (m,n-1,1), ...,(m,n-j,j),...(m,n- $\frac{n}{2}$,$\frac{n}{2}$) where $\frac{n}{2}$ was rounded up to an integer when necessary. Only sets with $m \geq n-j \geq j \geq 2$ were accepted as column totals. For the current example, (6,0,0);(5,1,0);(4,2,0),(4,1,1);(3,3,0), (3,2,1); and (2,4,0),(2,3,1),(2,2,2) were generated. Of these, only (2,2,2) was retained for column totals.

Formation of the sets of marginal totals. In phase two, the sets of row and column totals were read from the tape made during phase one. Each set of row totals was combined with every set of column totals to obtain sets of marginal totals for 2 X 3 contingency tables.

As each set was being formed, the expected cell frequencies were calculated. The set was discarded if one or more of the expectations was less than 1. The remaining sets were written to a second magnetic tape for use as marginal totals in stage 2.

For the case when N=6, the sets (4,2)X(2,2,2) and (3,3)X(2,2,2) were formed. However, since the first set yields an expected frequency less than 1, only (3,3)X(2,2,2) was written to tape.

Stage 2: Computation of Chi-square and Cumulative Exact Probabilities

In the second step of the computations, chi-square and cumulative

exact probabilities were calculated for each and every possible array of "observed" cell frequencies that could be generated for a given set of marginal totals. Although the computer program used for this stage is more general, the procedure is described for 2 X 3 tables only.

Before describing the procedure, it should be noted that expression (1,p.5) for the exact probability of an observed array can be rewritten for the 2 X 3 tables as

$$P_X = Q_X / R_X$$

where $$Q_X = \frac{\prod_{i=1}^{2}(r_i!)\ \prod_{j=1}^{3}(c_j!)}{N!}$$

and $$R_X = \prod_{i=1}^{2}\ \prod_{j=1}^{3}\ (x_{ij}!)$$

Now, for a given set of marginal totals, r_i and c_j, Q_X is a constant and needs to be determined only once for each and every R_X which varies depending upon the array of observed cell frequencies, x_{ij}. In order to avoid machine overflow and prevent loss of accuracy due to roundoff error, the actual computer computations were:

$$\log_{10}P_X = \log_{10}Q_X - \log_{10}R_X \tag{5}$$

where $$\log_{10}Q_X = \sum_{i=1}^{2} \log_{10}(r_i!) + \sum_{j=1}^{3} \log_{10}(c_j!) - \log_{10}(N!) \tag{6}$$

and $$\log_{10}R_X = \sum_{i=1}^{2} \sum_{j=1}^{3} \log_{10}(x_{ij}!) \tag{7}$$

The probability P_X was then obtained by raising 10 to the power $(\log_{10}P_X)$.

It should also be noted that the density function for the chi-square

distribution with exactly 2 degrees of freedom reduces to

$$f(X^2) = \frac{1}{2} e^{-\frac{1}{2}X^2}$$

Thus, for a 2 X 3 table with 2 degrees of freedom, the chi-square probability corresponding to a value for X^2 is quite easily and accurately computed as

$$P(X^2) = \int_0^{X^2} \frac{1}{2} e^{-\frac{1}{2}y} \, dy = e^{-\frac{1}{2}X^2} \qquad (8)$$

The procedure for each set of marginal totals was as follows:

(a) Upon reading a new set of marginal totals from tape, the expected frequencies were calculated using expression (4, p.9) and the logarithm of the constant term, Q_X, was obtained from expression (6).

(b) Next, all possible arrays of "observed" cell frequencies which would give the marginal totals were generated. This process was simplified by the fact that a 2 X 3 table has only two degrees of freedom.

Since the marginal totals were generated in such a manner that the last two columns had the smallest possible totals, initial values of c_2 and c_3 were assigned to "degrees of freedom" cells x_{12} and x_{13} respectively.[1] Successive sets were then generated for these two cells as $(c_2,c_3),(c_2-1,c_3),\ldots(c_2-n,c_3),\ldots(0,c_3)$, followed by $(c_2,c_3-1),(c_2-1,c_3-1),\ldots(c_2-n,c_3-1),\ldots(0,c_3-1)$ and so on until (0,0) was obtained. For example, given (3,3)X(2,2,2) as marginal totals, the sets (2,2),(1,2),(0,2),(2,1),(1,1),(0,1),(2,0),(1,0) and (0,0) were generated for cells (x_{12},x_{13}).

[1] Any two column totals could have been assigned to any two cells in one of the rows. However, the amount of computation is greatly reduced if the smallest column totals are chosen.

Of course, some of these sets did not lead to valid arrays of observed cell frequencies. But, they did lead to all arrays that were valid.

After x_{12} and x_{13} were filled, the remainder of the cell frequencies were obtained by subtraction from the marginal totals. If a negative cell frequency was encountered, the array was discarded. Otherwise, the logarithm of R_X was obtained from expression (7) and finally, the probability, P_X, of the given set of observations was computed by taking the antilogarithm of expression (5). The array of cell frequencies and P_X were saved in memory. Then, new values were obtained for x_{12} and x_{13} and the process was repeated until all possibilities were exhausted.

(c) After completion of part (b), the observations and P_X were sorted from minimum to maximum based on P_X. Obtaining the cumulative probabilities (significance values) was then a simple matter of addition. Finally, values were obtained for X^2 using expression (3, p.9) and the corresponding chi-square probability using expression (8). All of these values were written to magnetic tape for analysis in stage 3.

Stage 3: Summarization of the Results

During the last stage of the computations, exact cumulative probabilities were compared with corresponding chi-square probabilities. The results of these comparisons were summarized for each sample size and summary reports were printed. These reports are given in Appendix B.

Two types of probability comparisons were made. First, the absolute percentage errors of the chi-square probabilities were obtained from

$$\frac{|P(X^2) - P(E)|}{P(E)} \times 100$$

where P(E) is the exact cumulative probability and $P(X^2)$ is the corresponding chi-square probability. The mean percentage error was calculated in the .000-.005, .005-.010, .010-.050, .050-.100, .100-.150, .150-.200, and .200-1.00 probability regions. The number of times that the chi-square probabilities underestimated and overestimated the exact probabilities were also obtained in each of these regions.

Second, the extent of agreement between the exact and chi-square probabilities was determined by classifying each probability pair in a 7 X 7 table with classes equivalent to the seven probability regions described.

Program Validity

In order to check the validity of the computer programs, options were installed in the stage 2 and stage 3 programs to obtain printouts of complete probability distributions for given sets of marginal totals. The results of the computer were then verified by hand calculations. Examples of some stage 2 printouts are given in Appendix C.

One validity check that gave continued confidence in the results was the monitoring of the cumulative probability of obtaining all possible outcomes for a given set of marginal totals. This probability should, of course, be 1.0. Each summary report contains an accuracy-validity check line with the minimum and maximum total cumulative probability encountered among all sets of marginal totals summarized in the report. This print line shows that the exact probability totals were, in all cases, accurate to at least 10 significant digits.

Machine Precision

All computations were performed on the CDC 6400 digital computer at Lehigh University. Due to the large word size of this machine, 14 decimal digits were retained throughout the computations.

CHAPTER III

RESULTS

Introduction

Under the restriction that all of the expected frequencies were greater than or equal to one, 2,845 different sets of marginal totals were generated for values of N from 6 through 30. In addition, 878 and 1,516 sets were generated for Ns of 36 and 42 respectively. Exact and chi-square probabilities were calculated for all 2 X 3 contingency tables that could be generated from these sets of marginal totals. The chi-square probabilities were compared with the corresponding exact probabilities in the .000-.005, .005-.010, .010-.050, .050-.100, .100-.150, .150-.200, and .200-1.00 probability regions.

Three types of comparisons were made for each N. First, the mean absolute percentage error of chi-square was computed in each of the above regions. Second, the number of times that the chi-square test underestimated and overestimated the exact probability was obtained in each region. Finally, the chi-square and exact probability pairs were cross tabulated in a 7 X 7 table with classifications corresponding to the seven probability regions described. Summary reports for these comparisons are given in Appendix B.

Accuracy of the Chi-square Approximation

The means of the absolute percentage errors in selected significance regions for the chi-square approximation to the exact cumulative probability

are given in Table II. As expected, the chi-square approximation improved as the sample size, and therefore, the expected frequency, increased. The improvement was reasonably consistent in all regions except for the region from .005-.010.

Careful inspection of the mean percentage errors indicates that for N less than 19, the best approximations occur in the .010-.050 region. As N increases, there is a tendency for the region of best approximation to shift to the right. For N from 20 to 30, the approximation is best in the .050-.100 region while for an N of 36 or 42, the approximation is best in the .100-.150 region. However, it must be pointed out that this trend may be due to the fact that the absolute percentage errors were computed from

$$\frac{|P(X^2) - P(E)|}{P(E)} \times 100$$

Thus, the percentage error would decrease as P(E) increased even though the magnitudes of the corresponding average absolute errors might be the same.

Patterns of Overestimates and Underestimates

Table II also includes the number of times that chi-square underestimated and overestimated the exact probabilities in the specified significance regions. For the majority of the probability comparisons in the .005-.010 region, chi-square overestimates the exact probability. The converse is true in the remainder of the regions. Furthermore, the ratio of underestimates to overestimates consistently increases as the exact probability increases. Table III illustrates this trend clearly.

TABLE II

Means of absolute percentage errors and numbers of underestimates and overestimates* in selected regions for the chi-square approximation to the exact cumulative probabilities for 2 X 3 contingency tables.**

N	Region .005<P≤.010	.010<P≤.050	.050<P≤.100	.100<P≤.150	.150<P≤.200
8	- - -	35.9%(2/ 0)	12.9%(2/ 0)	- - -	- - -
9	- - -	68.9 (3/ 0)	- - -	70.0%(6/ 0)	- - -
10	15.1%(2/ 0)	30.4 (8/ 2)	56.7 (8/ 0)	5.2 (1/ 0)	50.5%(4/ 0)
11	32.6 (2/ 0)	20.0 (6/ 4)	43.1 (8/ 0)	28.1 (5/ 0)	58.2 (4/ 0)
12	52.3 (3/ 5)	38.1 (31/ 5)	33.7 (22/ 2)	53.7 (15/ 0)	32.0 (3/ 0)
13	9.8 (2/ 3)	22.8 (22/ 6)	39.1 (21/ 1)	37.9 (10/ 1)	32.3 (6/ 1)
14	36.9 (9/ 4)	24.6 (49/ 19)	35.4 (44/ 3)	32.4 (33/ 0)	37.8 (25/ 0)
15	48.8 (17/ 14)	24.3 (57/ 26)	36.4 (60/ 4)	32.8 (32/ 0)	28.8 (20/ 2)
16	26.5 (25/ 25)	22.9 (85/ 52)	30.8 (70/ 6)	31.7 (56/ 5)	34.5 (38/ 0)
17	29.0 (8/ 31)	22.8 (105/ 63)	27.9 (86/ 11)	30.4 (54/ 3)	29.4 (51/ 1)
18	30.2 (28/ 56)	25.5 (201/ 93)	38.7 (159/ 17)	29.6 (90/ 3)	29.1 (77/ 7)
19	32.5 (33/ 68)	23.7 (186/ 120)	24.6 (139/ 18)	24.9 (111/ 9)	26.3 (72/ 5)
20	34.1 (71/ 81)	22.3 (321/ 167)	23.7 (218/ 42)	26.7 (161/ 10)	28.2 (126/ 5)
21	32.4 (61/ 120)	23.6 (318/ 213)	23.7 (238/ 49)	25.7 (179/ 13)	23.2 (121/ 8)
22	28.5 (80/ 167)	22.2 (463/ 295)	21.7 (316/ 72)	23.9 (230/ 22)	25.0 (208/ 15)
23	32.3 (112/ 175)	22.7 (475/ 344)	21.0 (374/ 92)	22.5 (255/ 32)	21.4 (194/ 14)
24	33.0 (158/ 230)	22.9 (708/ 498)	22.7 (536/ 127)	22.6 (370/ 37)	22.9 (285/ 33)
25	30.8 (151/ 277)	22.5 (693/ 560)	19.8 (513/ 136)	21.0 (386/ 40)	20.7 (279/ 30)
26	32.3 (209/ 380)	21.6 (965/ 740)	19.6 (694/ 196)	20.3 (489/ 62)	21.2 (369/ 37)
27	34.5 (245/ 416)	22.9 (1060/ 845)	19.6 (786/ 215)	19.9 (537/ 84)	19.6 (399/ 54)
28	31.1 (316/ 537)	22.0 (1368/1098)	18.7 (961/ 299)	19.7 (668/ 126)	20.0 (546/ 62)
29	34.2 (322/ 620)	22.4 (1419/1219)	17.7 (984/ 334)	18.5 (765/ 125)	18.6 (543/ 75)
30	32.0 (442/ 774)	21.9 (1874/1570)	18.9 (1316/ 437)	19.1 (944/ 186)	19.1 (715/ 89)
36	32.4 (1032/1900)	21.6 (4146/3777)	16.8 (2864/1117)	16.2 (2012/ 465)	17.0 (1584/248)
42	32.3 (2150/3841)	21.3 (7986/7899)	15.6 (5374/2458)	14.9 (3782/1061)	14.8 (2926/614)

* Numbers in parenthesis as (underestimates/overestimates) where overestimates include cases of equality.

** All expected cell frequencies were greater than or equal to one.

TABLE III

Ratios of underestimates to overestimates in selected regions for the chi-square approximation to the exact cumulative probabilities for 2 X 3 contingency tables with N=12, 15, 18, 21, 24, 27, 30, 36, 42.*

	REGION				
N	$.005 < P \leq .010$	$.010 < P \leq .050$	$.050 < P \leq .100$	$.100 < P \leq .150$	$.150 < P \leq .200$
12	0.60	6.20	11.00	-	-
15	1.21	2.19	15.00	-	10.00
18	0.50	2.16	9.40	30.00	10.00
21	0.51	1.49	4.86	13.78	15.12
24	0.69	1.42	4.22	10.00	8.55
27	0.59	1.25	3.68	6.40	7.39
30	0.57	1.19	3.01	5.07	8.04
36	0.54	1.10	2.57	4.32	6.39
42	0.56	1.01	2.19	3.54	4.77

*All expected frequencies were greater than or equal to 1.

These results, for tables with expected frequencies greater than or equal to 1, suggest that there may be a somewhat systematic relationship between the chi-square and exact probabilities.

Some support for the above statement is given in Table IV which shows patterns of underestimation (-), overestimation (+), and equality (=) for the 14 exact distributions generated with an N of 12. In seven of these distributions, the chi-square test consistently underestimated the exact probabilities. It is interesting to note that either the row totals or the column totals are equal in all but one of these. On the other hand, the irregular patterns tend to occur when the row and column totals are unequal. This implies that the distribution of the marginal totals might be used to predict patterns of agreement.

TABLE IV

Patterns of underestimates (-), overestimates (+), and equality (=) for the chi-square approximation to exact probabilities in 14 distributions for 2 X 3 contingency tables with N=12.

Marginal Totals	Ordered Outcomes**
(8,2,2)X(6,6)	--------=
(7,3,2)X(6,6)	----------
(6,4,2)X(6,6)	++++----------=
(6,3,3)X(8,4)	----------=
(7,5)	++----------
(6,6)	++----------
(5,5,2)X(6,6)	++----------=
(5,4,3)X(8,4)	++--++--------
(7,5)	++++----------
(6,6)	++----------
(4,4,4)X(9,3)	---------=
(8,4)	----------
(7,5)	----------
(6,6)	----------=

* All expected frequencies were greater than or equal to 1.

**Ordered from smallest to largest (left to right) by exact cumulative probabilities for the outcome.

Returning to Table III, it is apparent that as N increases, the ratio of underestimates to overestimates decreases in all but the .005-.010 significance region. In that region, the ratio seems to have reached a limit near 0.55.

Extent of Agreement Between Chi-square and Exact Probabilities

In many practical applications of the chi-square test, the actual probability value is used only to determine if the null hypothesis, H_o, should be accepted or rejected at some specified level of significance. Thus, if the chi-square probability leads to the same decision as the

TABLE V

Percentages and numbers of times that the chi-square probability fell in a region below, the same as, and above that for the corresponding exact probability in the regions $P \leq .01$, $.01 < P \leq .05$, $.05 < P \leq .10$ and $.10 < P$ for 2 X 3 contingency tables.

N	$P_r(C) < P_r(E)$**	$P_r(C) = P_r(E)$	$P_r(C) > P_r(E)$
8	31.6% (6)	68.4%(13)	- -
9	27.3 (6)	72.7 (16)	- -
10	19.7 (12)	80.3 (49)	- -
11	27.6 (16)	72.4 (42)	- -
12	21.3 (44)	78.3 (162)	0.4%(1)
13	20.0 (32)	79.4 (127)	0.6 (1)
14	18.1 (72)	80.6 (321)	1.3 (5)
15	16.7 (82)	81.5 (401)	1.8 (9)
16	17.0 (136)	82.0 (658)	1.0 (8)
17	13.4 (117)	85.1 (741)	1.5 (13)
18	11.1 (178)	86.4 (1381)	2.5 (40)
19	10.2 (168)	88.1 (1444)	1.7 (28)
20	9.2 (244)	88.4 (2350)	2.4 (65)
21	8.9 (263)	88.9 (2630)	2.2 (66)
22	8.8 (366)	88.9 (3721)	2.3 (97)
23	7.3 (348)	90.3 (4272)	2.4 (113)
24	8.0 (557)	89.7 (6242)	2.3 (163)
25	6.7 (480)	90.9 (6544)	2.4 (172)
26	6.3 (628)	91.0 (9019)	2.6 (261)
27	6.3 (704)	91.2 (10248)	2.5 (277)
28	5.9 (865)	91.7 (13372)	2.4 (348)
39	5.2 (820)	92.1 (14574)	2.7 (430)
30	5.8 (1216)	91.6 (19288)	2.6 (540)
36	4.3 (2212)	93.1 (48412)	2.6 (1352)
42	3.5 (3903)	94.1 (104743)	2.4 (2679)

*All expected cell frequencies were greater than or equal to 1.

**Read: Chi-square probability region is below exact probability region.

corresponding exact probability, the chi-square test could be used regardless of the actual percentage error.

Table V reports the total percentages and numbers of agreement and disagreement for chi-square probabilities compared to exact probabilities in the four commonly used probability regions between 0.00-.010,

TABLE VI

Percentages and numbers of times that the chi-square probability fell in a region below, the same as, and above that for the corresponding exact probability in the regions $P \leq .05$ and $.05 < P$ for 2 X 3 contingency tables.

N	$P_r(C) < P_r(E)$**	$P_r(C) = P_r(E)$	$P_r(C) > P_r(E)$
8	10.5% (2)	89.5% (17)	- -
9	27.3 (6)	72.7 (16)	- -
10	13.1 (8)	86.9 (53)	- -
11	12.1 (7)	87.9 (51)	- -
12	12.1 (25)	87.9 (182)	- -
13	8.8 (14)	91.2 (146)	- -
14	7.5 (30)	92.2 (367)	0.3% (1)
15	7.9 (39)	92.1 (453)	- -
16	7.1 (57)	92.8 (744)	0.1 (1)
17	4.7 (41)	94.8 (826)	0.5 (4)
18	4.6 (74)	95.1 (1521)	0.3 (4)
19	4.0 (66)	95.6 (1568)	0.4 (6)
20	3.1 (82)	96.3 (2561)	0.6 (16)
21	3.1 (91)	96.5 (2855)	0.4 (13)
22	2.8 (117)	96.6 (4043)	0.6 (24)
23	3.0 (144)	96.4 (4564)	0.6 (25)
24	3.2 (220)	96.3 (6708)	0.5 (34)
25	2.6 (184)	96.7 (6959)	0.7 (53)
26	2.2 (222)	97.2 (9626)	0.6 (60)
27	2.5 (275)	96.9 (10887)	0.6 (67)
28	2.0 (296)	97.4 (14211)	0.5 (78)
29	1.8 (281)	97.4 (15421)	0.8 (122)
30	2.1 (441)	97.3 (20484)	0.6 (119)
36	1.5 (803)	97.8 (50847)	0.6 (326)
42	1.1 (1352)	98.4 (107233)	0.5 (678)

*All expected cell frequencies were greater than or equal to 1.

**Read: Chi-square probability region is below exact probability region.

.010-.050, .050-.100, and .100-1.00. The extent of agreement is quite good and steadily improves as N increases. As expected from the analysis of the ratios of underestimates to overestimates, the majority of the disagreements occur because the chi-square probability is in a region below that of the exact probability. In addition, the percentage of underestimates drops steadily while the percentage of overestimates seems to be reasonably constant at about 2.5.

Inspection of the summary reports in Appendix B indicates that the chi-square disagreements are seldom more than one probability region above or below the exact region. Consequently, the extent of agreement will improve for the case where a null hypothesis is to be accepted or rejected at some specified level of significance. This is illustrated in Table VI which reports the percentages and numbers of agreement and disagreement in the two regions of 0.00-.050 and .050-1.00. The extent of agreement is remarkable, even for tables with average expectations as low as 2. Once again, the majority of the disagreements are caused because the chi-square probability falls in a significance region below that of the exact probability.

CHAPTER IV

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

A frequent application of statistics in the behavorial sciences is that of testing the hypothesis that two variables are independent. It is often the case that the data can be meaningfully arranged in a two-way "contingency" table where the cells contain the observed frequencies.

Exact tests for contingency tables. The best known exact test of significance for contingency data is that developed by R. A. Fisher for the 2 X 2 table. Freeman and Halton (1951) give a general method, of which Fisher's is a special case, for computing exact probabilities for any size contingency table. In their derivation it is assumed that the marginal totals are fixed in repeated sampling.

The required calculations can be simplified by the use of logarithms or tables of binomial coefficients and, unless an exact probability is desired, it is often the case that the null hypothesis can be accepted or rejected without fully completing the computations. However, the method can still require a great deal of tedious computational work, especially for the larger tables.

The chi-square test for contingency tables. A well-known and simpler procedure is to calculate the statistic $X^2 = \sum\sum(x_{ij}-m_{ij})^2/m_{ij}$ where x_{ij} and m_{ij} are the observed and expected cell frequencies respectively. Under the null hypothesis that the two variables are independent within the sample, the expectations m_{ij} can be calculated

from the marginal totals of the contingency table. If none of the expected frequencies are "too small", X^2 will follow the chi-square distribution with (R-1)(C-1) degrees of freedom. The corresponding chi-square probability is then used as an approximation to the exact probability.

Recommendations about the minimum expectations to be allowed in the chi-square test are varied. One popular "rule of thumb" is that no cell should have an expected frequency less than 5. The most complete set of recommendations is given by Cochran (1954) who suggests that if 80 percent of the expectations are greater than 5, a minimum expectation as low as 1 is allowable.

The recommendations given above are somewhat arbitrary due to the lack of comprehensive research in this area, especially with respect to contingency tables. Sampling experiments by Lewontin and Felsenstein (1965) and Roscoe and Byars (1970) indicate that the distribution of X^2 is reasonably close to that of chi-square, even if there are expectations as low as 1 or 2. However, they do not provide evidence on the accuracy of the probabilities which result when X^2 is referred to the chi-square table.

Reports of comparisons between chi-square and exact probabilities seem to be limited to the related use of chi-square for testing goodness-of-fit to the multinomial distribution. The most comprehensive study of this kind is reported by Tate and Hyer (1969). Their results indicate that when the expectations are fewer than 10, the X^2 goodness-of-fit test is not satisfactory if close approximations to the exact probabilities are needed. On the other hand, when one is only interested

in rejecting or accepting some hypothesis at the .05 significance level, the X^2 test performs reasonably well with expectations as low as 1.

Purpose of this study. The purpose of this study was to investigate the accuracy of the chi-square probabilities when compared with exact cumulative probabilities for 2 X 3 contingency tables with small expectations.

Computational procedures. Expectations less than or equal to 5 were considered to be "small", primarily because of the published recommendations. Consequently, probability comparisons for tables with sample sizes less than or equal to 30 were thought to be of greatest importance. One restriction was imposed. Tables which had one or more expected frequencies less than 1 were eliminated.

The computations were performed in three stages involving four programs executed on the CDC 6400 digital computer at Lehigh University. The first stage in the computations was to generate sets of marginal totals for all of the tables to be considered. Even with the restriction that none of the expectations were less than one, 2,845 different sets of marginal totals were generated for values of N from 6 to 30. In addition, 878 sets for N=36 and 1,516 sets for N=42 were generated.

In the second stage of the computations, chi-square and exact cumulative probabilities were calculated for each and every possible array of "observed" cell frequencies that could be generated for a given set of marginal totals. In all, approximately 250,000 probability pairs were obtained.

During the last stage of the computations, the chi-square probabilities were compared with the corresponding exact probabilities. The

mean absolute percentage errors of chi-square were calculated in the .000-.005, .005-.010, .010-.050, .050-.100, .100-.150, .150-.200, and .200-1.00 probability regions. The number of times that the chi-square probabilities underestimated and overestimated the exact probabilities were also obtained in each of these regions. In addition, the extent of agreement between the exact and chi-square probabilities was determined by classifying each probability pair in a 7 X 7 table with classes equivalent to the seven probability regions described. The results of these comparisons were summarized for each sample size and summary reports were printed.

Presentation of results. The summary reports were further summarized into tables which show the results of the probability comparisons in the commonly used regions of significance for all of the sample sizes studied. Ratios of underestimates to overestimates were tabulated to illustrate some apparent trends in the numbers of underestimates and overestimates. Patterns of overestimates and underestimates for 14 distributions with N=12 were also presented.

Conclusions

It must be remembered that the following conclusions are based on data obtained from 2 X 3 contingency tables constructed in such a manner that all of the expected cell frequencies were greater than or equal to 1.

Comparisons of the chi-square probabilities of X^2 with the corresponding exact probabilities indicate that the X^2 approximation steadily improves as the sample size and therefore, the average

expected frequency, increases. The approximation is poorest in the .005-.010 significance region and tends to improve across regions as the exact probability increases. However, if close approximations to the exact probabilities are needed, the chi-square test may at times be poor. Since recommendations on the average minimum expectation depend upon the amount of error that the researcher is willing to accept, they will be left to the reader.

Numbers of underestimates and overestimates indicate that, for sample sizes of 8 through 42, the chi-square test is biased toward overestimation in the .005-.010 significance region and toward underestimation in the .010-.200 region. The relative amount of underestimation decreases as the sample size increases.

When one is interested in accepting or rejecting some hypothesis at, say, the .05 level of significance, the X^2 test seems to be quite satisfactory. In fact, the X^2 test will lead to the same decision as the exact probability at the .050 level better than 87 percent of the time if the average expectation is 2 or larger. When the average expectation is greater than or equal to 3, the extent of agreement is better than 95 percent. When there is disagreement, the chi-square probability tends to be in a lower region of significance than the exact probability.

A Suggestion for Further Research

The most intriguing result of this study is the tendency for the ratios of underestimates to overestimates of chi-square to increase as the exact probability increases and decrease as the sample size increases. This, coupled with the tentative conclusions about the

patterns of underestimation and overestimation for a limited number of exact distributions, suggests that a worthwhile correction might exist for the X^2 tests when applied to contingency tables. However, further evidence is needed in at least two areas. First, numbers and patterns of underestimates and overestimates should be investigated for tables with expectations less than one. Second, ratios of the mean percentage errors of the underestimates to overestimates should be investigated to see if there are trends similiar to those detected from the ratios of the numbers of underestimates to overestimates.

APPENDIX A

FORTRAN IV PROGRAMS FOR COMPUTATIONS

All computations were performed in three stages using four programs executed on the CDC 6400 digital computer at Lehigh University. In the first stage, row and column totals were generated by program COMB and combined to form sets of marginal totals by program TABLES. The second stage, program MARCHD, computed the exact and corresponding chi-square probabilities for all possible arrangements of cell frequencies within the limitations imposed by the marginal totals. In the last stage, the results of stage 2 were summarized by program SUMMARY to obtain the reports given in Appendix B.

Since some of the subroutines are common to several of the programs, listings of the four main programs are given first. These are followed by a listing of all of the subroutines. The subroutines required by each of the programs can be determined from the "EXTERN" section of the main documentation paragraph in the beginning of each program and subroutine.

Subroutine CARDIN and functions LRD and ICRD are used in program COMB to provide free form input of the data cards. These externals can easily be replaced by a single FORTRAN formatted read statement.

The data generated by the programs was stored on magnetic tape by means of a CDC Scope System program called UPDATE. The *DECK and *WEOR cards mentioned in the programs are control cards for UPDATE. Since UPDATE was used only so that parts of the data files could be retrieved easily, its use could be completely eliminated.

```
      PROGRAM COMB(INPUT,OUTPUT,COMBIN,
     A    TAPE1=INPUT,TAPE2=OUTPUT,TAPE3=COMBIN,TAPE7,TAPE8)
C ------------------------------------------------------------------
CCP       PROGRAM: COMBS - STAGE 1, PHASE 1 FOR DISSERTATION OF D. L
CC                 MARCH
CCP       PURPOSE: PROGRAM COMBS GENERATES COMBINATIONS OF 2,3,4,5
CC                 AND/OR 6 DIGITS WHICH SUM TO A GIVEN N.  COMBIN
CC                 WHICH DO NOT HAVE A 0 OR A 1 CLASS ARE WRITTEN TO
CC                 DISK OR TAPE FOR PHASE 2.
CCP       METHOD:  SEE TEXT OF DISSERTATION.
CCP       EXTERN:  -NONE-
CCP       LANG:    FORTRAN IV
CCP       AUTHOR:  D. L. MARCH, MARCH, 1970
C ------------------------------------------------------------------
      DIMENSION IBUFF(6),JBUFF(24),MYFORM(5)
      DATA MYFORM(1),MYFORM(2)/10H(7H*DECK R,  10H,I1,1HN,     /
      DATA MYFORM(4),MYFORM(5)/10HX,I1,I3,   ,   10HI4)        /
      DATA IF1,IF2,IF3/10H2H00,I1,60  ,  10H1H0,I2,60   ,
     A            10HI3,60      /
C                                     SET UP FREE FORM INPUT
      ND=0
      CALL CARDIN(2,1,0,0)
C                                     WRITE PAGE HEADER
      WRITE(2,1)
    1 FORMAT(1H1,/,10X,*COMBINATION GENERATOR --- D. L. MARCH*)
C                                     READ TYPE AND CHECK FOR END OF JOB
   10 NTYPE=LRD(ND)
      IF(NTYPE.EQ.3HEND) STOP
C                                     GET NUMBER OF ROWS AND TOTAL SIZE
      NROWS=ICRD(ND)
      NTOT=ICRD(ND)
C                                     SET UP VARIABLE FORMAT FOR *DECK
      MYFORM(3)=IF1
      IF(NTOT.LT.10) GO TO 15
      MYFORM(3)=IF2
      IF(NTOT.LT.100) GO TO 15
      MYFORM(3)=IF3
C                                     WRITE OUT PARAMETER LINE
   15 WRITE(2,2) NROWS,NTOT
    2 FORMAT(//,10X,*COMBINATIONS FOR*,I2,* ROWS WITH TOTAL SUM OF*,
    3 FORMAT(24I3,I1,I3,I4)
C                                     ZERO THE BUFFERS
      DO 20 I=1,6
      IBUFF(I)=0
   20 CONTINUE
      DO 30 I=1,24
      JBUFF(I)=0
   30 CONTINUE
C                                     SET UP INITIAL COMBINATION
      REWIND 7
      WRITE(7,3) NTOT
```

```
      JT=7
      NJ=1
      IT=8
      NI=0
C -----------------------------------------------------------------
C                                     IT = READ TAPE FOR PAST COMBIN.
C                                     NI = NUMBER OF COMBINATIONS ON IT
C                                     JT = WRITE TAPE FOR NEW COMBINATI
C                                     NJ = COUNTER FOR NUMBER WRITTEN
C                                          THE ABOVE ARE SWITCHED AFTER
C                                     COMPLETION OF EACH COLUMN
C                                     PERFORM FOLLOWING FOR NUMBER ROWS
      DO 100 I=2,NROWS
      NCARD=1
      NCOMB=0
      JCOUNT=0
C                                     WRITE *DECK CARD
      WRITE(3,MYFORM)      I,NTOT,I,NTOT,NCARD
      NCARD=NCARD+1
      WRITE(3,7) I,NTOT,I,NTOT,NCARD
    7 FORMAT(7H DECK R,I1,1HN,I3,60X,I1,I3,I4)
      NCARD=NCARD+1
      REWIND IT
      REWIND JT
C                                     SWITCH TAPE UNITS
      ISW=IT
      IT=JT
      JT=ISW
      NI=NJ
      NJ=0
      IN=I-1
C                                     COMPUTE NEW COMBINATIONS FROM EACH
C                                     ORIGINAL COMBINATION.
      DO 80 J=1,NI
C                                     READ A COMBINATION
      READ(IT,3) (IBUFF(K),K=1,IN)
C                                     ASSUMING SYMMETRICAL COMBINATIONS
C                                     EXCLUDED, COMPUTE CUTOFF POINT
C                                     FOR NEW GENERATION
      KDO=IBUFF(IN)*(NROWS-IN)/(NROWS-IN+1)
C                                     COMPUTE NEW COMBINATIONS TO THE
C                                     ABOVE CUTOFF
      DO 60 K=1,KDO
      IBUFF(IN)=IBUFF(IN)-1
      IBUFF(I)=K
C                                     SAVE COMBINATION
      WRITE(JT,3) (IBUFF(L),L=1,I)
      NJ=NJ+1
C                                     CHECK FOR NEW ELEMENTS GREATER
C                                     PREVIOUS ELEMENTS -- IF NOT IS A
C                                     COMBINATION WHICH WILL NOT APPEAR
```

```
C                                       LATER
      DO 40 L=1,IN
      IF(IBUFF(L).LT.IBUFF(L+1)) GO TO 55
C                                       EXCLUDE IF 0 OR 1 IN COMBINATION
      IF(IBUFF(L+1).EQ.0) GO TO 55
      IF(IBUFF(L+1).EQ.1) GO TO 55
   40 CONTINUE
C                                       SAVE IN CARD BUFFER TO WRITE OUT
      DO 50 L=1,I
      JBUFF(JCOUNT+L)=IBUFF(L)
   50 CONTINUE
      NCOMB=NCOMB+1
      JCOUNT=JCOUNT+6
C                                       IF CARD IS FULL WRITE IT
      IF(JCOUNT.LT.24) GO TO 55
      WRITE(3,3) (JBUFF(L),L=1,24),I,NTOT,NCARD
      JCOUNT=0
      NCARD=NCARD+1
   55 CONTINUE
   60 CONTINUE
   80 CONTINUE
C                                       END OF COLUMN PASS - TERMINATE
C                                       WITH A -99 CARD AND WRITE TOTALS
C                                       TO PRINTER
      IF(JCOUNT.EQ.0) GO TO 90
      JCOUNT=JCOUNT+1
      DO 85 L=JCOUNT,24
      JBUFF(L)=0
   85 CONTINUE
      WRITE(3,3) (JBUFF(L),L=1,24),I,NTOT,NCARD
      NCARD=NCARD+1
   90 WRITE(3,4) I,NTOT,NCARD
    4 FORMAT(5H-99  ,67X,I1,I3,I4)
      WRITE(2,5) I,NCOMB,NCARD
    5 FORMAT(15X,I1,* ROWS WITH *,I4,* COMBINATIONS ON*,I4,* CARDS*)
C                                       END OF PROCEEDURE
  100 CONTINUE
C -------------------------------------------------------------------
C                                       GO SEE IF MORE TO GENERATE
      GO TO 10
      END



      PROGRAM TABLES(COMBS,OUTPUT,TABLE,TAPE5=COMBS,TAPE6=OUTPUT,
     A    TAPE7=TABLE)
C -------------------------------------------------------------------
CCP       PROGRAM: TABLES - STAGE 1, PHASE 2 FOR DISSERTATION OF D.
CC                 MARCH.
CCP       PURPOSE: PROGRAM TABLES COMBINES ROW TOTALS AND COLUMN TOT
```

```
CC                TO FORM SETS OF MARGINAL TOTALS FOR 2X2 THRU 6X6
CC                TABLES.
CCP       METHOD:  SEE TEXT OF DISSERTATION.
CCP       EXTERN:  -NONE-
CCP       LANG:    FORTRAN IV
CCP       AUTHOR:  D. L. MARCH, MARCH, 1970
C ------------------------------------------------------------------
      DIMENSION BUFF(6,4),TABLE(6,3000)
          DIMENSION MYFORM(3)
          DATA MYFORM(1),MYFORM(2)/10H(7H*DECK T  ,10H,2I1,      /
C ------------------------------------------------------------------
C                                     READ TYPE - SYMM OR SKEW
          READ(5,73) NTYPE
   73     FORMAT(A10)
C                                     READ FIRST CARD OF ROW TOTALS
     1    READ(5,10) NR,NT
    10    FORMAT(7X,I1,X,I3)
C                                     IF END-OF-FILE AM DONE
          IF(EOF,5) 20,30
    20    STOP
C                                     INITIALIZE COUNTERS
    30    NTABLE=0
          NTOUT=0
C ------------------------------------------------------------------
C                                     THIS BLOCK READS IN THE ROW TOTAL
C                                     TABLE.  NOTE THAT IF ANY TOTAL
C                                     IS 1 THE COMBINATION IS REJECTED
C                                     BECAUSE THE EXPECTED FREQUENCY
C                                     BE LESS THAN 1.
    40    READ(5,50) ((BUFF(I,J),I=1,6),J=1,4)
    50    FORMAT(24F3.0)
          IF(BUFF(1,1).EQ.-99.0) GO TO 100
    60    DO 80 J=1,4
          IF(BUFF(1,J).EQ.0.0) GO TO 40
          IF(BUFF(NR,J).EQ.1.0) GO TO 80
          NTABLE=NTABLE+1
          DO 70 I=1,6
          TABLE(I,NTABLE)=BUFF(I,J)
    70    CONTINUE
    80    CONTINUE
          GO TO 40
C ------------------------------------------------------------------
C                                     READ FIRST CARD OF COLUMN TOTALS
C                                     BE SURE SAMPLE SIZES AGREE
   100    CONTINUE
          NC=NR
          IF(NTYPE.EQ.4HSYMM) GO TO 120
          READ(5,10) NC,I
          IF(I.EQ.NT) GO TO 120
          WRITE(6,110) NR,NT,NC,I
   110    FORMAT(1H1,10X,*TOTAL MISMATCH FOR *,2(1H(,I3,1H,,I3,1H)))
```

```
      STOP 777
C                                     SET UP VARIABLE FORMAT FOR *DECK
  120 MYFORM(3)=10H1HN,I3)
      IF(NT.LT.100) MYFORM(3)=10H2HN0,I2)
      IF(NT.LT.10) MYFORM(3)=10H3HN00,I1)
      WRITE(7,MYFORM) NR,NC,NT
      WRITE(7,130) NR,NC,NT
  130 FORMAT(7H DECK T,2I1,1HN,I3)
C                                     SET UP INITIAL BOT OF TABLE POINT
      IF(NR.NE.NC) NBOT=1
      IF(NR.EQ.NC) NBOT=0
C ------------------------------------------------------------------
C                                     GENERATE TABLES EXCLUDING THOSE
C                                     FE S LESS THAN 1
  140 CONTINUE
      IF(NTYPE.EQ.4HSYMM) GO TO 3000
      READ(5,50) ((BUFF(I,J),I=1,6),J=1,4)
      IF(BUFF(1,1).NE.-99.0) GO TO 170
C                                     END OF INPUT --- WRITE *WEOR
  143 CONTINUE
  150 WRITE(7,155)
  155 FORMAT(3H-99)
      WRITE(6,160) NR,NC,NT,NTOUT
  160 FORMAT(/,10X,I1,1HX,I1,* FOR SIZE*,I4,* --- NUMBER OF TABL
     A     *,I4)
      GO TO 1
C                                     J COMBINATIONS IN BUFFER
  170 DO 500 J=1,4
      IF(BUFF(1,J).EQ.0.0) GO TO 140
      IF(BUFF(NC,J).EQ.1.0) GO TO 500
C                                     MATCH INPUT WITH EACH REMAINING
C                                     TABLE COMBINATION
      DO 200 N=NBOT,NTABLE
C                                     COMPUTE FE
      DO 190 I=1,NR
      DO 180 K=1,NC
      FE=TABLE(I,N)*BUFF(K,J)/FLOAT(NT)
      IF(FE.LT.1.0) GO TO 200
  180 CONTINUE
  190 CONTINUE
C                                     HAVE A GOOD TABLE --- WRITE OUT
      WRITE(7,50) (TABLE(I,N),I=1,6), (BUFF(I,J),I=1,6)
      NTOUT=NTOUT+1
  200 CONTINUE
      IF(NTYPE.EQ.4HSYMM) GO TO 3000
  500 CONTINUE
C                                     GO BACK FOR A NEW CARD
      GO TO 140
C ------------------------------------------------------------------
C                                     SET UP BUFFER FOR SYMMETRICAL COM
 3000 CONTINUE
```

```
      NBOT=NBOT+1
      IF(NBOT.GT.NTABLE) GO TO 143
      DO 3010 KX=1,6
      BUFF(KX,1)=TABLE(KX,NBOT)
3010  CONTINUE
      GO TO 170
      END
```

```
      PROGRAM MARCHD(COMPILE,RESULT,OUTPUT,TAPE5=COMPILE,TAPE6=OUTPU
     A    TAPE7=RESULT)
C ------------------------------------------------------------------
CCP       PROGRAM: MARCHD - STAGE 2 FOR DISSERTATION OF D. L. MARCH
CCP       PURPOSE: PROGRAM MARCHD COMPUTES CHI-SQUARE, THE CHI-SQUAR
CC                 PROBABILITY, THE EXACT PROBABILITY, AND THE EXACT
CC                 CUMULATIVE PROBABILITY FOR EACH AND EVERY POSSIBLE
CC                 ARRAY OF OBSERVED CELL FREQUENCIES THAT CAN BE
CC                 GENERATED FROM A GIVEN SET OF MARGINAL TOTALS.
CC                 THE PROGRAM WILL HANDLE 2X2 THRU 2X6 TABLES.
CC                 HOOKS ARE PROVIDED FOR 3X3 THRU 3X6 TABLES.
CCP       NOTE:    BY SETTING MACHINE PSEUDO SWITCH 1, EACH EXACT
CC                 DISTRIBUTION COMPUTED WILL BE PRINTED.
CCP       METHOD:  SEE TEXT OF DISSERTATION.
CCP       EXTERN:  CALFE - SUBROUTINE TO COMPUTE EXPECTED FREQUENCIES
CC                      FROM THE MARGINAL TOTALS
CC                 CALRXL - SUBROUTINE TO OBTAIN THE LOG OF THE R.X
CC                      TERM FOR THE METHOD OF FREEMAN AND HALTON.
CC                 SORTER - SUBROUTINE TO STORE AND SORT ARRAYS
CC                      OF CELL FREQUENCIES AND EXACT PROBABILITIES
CC                 CELLFIL - SUBROUTINE TO GET AN ARRAY OF VALID CELL
CC                      FREQUENCIES.
CC                 CALEP - SUBROUTINE TO OBTAIN THE FINAL VALUE OF
CC                      THE EXACT PROBABILITY FOR AN ARRAY.
CC                 UNPACK - SUBROUTINE TO UNPACK 2 DIGIT NUMBERS
CC                      FROM A PACKED WORD.
CC                 CALCHI - SUBROUTINE TO COMPUTE X-SQUARE AND GET
CC                      CORRESPONDING CHI-SQUARE PROBABILITY.
CC                 B2DUOD - SUBROUTINE TO PACK A 60 BIT WORD INTO
CC                      12 CHARACTERS.
CCP       LANG:    FORTRAN IV
CCP       AUTHOR:  D. L. MARCH, APRIL, 1970
C ------------------------------------------------------------------
      COMMON/MATRIX/ NROW,NRTOT(6),NCOL,NCTOT(6)
     A    ,NTOTAL,IFO(3,6),FE(3,6)
      COMMON/SORT/ LENGTH,INDEX(2000)
     A    ,PROBS(2000),CELLS(3,2000)
      COMMON/CHICOM/ XSQ,PXSQ,NDF
      DIMENSION IOUT(8)
      DIMENSION MYFORM(3)
      DATA MYFORM(1),MYFORM(2)/10H(7H*DECK V      ,      10H,2I1,
      DATA IORD,IOWR,IORES/5,6,7/
C ------------------------------------------------------------------
C                                     GET THE LIST SWITCH SETTING
      CALL SSWTCH(1,JSW1)
C                                     READ THE DECK CARD
    1 CONTINUE
      READ(IORD,10) NROW,NCOL,NTOTAL
   10 FORMAT(7X,2I1,X,I3)
      IF(EOF,IORD) 20,50
C                                     END OF INPUT
```

```
   20 CALL REMARK(27H ** EOF-NORMAL TERMINATION  )
      STOP
   50 CONTINUE
      MYFORM(3)=10H2HN0I2)
      IF(NTOTAL.LT.10) MYFORM(3)=10H3HN00I1)
C                                     WRITE *DECK AND DECK CARDS
      WRITE(IORES,MYFORM) NROW,NCOL,NTOTAL
      WRITE(IORES,60) NROW,NCOL,NTOTAL
   60 FORMAT(7H DECK V,2I1,1HN,I3)
C ------------------------------------------------------------------
  100 CONTINUE
C                                     READ ROW AND COLUMN TOTALS
      READ(IORD,110) (NRTOT(I),I=1,6),(NCTOT(I),I=1,6)
  110 FORMAT(24I3)
C                                     CHECK FOR END OF THIS DECK
      IF(NRTOT(1).EQ.-99) WRITE(IORES,115)
  115 FORMAT(3H-99)
      IF(NRTOT(1).EQ.-99) GO TO 1
C                                     WRITE ROW AND COLUMN TOTALS
      WRITE(IORES,110) (NRTOT(I),I=1,6),(NCTOT(I),I=1,6)
C                                     ZERO THE FE MATRIX
      DO 120 I=1,3
      DO 120 J=1,6
      FE(I,J)=0.0
  120 CONTINUE
C                                     COMPUTE THE FE S
      CALL CALFE
C                                     SEE IF PRINTOUT IS DESIRED
      IF(JSW1.EQ.2) GO TO 200
C                                     YES...DO THE PRINT
      WRITE(IOWR,160) NROW,NCOL,NTOTAL,
     A     (NRTOT(I),I=1,NROW)
  160 FORMAT(1H1,/,1H0,4X,I1,1HX,I1,16H TABLE WITH N = ,I2,/,
     A     9X,13HROW TOTALS = ,6I3)
      WRITE(IOWR,161) (NCTOT(I),I=1,NCOL)
  161 FORMAT(9X,13HCOL TOTALS = ,6I3)
      WRITE(IOWR,162)
  162 FORMAT(1H0)
      WRITE(IOWR,165) (FE(1,J),J=1,NCOL)
  165 FORMAT(5X,21HEXPECTED FREQUENCIES: ,2X,6F9.3)
      DO 175 I=2,NROW
      WRITE(IOWR,170) (FE(I,J),J=1,NCOL)
  170 FORMAT(28X,6F9.3)
  175 CONTINUE
      WRITE(IOWR,180)
  180 FORMAT(1H0,7X,8H ROW  1 ,2X,8H ROW  2 ,2X,
     A     10H  EXACTP  ,2X,
     B     10H   SUMP   ,2X,
     C     10H    XSQ   ,2X,
     D     10H   PCHI    ,/,
     E     6X,2(2X,8H-------- ),4(2X,10H---------- ))
```

```
C ------------------------------------------------------------------
  200 CONTINUE
C                                       COMPUTE RXL TERM OF EXACTP
      CALL CALRXL
C                                       INITIALIZE THE SORT
      CALL SORTER(1)
C                                       MAKE CALLS TO FILL CELLS AND FIND
C                                       EXACT P
      IGO=1
  250 CONTINUE
      CALL CELLFIL(IGO,IEND)
C                                       CHECK FOR DONE
      IF(IEND.NE.0) GO TO 300
      CALL CALEP
C                                       STORE THE CELLS AND EXACT P
      CALL SORTER(2)
      GO TO 250
C ------------------------------------------------------------------
  300 CONTINUE
C                                       SORT THE RESULTS
      CALL SORTER(3)
      SUMP=0
      EXACTP=10.0
C                                       CALCULATE FINAL RESULTS
      DO 390 I=1,LENGTH
C                                       TEST FOR EQUAL...ALLOW FOR
C                                       FLOATING POINT ROUND OFF ERROR AT
C                                       SIXTH DIGIT
      PD=ABS(PROBS(I)-EXACTP)/EXACTP
      IF(PD.LE.1.0E-5) GO TO 350
      EXACTP=PROBS(I)
      SUMP=SUMP+EXACTP
      JBOT=I+1
      IF(JBOT.GT.LENGTH) GO TO 350
      DO 320 J=JBOT,LENGTH
      PD=ABS(PROBS(J)-EXACTP)/EXACTP
      IF(PD.GE.1.0E-5) GO TO 350
      SUMP=SUMP+EXACTP
  320 CONTINUE
  350 CONTINUE
      J=INDEX(I)
      CALL UNPACK(IFO,CELLS(1,J))
C                                       GET CHI SQUARE VALUES
      CALL CALCHI
C                                       CONVERT BINARY TO 12 CHARACTERS
      CALL B2DUOD(EXACTP,IOUT(1))
      CALL B2DUOD(SUMP,IOUT(3))
      CALL B2DUOD(XSQ,IOUT(5))
      CALL B2DUOD(PXSQ,IOUT(7))
C                                       WRITE VALUES TO RESULT
      WRITE(IORES,360) ((IFO(J,K),K=1,6),J=1,2)
```

```
     A     ,(IOUT(J),J=1,8)
  360 FORMAT(12I2,4(A10,A2))
C                                        SEE IF SHOULD BE PRINTED
      IF(JSW1.EQ.2) GO TO 390
      WRITE(IOWR,380) I,((IFO(J,K),K=1,3),J=1,2),
     A     EXACTP,SUMP,XSQ,PXSQ
  380 FORMAT(X,I5,2X,2(I2,X),I2,2X,2(I2,X),I2,4(2X,E10.3))
  390 CONTINUE
C                                        SIGNAL END OF TABLE WITH -9 CARD
      WRITE(IORES,400)
  400 FORMAT(2H-9)
      GO TO 100
      END
```

```
      PROGRAM SUMMARY(COMPILE,OUTPUT,TAPE5=COMPILE,TAPE6=OUTPUT)
C ------------------------------------------------------------------
CCP       PROGRAM: SUMMARY - STAGE 3 FOR DISSERTATION OF D. L. MARCH
CCP       PURPOSE: PROGRAM SUMMARY SUMMARIZES THE RESULTS OF STAGE
CC                 2 COMPUTATIONS FOR EACH N BY (1) COMPUTING ABS
CC                 PERCENTAGE ERRORS FOR CHI-SQUARE PROBABILITIES
CC                 RESPECT TO EXACT PROBABILITIES AND (2) CROSS
CC                 TABULATING CHI-SQUARE AND EXACT PROBABILITIES.
CC                 BOTH OPERATIONS ARE SUMMARIZED FOR SEVEN
CC                 PROBABILITY REGIONS.  THE PROGRAM WILL HANDLE
CC                 RESULTS FOR 2X2 THRU 2X6 AND 3X3 THRU 3X6 TABLES.
CCP       METHOD:  SEE TEXT OF DISSERTATION.
CCP       EXTERN:  CALFE - SUBROUTINE TO COMPUTE EXPECTED FREQUENCIES
CC                      FROM THE MARGINAL TABLES.
CC                 DUOD2B - SUBROUTINE TO CONVERT 12 CHARACTER PACKED
CC                      DATA INTO A 60 BIT FLOATING POINT NUMBER.
CC                 RANGE - SUBROUTINE TO DETERMINE THE SIGNIFICANCE
CC                      REGION FOR A PROBABILITY.
CCP       LANG:    FORTRAN IV
CCP       AUTHOR:  D. L. MARCH, APRIL, 1970
C ------------------------------------------------------------------
      COMMON/MATRIX/ NROW,NRTOT(6),NCOL,NCTOT(6)
     A     ,NTOTAL,IFO(3,6),FE(3,6)
      DIMENSION IN(8)
      DIMENSION LHIGH(7),LLOW(7),XM(7),ICROSS(7,7),LNUM(7)
      DIMENSION IR(7)
      DIMENSION MYFORM(12)
      DATA MYFORM(1)/10H(1H0,4X,  /
      DATA MYFORM(2)/10H1H.,A10,  /
      DATA MYFORM(3)/10H3X,1H(,I1,/
      DATA MYFORM(4)/10H1H)       /
      DATA MYFORM(12)/10H)          /
      DATA (IR(I),I=1,7)/
     1     10H000≤P≤.005
     2    ,10H005<P≤.010
     3    ,10H010<P≤.050
     4    ,10H050<P≤.100
     5    ,10H100<P≤.150
     6    ,10H150<P≤.200
     7    ,10H200<P     /
      DATA IORD,IOWR/5,6/
C ------------------------------------------------------------------
C                                     GET THE LIST SWITCH SETTING
      CALL SSWTCH(1,JSW1)
C                                     READ THE DECK CARD
    1 CONTINUE
      READ(IORD,10) ITYP,NROW,NCOL,NTOTAL
   10 FORMAT(A5,2X,2I1,X,I3)
      IF(EOF,IORD) 20,50
C                                     END OF INPUT
   20 CALL REMARK(27H ** EOF-NORMAL TERMINATION  )
```

```
      STOP
   50 CONTINUE
C                                     BE SURE READ IS FROM COMPILE FILE
      IF(ITYP.EQ.5H*DECK ) GO TO 1
C                                     INITIALIZE ALL COUNTERS
      AMIN=2.0
      AMAX=0.0
      LTABS=0
      LPROBS=0
      NDG=0
      NADG=0
      NBDG=0
      DO 60 I=1,7
      LHIGH(I)=0
      LLOW(I)=0
      XM(I)=0.0
      DO 60 J=1,7
      ICROSS(I,J)=0
   60 CONTINUE
      GO TO 105
C ------------------------------------------------------------------
  100 CONTINUE
C                                     COMPUTE AMIN AND AMAX
      IF(SUMP.LT.AMIN) AMIN=SUMP
      IF(SUMP.GT.AMAX) AMAX=SUMP
  105 CONTINUE
C                                     READ ROW AND COLUMN TOTALS
      READ(IORD,110) (NRTOT(I),I=1,6),(NCTOT(I),I=1,6)
  110 FORMAT(24I3)
C                                     CHECK FOR END OF THIS DECK
      IF(NRTOT(1).EQ.-99) GO TO 500
C                                     ADD TO NUMBER OF TABLES
      LTABS=LTABS+1
C                                     ZERO THE FE MATRIX
      DO 120 I=1,3
      DO 120 J=1,6
      FE(I,J)=0.0
      IFO(I,J)=0
  120 CONTINUE
C                                     COMPUTE THE FE S
      CALL CALFE
C                                     SEE IF PRINTOUT IS DESIRED
      IF(JSW1.EQ.2) GO TO 200
C                                     YES...DO THE PRINT
      WRITE(IOWR,160) NROW,NCOL,NTOTAL
     A     ,(NRTOT(I),I=1,3)
     B     ,(NCTOT(I),I=1,6)
  160 FORMAT(1H1,4X,4HNR =,I2,4X,4HNC =,I2,4X,4HNT =,I3,
     A     4X,8HNRTOT = ,2(I2,X),I2,
     B     4X,8HNCTOT = ,5(I2,X),I2,//)
      WRITE(IOWR,165) (FE(1,J),J=1,NCOL)
```

```
  165 FORMAT(5X,21HEXPECTED FREQUENCIES: ,2X,6F9.3)
      DO 175 I=2,NROW
      WRITE(IOWR,170) (FE(I,J),J=1,NCOL)
  170 FORMAT(28X,6F9.3)
  175 CONTINUE
      WRITE(IOWR,180)
  180 FORMAT(///,16X,5HROW 1,16X,5HROW 2,16X,5HROW 3,13X,6HEXACTP,
     A     11X,4HSUMP,12X,3HXSQ,13X,4HPCHI ,/,
     B     6X,3(4X,17H---------------- ),4(4X,12H------------ ))
C ------------------------------------------------------------------
  200 CONTINUE
C                                       READ THE VALUES
      READ(IORD,210) ((IFO(J,K),K=1,6),J=1,2),(IN(J),J=1,8)
  210 FORMAT(12I2,4(A10,A2))
C                                       CHECK FOR END OF TABLE
      IF(IFO(1,1).EQ.-9) GO TO 100
      IF(NROW.LE.2) GO TO 220
C                                       RESTORE THE THIRD ROW
      DO 215 I=1,NCOL
      IFO(3,I)=NCTOT(I)-IFO(1,I)-IFO(2,I)
  215 CONTINUE
  220 CONTINUE
C                                       ADD TO NUMBER OF PROBABILITIES
      LPROBS=LPROBS+1
C                                       CONVERT FROM CHARACTERS TO NUMBER
      CALL DUOD2B(IN(1),EXACTP)
      CALL DUOD2B(IN(3),SUMP)
      CALL DUOD2B(IN(5),XSQ)
      CALL DUOD2B(IN(7),PXSQ)
C ------------------------------------------------------------------
  300 CONTINUE
C                                       GET CROSS TAB SUBS
      CALL RANGE(SUMP,LEP)
      CALL RANGE(PXSQ,LCP)
      IF(LEP.EQ.LCP) NDG=NDG+1
      IF(LEP.LT.LCP) NBDG=NBDG+1
      IF(LEP.GT.LCP) NADG=NADG+1
      ICROSS(LEP,LCP)=ICROSS(LEP,LCP)+1
      DIFF=SUMP-PXSQ
C                                       ADD PERCENTAGE DIFFERENCE
      XM(LEP)=XM(LEP)+ABS(DIFF)/SUMP
C                                       TALLEY HIGH/LOW VALUES
      IF(DIFF.GT.0.0) LLOW(LEP)=LLOW(LEP)+1
      IF(DIFF.LE.0.0) LHIGH(LEP)=LHIGH(LEP)+1
C                                       SEE IF SHOULD BE PRINTED
      IF(JSW1.EQ.2) GO TO 390
      WRITE(IOWR,380) I,((IFO(J,K),K=1,6),J=1,3),
     A     EXACTP,SUMP,XSQ,PXSQ
  380 FORMAT(X,I5
     A     ,4X,5(I2,X),I2
     B     ,4X,5(I2,X),I2
```

```
     C    ,4X,5(I2,X),I2
     D    ,4(4X,E12.4))
  390 CONTINUE
C                                        GO FOR NEXT CASE
      GO TO 200
C ---------------------------------------------------------------
  500 CONTINUE
      XMTOT=0.0
      LLTOT=0
      LHTOT=0
C                                        COMPUTE TOTALS AND MEANS
      DO 510 I=1,7
      XMTOT=XMTOT+XM(I)
      LNUM(I)=LLOW(I)+LHIGH(I)
      IF(LNUM(I).GT.0)
     A    XM(I)=100.0*XM(I)/FLOAT(LNUM(I))
      LLTOT=LLTOT+LLOW(I)
      LHTOT=LHTOT+LHIGH(I)
  510 CONTINUE
      XMTOT=100.0*XMTOT/FLOAT(LPROBS)
C                                        WRITE PAGE HEADER FOR SUMMARY
      WRITE(IOWR,520) LTABS,NROW,NCOL,NTOTAL
  520 FORMAT(1H1,4X,12HSUMMARY FOR ,I4,I3,1HX,I1,
     A    17H TABLES WITH N = ,I2,/)
C                                        WRITE NOTE LINE
      WRITE(IOWR,525)
  525 FORMAT(5X,
     AALL EXPECTED CELL FREQUENCIES WERE ≥ 1
     B    )
C                                        WRITE ACCURACY CHECK
      WRITE(IOWR,527) AMIN,AMAX
  527 FORMAT(1H0,4X,16HACCURACY CHECK:  ,2X,F10.2,14H ≤ SUM P(E) ≤ ,
     A    F10.2,//)
C                                        WRITE SUMMARY HEADER
      WRITE(IOWR,530)
  530 FORMAT(1H0,6X,7HRANGE E,7X,3HNO.,3X,5H C<E ,4X,4HC≥E ,5X,4HMEAN
     A    5X,11H-----------,3X,5H-----,2(3X,5H-----),3X,7H------- )
C                                        WRITE SUMMARY
      DO 550 I=1,7
      WRITE(IOWR,540) IR(I),LNUM(I),LLOW(I),LHIGH(I),XM(I)
  540 FORMAT(1H0,4X,1H.,A10,3I8,F10.2)
  550 CONTINUE
      WRITE(IOWR,555)
  555 FORMAT(16X,3(3X,5H-----),3X,7H------- )
C                                        WRITE TOTAL SUMMARY LINE
      WRITE(IOWR,560) LPROBS,LLTOT,LHTOT,XMTOT
  560 FORMAT(1H0,10X,5HTOTAL,3I8,F10.2)
C                                        WRITE CROSS TABS HEADER
      WRITE(IOWR,570) (I,I=1,7)
  570 FORMAT(1H0,/,1H0,/,1H0,4X,10HCHI-SQUARE ,
     A    23X,19HEXACT PROBABILITIES ,/,
```

```
     B     5X,13HPROBABILITIES ,3X,
     C     7(4X,1H(,I1,1H)),/,
     D     22X,7(2X,5H-----))
C                                        WRITE CROSS TABS MATRIX
      DO 590 J=1,7
C                                        SET UP FORMAT TO SUPPRESS ZEROS
      DO 580 K=1,7
      L=K+4
      MYFORM(L)=10H,I7
      IF(ICROSS(K,J).LE.0) MYFORM(L)=10H,A1,6X
  580 CONTINUE
      WRITE(IOWR,MYFORM) IR(J),J,(ICROSS(I,J),I=1,7)
  590 CONTINUE
C                                        COMPUTE DIAGONAL PERCENTS
      XDG=100.0*FLOAT(NDG)/FLOAT(LPROBS)
      XADG=100.0*FLOAT(NADG)/FLOAT(LPROBS)
      XBDG=100.0*FLOAT(NBDG)/FLOAT(LPROBS)
C                                        WRITE THE RESULTS
      WRITE(IOWR,600) NADG,XADG,NDG,XDG,NBDG,XBDG
  600 FORMAT(1H0,//,44X,3HNO.,4X,7HPERCENT,/,
     A     43X,5H-----,3X,7H-------,//,
     B     37X,3HC<E,I8,F10.2,//,
     B     37X,3HC=E,I8,F10.2,//,
     B     37X,3HC>E,I8,F10.2)
      GO TO 1
      END
```

```
      SUBROUTINE RANGE(X,L)
C -----------------------------------------------------------------
CCP       CALL:     CALL RANGE(X,L)
CCP       PURPOSE:  SUBROUTINE RANGE CLASSIFIES A GIVEN PROBABILITY
CC                  INTO ONE OF SEVEN PROBABILITY REGIONS.
CCP       ARG-IN:   X - THE PROBABILITY VALUE
CCP       ARG-OUT:  L - THE REGION NUMBER FROM 1 TO SEVEN FOR REGIONS
CC                            P≤.005
CC                       .005<P≤.010
CC                       .010<P≤.050
CC                       .050<P≤.100
CC                       .100<P≤.150
CC                       .150<P≤.200
CC                       .200<P
CCP       COMMON:   -NONE-
CCP       IO:       -NONE-
CCP       METHOD:   THE VALUE OF X IS COMPARED AGAINST SUCCESSIVELY
CC                  LARGER UPPER LIMITS UNTIL IT IS FOUND TO BE LESS
CC                  THAN THE UPPER LIMIT OF A REGION.
CCP       EXTERN:   -NONE-
CCP       AUTHOR:   D. L. MARCH, APRIL, 1970
CCP       LANG:     FORTRAN IV
C -----------------------------------------------------------------
      L=1
      IF(X.LE.0.005) RETURN
      L=2
      IF(X.LE.0.01) RETURN
      L=3
      IF(X.LE.0.05) RETURN
      L=4
      IF(X.LE.0.10) RETURN
      L=5
      IF(X.LE.0.15) RETURN
      L=6
      IF(X.LE.0.20) RETURN
      L=7
      RETURN
      END



      IDENT     B2DUOD
      SPACE     2
* -----------------------------------------------------------------
**P       CALL:     CALL B2DUOD(IN,OUT)
**P       PURPOSE:  SUBROUTINE B2DUOD CONVERTS A 60 BIT WORD TO 12 HO
**                  ERITH CHARACTERS AND PUTS THE CHARACTERS INTO TWO
**                  OUTPUT WORDS IN A10,A2 FORMAT.  THIS IS USED TO
**                  RETAIN ALL 60 BITS OF ACCURACY BETWEEN RUNS.  THE
**                  RESULTS OF THIS ROUTINE ARE REVERSED BY SUBROUTIN
```

```
**                  DUOD29.
**P       ARG-IN:   IN - THE 60 BIT WORD TO BE CONVERTED.
**P       ARG-OUT:  OUT - A TWO WORD ARRAY TO RECEIVE THE CHARACTERS
**                       IN A10, A2 FORMAT.
**P       METHOD:   THE 60 BITS ARE EXTRACTED 5 BITS AT A TIME AND PUT
**                  IN THE OUTPUT WORDS.  A FUDGE OF ALL 4.S IS ADDED
**                  TO EACH WORD SO THAT ZERO BYTES ARE NOT POSSIBLE
**                  IN THE CHARACTERS.  THIS IS DONE BECAUSE OF THE
**                  SPECIAL MEANING OF ZERO BYTES TO SCOPE INPUT-OUTP
**P       LANG:     COMPASS (CP) - CDC 6400
**P       AUTHOR:   DAVID L. MARCH, JUNE 1970
* ---------------------------------------------------------------------
          SPACE     2
          ENTRY     B2DUOD
 B2DUOD   PS
          SA1       B1                GET INPUT WORD
          MX0       55
          BX6       -X6*X6
          SB6       9
 LOOP     LX1       5
          LX6       6
          BX2       -X0*X1            MASK 5 BITS FROM INPUT
          BX6       X2+X6             ADD TO OUTPUT
          SB6       B6-1
          GE        B6,B0,LOOP
          SA5       =10HAAAAAAAAAA    FUDGE FACTOR TO BE SURE NO ZEROS
          IX6       X5+X6             ADD FUDGE FACTOR
          SA6       B2                STORE FIRST 10 CHARACTERS
          LX1       5                 GET LAST 10 BITS AS TWO CHARACTER
          BX6       -X0*X1
          LX1       5
          LX6       6
          BX2       -X0*X1
          BX6       X2+X6
          LX6       48                CONVERT TO A FORMAT
          IX6       X5+X6             ADD FUDGE FACTOR
          SA6       B2+1              STORE LAST 2 CHARACTERS
          EQ        B2DUOD
          END



          IDENT     DUOD29
          SPACE     2
* ---------------------------------------------------------------------
**P       CALL:     DUOD28(IN,OUT)
**P       PURPOSE:  SUBROUTINE DUOD28 REVERSES THE ACTION OF SUB
**                  B2DUOD.  THUS, IT CONVERTS 12 CHARACTERS OF INPUT
**                  INTO ONE 60 BIT WORD.
**P       ARG-IN:   IN - THE 12 CHARACTERS IN TWO WORDS AS A10,A2
```

```
**P       ARG-OUT: OUT - THE RESULTING 60 BIT WORD
**P       LANG:    COMPASS (CP) - CDC 6400
**P       AUTHOR:  DAVID L. MARCH, JUNE 1970
* ------------------------------------------------------------
          SPACE    2
          ENTRY    DUOD2B
 DUOD2B   PS
          SA1      B1              GET FIRST 10 CHARS
          MX0      54
          BX6      -X6*X6
          SA5      =10HAAAAAAAAAA
          IX1      X1-X5           SUBTRACT OUT FUDGE FACTOR
          SB6      9
 LOOP     LX1      6
          LX6      5
          BX2      -X0*X1
          BX6      X2+X6
          SB6      B6-1
          GE       B6,B0,LOOP
          SA1      B1+1
          IX1      X1-X5
          LX1      6
          LX6      5
          BX2      -X0*X1
          BX6      X2+X6
          LX1      6
          LX6      5
          BX2      -X0*X1
          BX6      X2+X6
          SA6      B2              STORE THE RESULT
          EQ       DUOD2B
          END
```

```
      SUBROUTINE CELLFIL(IGO,IEND)
C ------------------------------------------------------------
CCP       CALL:    CALL CELLFIL(IGO,IEND)
CCP       PURPOSE: SUBROUTINE CELLFIL OBTAINS THE NEXT VALID ARRAY
CC                 OF CELL FREQUENCIES THAT CAN BE GENERATED SUBJECT
CC                 THE LIMITATIONS IMPOSED BY THE MARGINAL TOTALS.
CCP       ARG-IN:  IGO - ACTION TO BE PERFORMED WHERE:
CC                      1 = INITIALIZE AND OBTAIN FIRST ARRAY
CC                      2 = OBTAIN NEXT SET OF CELL FREQUENCIES
CC                 IEND - SIGNALS END OF ARRAYS WHERE
CC                      0 = ARRAY RETURNED
CC                      1 = NO MORE ARRAYS
CCP       COMMON:  /MATRIX/ - WHICH CONTAINS ROW AND COLUMN TOTALS
CCP       IO:      -NONE-
CCP       METHOD:  SEE TEXT OF DISSERTATION
```

```
CCP       EXTERN:  -NONE-
CCP       LANG:    FORTRAN IV
CCP       AUTHOR:  D. L. MARCH, MARCH, 1970
C ------------------------------------------------------------------
      COMMON/MATRIX/ NROW,NRTOT(6),NCOL,NCTOT(6)
     A    ,NTOTAL,IFO(3,6),FE(3,6)
C ------------------------------------------------------------------
      IF(IGO.EQ.2) GO TO 200
C                                     INITIALIZE
  100 CONTINUE
      IEND=0
      IGO=2
C                                     ZERO THE MATRICES
      DO 110 I=1,3
      DO 110 J=1,6
      IFO(I,J)=0
  110 CONTINUE
C                                     SET THE RIGHT CELLS OF ROW ONE
C                                     THESE ARE DEGREES OF FREEDOM
      ISUM=0
      DO 120 I=2,NCOL
      IFO(1,I)=NCTOT(I)
      ISUM=ISUM+NCTOT(I)
  120 CONTINUE
C                                     GO FILL REST OF CELLS
      GO TO 300
C ------------------------------------------------------------------
C                                     SUBSEQUENT CALL
  200 CONTINUE
C                                     GENERATE NEW CELLS BY SUCCESSIVE
C                                     SUBTRACTIONS OF ONE FROM DEGREES
C                                     FREEDOM CELLS
      DO 220 I=2,NCOL
      IFO(1,I)=IFO(1,I)-1
      ISUM=ISUM-1
      IF(IFO(1,I).GE.0) GO TO 300
C                                     CELL IS NEGATIVE...RESTORE IT AND
C                                     LOOK AT ONE TO THE RIGHT
      IFO(1,I)=NCTOT(I)
      ISUM=ISUM+NCTOT(I)+1
  220 CONTINUE
C                                     UNABLE TO FILL DEGREES OF FREEDOM
C                                     SIGNAL END OF CELLS
      IEND=1
      RETURN
C                                     HERE TO FILL OTHER CELLS
  300 CONTINUE
C                                     CHECK CELL (1,1)
      IFO(1,1)=NRTOT(1)-ISUM
      IF(IFO(1,1).LT.0) GO TO 200
C                                     FILL AND CHECK ROW TWO CELLS
```

```
      DO 320 I=1,NCOL
      IFO(2,I)=NCTOT(I)-IFO(1,I)
      IF(IFO(2,I).LT.0) GO TO 200
  320 CONTINUE
C                                     CELLS ARE ALL GOOD
      IFND=0
      RETURN
      END



      SUBROUTINE CALEP
C -------------------------------------------------------------------
CCP       CALL:    CALL CALEP
CCP       PURPOSE: SUBROUTINE CALCULATES THE EXACT PROBABILITY FOR
CC                 AN ARRAY OF CELL FREQUENCIES.  THE ROUTINE EXPECT
CC                 THE LOG OF THE CONSTANT R.X TERM TO BE IN COMMON
CC                 /EXACT/.
CCP       ARG-IN:  -NONE-
CCP       ARG-OUT: -NONE-
CCP       COMMON:  /MATRIX/ - THE CELL FREQUENCIES AND MARGINAL TOTAL
CC                 /EXACT/ - R.X AND THE RESULT OF THIS ROUTINE.
CCP       METHOD:  SEE TEXT OF THE DISSERTATION.
CCP       EXTERN:  FACTLOG - FUNCTION TO RETURN THE LOG TO THE BASE
CC                      10 OF N FACTORIAL.
CCP       LANG:    FORTRAN IV
CCP       AUTHOR:  D. L. MARCH, APRIL, 1970
C -------------------------------------------------------------------
      COMMON/MATRIX/ NROW,NRTOT(6),NCOL,NCTOT(6)
     A    ,NTOTAL,IFO(3,6),FE(3,6)
      COMMON/EXACT/ RXLOG,EXACTP
C -------------------------------------------------------------------
C                                     SUM LOGS OF CELL FREQ S
      EXACTP=0.0
      DO 10 I=1,NROW
      DO 10 J=1,NCOL
      EXACTP=EXACTP+FACTLOG(IFO(I,J))
   10 CONTINUE
C                                     GET LOG OF RESULT
      EXACTP=RXLOG-EXACTP
C                                     CONVERT TO A NUMBER
      EXACTP=10.0**EXACTP
      RETURN
      END
```

```
      SUBROUTINE CALFE
C -------------------------------------------------------------------
CCP       CALL:     CALL CALFE
CCP       PURPOSE:  SUBROUTINE CALFE CALCULATES THE EXPECTED FREQ
CC                  FOR A CONTINGENCY TABLE FROM THE MARGINAL TOTALS
CC                  AND THE TOTAL SAMPLE SIZE
CCP       ARG-IN:   -NONE- (BUT SEE COMMON DESCRIPTION)
CCP       ARG-OUT:  -NONE- (BUT SEE COMMON DESCRIPTION)
CCP       COMMON:   /MATRIX/ - A LABELED COMMON BLOCK CONTAINING
CC                       NROW - THE NUMBER OF ROWS IN THE TABLE
CC                       NCOL - THE NUMBER OF COLUMNS IN THE TABLE
CC                       NRTOT - THE ROW TOTALS
CC                       NCTOT - THE COLUMN TOTALS
CC                       NTOTAL - THE TOTAL SAMPLE SIZE
CC                       FE - THE EXPECTED FREQ MATRIX
CCP       IO:       -NONE-
CCP       METHOD:   THE EXPECTED FREQUENCIES ARE CALCULATED FROM THE
CC                  MARGINAL TOTALS AND THE TOTAL SAMPLE SIZE.
CCP       EXTERN:   FLOAT - AN IN-LINE FUNCTION TO FLOAT A NUMBER
CCP       LANG:     FORTRAN IV
CCP       AUTHOR:   DAVID L. MARCH, MAY 1970
C -------------------------------------------------------------------
      COMMON/MATRIX/ NROW,NRTOT(6),NCOL,NCTOT(6)
     A     ,NTOTAL,IFO(3,6),FE(3,6)
C -------------------------------------------------------------------
C                                       FLOAT TOTAL SAMPLE SIZE
      XTOT=FLOAT(NTOTAL)
C                                       COMPUTE FE AS PRODUCT OF MARGINAL
C                                       TOTALS DIVIDED BY TOTAL SAMPLE
      DO 100 I=1,NROW
      DO  50 J=1,NCOL
      FE(I,J)=FLOAT(NRTOT(I))*FLOAT(NCTOT(J))/XTOT
   50 CONTINUE
  100 CONTINUE
      RETURN
      END



      SUBROUTINE NORMAL(ZIN,SVAL)
C -------------------------------------------------------------------
CCP       CALL:     CALL NORMAL(ZIN,SVAL)
CCP       PURPOSE:  SUBROUTINE NORMAL CALCULATES THE NORMAL PROB
CC                  FOR THE UNIT NORMAL CURVE.  THE INTEGRAL IS EVAL-
CC                  UATED FOR THE AREA BETWEEN ZERO AND THE VALUE OF
CC                  THE GIVEN ORDINATE.  NOTE THAT THE ABSOLUTE VALUE
CC                  OF THE ORDINATE IS USED IN THE COMPUTATIONS.
CCP       ARG-IN:   ZIN - THE NORMAL ORDINATE.
CCP       ARG-OUT:  SVAL - THE NORMAL PROBABILITY.
```

```
CCP        COMMON:  -NONE-
CCP        IO:      -NONE-
CCP        METHOD:  THE METHOD IS A COMBINATION OF LOOKING UP THE ARE
CC                  FOR Z PRIME CLOSEST TO Z AND THEN USING SIMPSONS
CC                  TO CORRECT TO THE EXACT VALUE.  IF THE ABSOLUTE
CC                  OF +ZIN+ IS GREATER THAN 10, THE PROBABILITY IS
CC                  EQUAL TO 0.5.
CCP        NOTE:    TESTS ON THE CDC 6400 INDICATE THAT THE ANSWERS
CC                  ARE ACCURATE TO AT LEAST 9 SIGNIFICANT DIGITS.
CCP        EXTERN:  -NONE-
CCP        LANG:    FORTRAN IV
CCP        AUTHOR:  DAVID L. MARCH, MARCH 1970
C ------------------------------------------------------------------
           DIMENSION Z(9),EVAL(11)
C                                       TABLE OF PROBABILITIES FOR Z OF
CC                                      0.0, 0.5, 1.0, ......., 5.0
           DATA     (EVAL(I),I=1,11)/
      A             0.0,
      B             0.191462461274,
      C             0.341344746069,
      D             0.433192798731,
      E             0.477249868052,
      F             0.493790334674,
      G             0.4986501020,
      H             0.4997673709,
      I             0.4999683288,
      J             0.4999966023,
      K             0.4999997133/
           DATA     E/2.718281828459/
           DATA     ESP/0.39894228040143/
C ------------------------------------------------------------------
           SUP=ABS(ZIN)
           ADD=1.0
           IF(SUP.GT.5.24) GO TO 100
C                                       DETERMINE CLOSEST TABLE VALUE
           J=SUP/0.25
           J=(J+1)/2
C                                       DETERMINE ORDINATE OF CLOSEST TAB
           S=FLOAT(J)*0.5
           A=EVAL(J+1)
C                                       IS INPUT .LT., .EQ., OR .GT.
C                                       TABLE --- TAKE APPROPRIATE ACTION
           IF(SUP-S) 10,20,30
    10     ADD=SUP
           SUP=S
           S=ADD
           ADD=-1.0
           GO TO 30
    20     SVAL=A
           RETURN
C                                       HERE TO INTEGRATE BY SIMPSONS RULE
```

```
   30     H=(SUP-S)/8.0
          DO 40 I=1,9
          Z(I)=1.0/(E**(S*S/2.0))
          S=S+H
   40     CONTINUE
          SVAL=4.0*(Z(2)+Z(4)+Z(6)+Z(8))
     A                +2.0*(Z(3)+Z(5)+Z(7))
     B                +Z(1)+Z(9)
          SVAL=(H/3.0)*SVAL*ESP*ADD+A
          RETURN
  100     SVAL=0.5
          IF(SUP.GT.10.0) RETURN
          S=5.0
          A=EVAL(11)
          GO TO 30
          END



          SUBROUTINE CHISQ(SCHI,NUS,SPCHI)
C ---------------------------------------------------------------
CCP       CALL:     CALL CHISQ(SCHI,NUS,SPCHI)
CCP       PURPOSE:  SUBROUTINE CHISQ CALCULATES THE AREA TO THE RIGHT
CC                  OF CHI-SQUARE UNDER THE CHI-SQUARE DISTRIBUTION
CC                  FUNCTION.
CCP       ARG-IN:   SCHI - THE CHI-SQUARE VALUE
CC                  NUS - THE CORRESPONDING DEGREES OF FREEDOM
CCP       ARG-OUT:  SPCHI - THE CHI-SQUARE PROBABILITY.
CCP       COMMON:   -NONE-
CCP       IO:       -NONE-
CCP       METHOD:   TWO SERIES EXPANSIONS ARE USED, ONE FOR ↓NUS↓
CC                  EVEN AND ONE FOR ↓NUS↓ ODD.  IF ↓NUS↓ EQUALS ONE
CC                  OR TWO THE COMPUTATION IS DIRECT.  FOR ↓NUS↓ GREA
CC                  THAN 30 A CUBE ROOT APPROXIMATION IS USED.
CCP       REF:      EQUATIONS 26.4.4 AND 26.4.5, ABRAMOWITZ AND STEGUN
CC                  *HANDBOOK OF METHEMATICAL FUNCTIONS*, NATIONAL
CC                  BUREAU OF STANDARDS APPLIED MATHEMATICS SERIES 55
CC                  1964.
CCP       EXTERN:   NORMAL - SUBROUTINE TO COMPUTE THE NORMAL
CC                       PROBABILITY.
CCP       LANG:     FORTRAN IV
CCP       AUTHOR:   DAVID L. MARCH, MARCH 1970
C ---------------------------------------------------------------
          DATA      ES/2.71828182846/
          DATA      SPI/0.79788456080286/
C ---------------------------------------------------------------
C                                   CHECK FOR SPECIAL VALUES OF NUS
          IF(NUS.GT.30) GO TO 100
          IF(NUS.EQ.1) GO TO 200
C                                   COMPUTE NORMAL CURVE ORDINATE
```

```
      Z=1.0/(ES**(SCHI/2.0))
      IF(NUS.EQ.2) GO TO 300
C                                     DETERMINE IF NUS IS EVEN OR ODD
C                                     AND SET APPROPRIATE PARAMETERS
      NS=(NUS+1)/2
      MS=NUS/2
      IF(NS.EQ.MS) GO TO 10
      NS=1
      SN=1.0
      GO TO 20
   10 NS=2
      SN=0.0
C                                     PERFORM SERIES EXPANSION
   20 NS=(NUS-NS)/2
      S=0.0
      SZ=1.0
      DO 30 I=1,NS
      ZS=I
      SZ=SZ*(2.0*ZS-SN)
      S=S+(SCHI**ZS)/SZ
   30 CONTINUE
      IF(SN.NE.0.0) GO TO 40
C                                     PERFORM FINAL COMPUTATIONS FOR EV
      SPCHI=Z*(1.0+S)
      RETURN
C                                     PERFORM FINAL COMP FOR ODD NUS
   40 XCHI=SQRT(SCHI)
      S=S*Z/XCHI
      CALL NORMAL(XCHI,VAL)
      SPCHI=SPI*S+2.0*(0.5-VAL)
      RETURN
C ------------------------------------------------------------------
C                                     NUS GREATER THAN 30
C                                     USE CUBE ROOT APPROXIMATION
  100 SZ=NUS
      ZS=2.0/(9.0*SZ)
      XCHI=((SCHI/SZ)**(1.0/3.0)-1.0+ZS)/SQRT(ZS)
      CALL NORMAL(XCHI,VAL)
      IF(XCHI.LT.0.0) GO TO 110
      SPCHI=0.5-VAL
      RETURN
  110 SPCHI=VAL+0.5
      RETURN
C ------------------------------------------------------------------
C                                     DIRECT COMPUTATIONS
C                                     NUS = 1
  200 XCHI=SQRT(SCHI)
      CALL NORMAL(XCHI,VAL)
      SPCHI=2.0*(0.5-VAL)
      RETURN
C                                     NUS = 2
```

```
  300     SPCHI=Z
          RETURN
          END



      SUBROUTINE CALCHI
C -------------------------------------------------------------------
CCP       CALL:     CALL CALCHI
CCP       PURPOSE:  SUBROUTINE CALCHI CALCULATES THE VALUE OF CHI
CC                  SQUARE FOR A CONTINGENCY TABLE AND OBTAINS THE
CC                  CHI SQUARE PROBABILITY.
CCP       ARG-IN:   -NONE- (BUT SEE THE COMMON DESCRIPTION BELOW)
CCP       ARG-OUT:  -NONE- (BUT SEE THE COMMON DESCRIPTION BELOW)
CCP       COMMON:   /MATRIX/ - THE MATRIX OF EXPECTED AND OBSERVED
CC                       AND OTHER PARAMETERS
CC                  /CHICOM/ - THE RESULTS OF THIS SUBROUTINE WHERE:
CC                       XSQ - THE VALUE OF CHI-SQUARE FOR THE TABLE
CC                       PXSQ - THE CORRESPONDING CHI-SQUARE PROB
CC                       NDF - THE NUMBER OF DEGREES OF FREEDOM
CCP       IO:       -NONE-
CCP       METHOD:   THE VALUE OF CHI-SQUARE IS COMPUTED USING THE
CC                  FAMILIAR RELATIONSHIP.  THEN THE DEGREES OF
CC                  FREEDOM ARE COMPUTED AND THE PROBABILITY IS
CC                  FROM SUBROUTINE CHISQ.
CCP       EXTERN:   CHISQ - A SUBROUTINE TO COMPUTE THE CHI-SQUARE
CC                       PROBABILITY
CC                  FLOAT - AN IN-LINE FUNCTION TO FLOAT A NUMBER
CCP       LANG:     FORTRAN IV
CCP       AUTHOR:   DAVID L. MARCH, MAY 1970
C -------------------------------------------------------------------
      COMMON/MATRIX/ NROW,NRTOT(6),NCOL,NCTOT(6)
     A     ,NTOTAL,IFO(3,6),FE(3,6)
      COMMON/CHICOM/ XSQ,PXSQ,NDF
C -------------------------------------------------------------------
C                                     INITIALIZE THE VALUE OF CHI-SQ
      XSQ=0.0
C                                     COMPUTE CHI-SQUARE
      DO 100 I=1,NROW
      DO  50 J=1,NCOL
      XTOP=FLOAT(IFO(I,J))-FE(I,J)
      XSQ=XSQ+XTOP*XTOP/FE(I,J)
   50 CONTINUE
  100 CONTINUE
C                                     COMPUTE NUMBER OF DEGREES OF FREE
      NDF=(NROW-1)*(NCOL-1)
C                                     GET THE CHI-SQUARE PROBABILITY
      CALL CHISQ(XSQ,NDF,PXSQ)
      RETURN
      END
```

```
      FUNCTION FACT(N)
C -------------------------------------------------------------------
CCP       FUNCTION:  FACT(N)
CCP       PURPOSE: FUNCTION FACT RETURNS THE FLOATING POINT FACTORIAL
CC                 OF N WHERE N IS GREATER OR EQUAL TO 0 AND LESS
CC                 THAN OR EQUAL TO 100
CCP       ARG-IN:  N - THE INTEGER FOR WHICH THE FACTORIAL IS DESIRED
CCP       ARG-OUT: FACT - N FACTORIAL
CCP       COMMON:  -NONE-
CCP       IO:      -NONE-
CCP       METHOD:  ON THE FIRST REFERENCE, A TABLE OF FACTORIALS IS
CC                 COMPUTED.  THEREAFTER, THE RESULT IS LOOKED UP IN
CC                 THE TABLE.
CCP       EXTERN:  FLOAT - AN IN-LINE FUNCTION TO FLOAT AN INTEGER
CCP       LANG:    FORTRAN IV
CCP       AUTHOR:  DAVID L. MARCH, MAY 1970
C -------------------------------------------------------------------
      DIMENSION TABLE(101)
      DATA TABLE(1)/1.0/
      DATA TFLAG/0/
C -------------------------------------------------------------------
C                                       WAS THE TABLE GENERATED
      IF(TFLAG.EQ.0) GO TO 100
   10 CONTINUE
C                                       TABLE GENERATED...GET ANSWER
      FACT=TABLE(N+1)
      RETURN
C -------------------------------------------------------------------
C                                       GENERATE THE FACTORIAL TABLE
  100 CONTINUE
      DO 110 I=1,100
      TABLE(I+1)=TABLE(I)*FLOAT(I)
  110 CONTINUE
C                                       SET FLAG...TABLE GENERATED
      TFLAG=1
C                                       GO GET FIRST RESULT
      GO TO 10
      END



      FUNCTION FACTLOG(N)
C -------------------------------------------------------------------
CCP       FUNCTION:  FACTLOG(N)
CCP       PURPOSE: FUNCTION FACTLOG RETURNS THE LOG TO THE BASE 10
CC                 FOR N FACTORIAL WHERE N IS GREATER THAN OR EQUAL
CC                 TO 0 AND LESS THAN OR EQUAL TO 100
CCP       ARG-IN:  N - THE INTEGER FOR WHICH THE LOG FACTORIAL IS
CC                      DESIRED
```

```
CCP       ARG-OUT: FACTLOG - THE LOG TO THE BASE 10 OF N FACTORIAL
CCP       COMMON:  -NONE-
CCP       IO:      -NONE-
CCP       METHOD:  ON THE FIRST REFERENCE, A TABLE OF LOGS OF THE
CC                 FACTORIALS IS COMPUTED.  THEREAFTER, THE RESULT
CC                 IS LOOKED UP IN THE TABLE.
CCP       EXTERN:  FLOAT - AN IN-LINE FUNCTION TO FLOAT A NUMBER
CC                 ALOG10 - A FUNCTION TO COMPUTE THE LOG TO THE
CC                      BASE 10 OF A FLOATING POINT NUMBER.
CC                 FACT - A FUNCTION TO RETURN THE FLOATING POINT
CC                      FACTORIAL OF AN INTEGER
CCP       LANG:    FORTRAN IV
CCP       AUTHOR:  DAVID L. MARCH, MAY 19700
C ------------------------------------------------------------------
      DIMENSION TABLEL(101)
      DATA TFLAG/0/
C ------------------------------------------------------------------
C                                       WAS THE TABLE GENERATED
      IF(TFLAG.EQ.0) GO TO 100
   10 CONTINUE
C                                       TABLE GENERATED...GET ANSWER
      FACTLOG=TABLEL(N+1)
      RETURN
C ------------------------------------------------------------------
C                                       GENERATE THE LOG TABLE
  100 CONTINUE
      DO 110 I=1,101
      X=FACT(I-1)
      TABLEL(I)=ALOG10(X)
  110 CONTINUE
C                                       SET FLAG...TABLE GENERATED
      TFLAG=1
C                                       GO GET FIRST ANSWER
      GO TO 10
      END



      SUBROUTINE CALRXL
C ------------------------------------------------------------------
CCP       CALL:    CALL CALRXL
CCP       PURPOSE: SUBROUTINE CALRXL COMPUTES THE LOG TO THE BASE 10
CC                 THE RX TERM DESCRIBED BY FREEMAN AND HALTON FOR
CC                 OBTAINING THE EXACT PROBABILITY OF A CONTINGENCY
CC                 TABLE.
CCP       ARG-IN:  -NONE- (BUT SEE COMMON DESCRIPTIONS)
CCP       ARG-OUT: -NONE- (BUT SEE COMMON DESCRIPTIONS)
CCP       COMMON:  /EXACT/ - PARAMETERS FOR THE EXACT PROBABILITY
CC                 /MATRIX/ - THE ROW AND COLUMN TOTALS
CCP       METHOD:  RX IS OBTAINED BY ADDING THE LOGS OF THE MARGINAL
```

```
CC                TOTAL FACTORIALS AND SUBTRACTING THE LOG
CC                OF THE TOTAL SAMPLE SIZE FACTORIAL.
CCP       EXTERN:  FACTLOG - A FUNCTION TO RETURN THE LOG TO THE
CC                      BASE 10 OF THE FACTORIAL OF AN INTEGER
CCP       LANG:    FORTRAN IV
CCP       AUTHOR:  DAVID L. MARCH, JUNE 1970
C -------------------------------------------------------------------
      COMMON/EXACT/ RXLOG,EXACTP
      COMMON/MATRIX/ NROW,NRTOT(6),NCOL,NCTOT(6)
     A    ,NTOTAL,IFO(3,6),FE(3,6)
C -------------------------------------------------------------------
C                                       INITIALIZE RESULT TO NEGATIVE OF
C                                       LOG FACTORIAL OF TOTAL SAMPLE SIZE
      RXLOG=-FACTLOG(NTOTAL)
C                                       SUM THE LOGS OF THE ROW TOTALS
      DO 10 I=1,NROW
      RXLOG=RXLOG+FACTLOG(NRTOT(I))
   10 CONTINUE
C                                       ADD THE LOGS OF THE COLUMN TOTALS
      DO 20 I=1,NCOL
      RXLOG=RXLOG+FACTLOG(NCTOT(I))
   20 CONTINUE
      RETURN
      END



      SUBROUTINE SORTER(IGO)
C -------------------------------------------------------------------
CCP       CALL:    CALL SORTER(IGO)
CCP       PURPOSE: SUBROUTINE SORTER STORES THE CELL FREQ AND PROBAB
CC                 FOR SORTING.  IT ALSO DOES THE SORTING WHEN
CC                 PROPERLY CALLED.
CCP       ARG-IN:  IGO - THE ACTION TO BE PERFORMED WHERE:
CC                      1 = INITIALIZE
CC                      2 = STORE ARRAY IFO AND EXACTP
CC                      3 = TAG SORT BASED ON EXACTP
CCP       ARG-OUT: -NONE-
CCP       COMMON:  /MATRIX/ - THE CELL FREQUENCIES
CC                 /EXACT/ - THE EXACT PROBABILITY
CC                 /SORT/ - THE SORT AND STORAGE ARRAYS
CCP       IO:      -NONE-
CCP       METHOD:  THE STORAGE IS STRAIGHT FORWARD ALTHOUGH IT
CC                 SHOULD BE NOTED THAT THE CELL FREQ ARE PACKED INTO
CC                 WORDS.  THE SORT IS A BINARY SORT WITH THE INDEX
CC                 AND PROBABILITIES ONLY BEING MOVED.  THE INDEX IS
CC                 THE LOCATION OF THE PACKED CELL FREQ.
CCP       REF:     SORT -  STUART, F. 1970.  *FORTRAN PROGRAMMING*,
CC                 WILEY AND SONS, PAGE 294
CCP       EXTERN:  PACK - SUBROUTINE TO PACK CELLS INTO 6 WORDS
```

```
CC                  REMARK - SUBROUTINE TO WRITE MESSAGE TO DAYFILE
CC                  ABORT - SUBROUTINE TO ABORT THE JOB
CCP       LANG:     FORTRAN IV
CCP       AUTHOR:   DAVID L. MARCH, JUNE 1970
C ------------------------------------------------------------------
      COMMON/MATRIX/ NROW,NRTOT(6),NCOL,NCTOT(6)
     A     ,NTOTAL,IFO(3,5),FF(3,6)
      COMMON/EXACT/ RXLOG,EXACTP
      COMMON/SORT/ LENGTH,INDEX(2000)
     A     ,PROBS(2000),CELLS(3,2000)
      DATA LENGTH/2000/
      DATA LMAX/2000/
C ------------------------------------------------------------------
C                                     GO TO APPROPRIATE SECTION
      GO TO (100,200,300) IGO
C ------------------------------------------------------------------
  100 CONTINUE
C                                     INITIALIZE INDEX...ONLY PART LAST
C                                     CHANGED
      IF(LENGTH.EQ.0) RETURN
      DO 110 I=1,LENGTH
      INDEX(I)=I
  110 CONTINUE
C                                     SET LENGTH TO ZERO...NO POSITIONS
C                                     FILLED
      LENGTH=0
      RETURN
C ------------------------------------------------------------------
  200 CONTINUE
C                                     STORE CELLS AND PROB...ADD TO LEN
      LENGTH=LENGTH+1
C                                     TEST FOR OVERFLOW
      IF(LENGTH.GT.LMAX) GO TO 250
C                                     STORE PROB
      PROBS(LENGTH)=EXACTP
C                                     PACK AND STORE CELL FREQ
      CALL PACK(IFO,CELLS(1,LENGTH))
      RETURN
C                                     HERE IF OVERFLOW
  250 CONTINUE
      CALL REMARK(18H  ***SORT OVERFLOW )
      CALL REMARK(16H  ***RUN ABORTED )
      CALL ABORT
C ------------------------------------------------------------------
  300 CONTINUE
C                                     MARK END OF LIST
      PROBS(LENGTH+1)=10.0
C                                     DO THE BINARY SORT
      M=LENGTH
  320 M=M/2
      IF(M.EQ.0) RETURN
```

```
      K=LENGTH-M
      J=1
  330 I=J
  340 L=I+M
      IF(PROBS(I).LE.PROBS(L)) GO TO 350
C                                       SWITCH
      A=PROBS(I)
      PROBS(I)=PROBS(L)
      PROBS(L)=A
      N=INDEX(I)
      INDEX(I)=INDEX(L)
      INDEX(L)=N
      I=I-M
      IF(I.GE.1) GO TO 340
  350 J=J+1
      IF(K.LT.J) 320,330
      END
```

```
          IDENT    PACK
          SPACE    2
* ---------------------------------------------------------------------
**P        CALL:    CALL PACK(IN,OUT)
**P        PURPOSE: SUBROUTINE PACK PACKS AN INPUT ARRAY CONTAINING 2
**                  DIGIT BINARY NUMBERS INTO AN OUTPUT ARRAY WHICH
**                  RECEIVE 6 NUMBERS PER WORD.  THE ROUTINE IS
**                  SET TO PACK 18 NUMBERS OF INPUT INTO 3 WORDS
**                  OF OUTPUT.
**P        ARG-IN:  IN - THE INPUT ARRAY WITH A 2 DIGIT NUMBER IN EACH
**                       WORD
**P        ARG-OUT: OUT - THE OUTPUT ARRAY WITH SIX 2 DIGIT NUMBERS IN
**                       EACH WORD.
**P        AUTHOR:  DAVID L. MARCH, JUNE 1970
* ---------------------------------------------------------------------
          SPACE    2
 N.ROW    EQU      3                NUMBER OF ROWS TO BE PACKED
          SPACE    2
          ENTRY    PACK
 PACK     PS
          SB7      N.ROW-1
          SA1      B1-1             SET INPUT ADDRESS
 PL1      BX6      -X6*X6           OUTER LOOP ... ZERO RESULT REG
          SB6      5                INNER LOOP COUNT
 PL2      SA1      A1+1             INNER LOOP ... GET NEXT INPUT
          LX6      10               SHIFT RESULT
          BX6      X1+X6            ADD IN INPUT
          SB6      B6-1             TEST FOR END OF INNER LOOP
          GE       B6,B0,PL2
          SA6      B2               STORE RESULT
```

```
          SB7       B7-1                TEST FOR END OF OUTER LOOP
          LT        B7,B0,PACK
          SB2       B2+1                SET NEXT OUTPUT ADDRESS
          EQ        PL1
          SPACE     2
* ----------------------------------------------------------------------
**P       CALL:     CALL UNPACK(OUT,IN)
**P       PURPOSE:  SUBROUTINE UNPACK IS THE INVERSE OF SUBROUTINE PA
* ----------------------------------------------------------------------
          SPACE     2
          ENTRY     UNPACK
 UNPACK   PS
          MX0       50
          SB7       N.ROW-1
          SA5       B1-1                FUDGE TO SET OUTPUT ADDRESS
          BX6       X5
          SA6       A5
          SA2       B2-1                SET INPUT ADDRESS
 UL1      SA2       A2+1                OUTER LOOP ... GET INPUT WORD
          SB6       5
 UL2      LX2       10                  SHIFT NEXT NUMBER INTO POSITION
          BX6       -X0*X2              MASK NUMBER
          SA6       A6+1                STORE NUMBER
          SB6       B6-1                TEST FOR INNER LOOP DONE
          GE        B6,B0,UL2
          SB7       B7-1                TEST FOR OUTER LOOP DONE
          GE        B7,B0,UL1
          EQ        UNPACK
          END
```

APPENDIX B

SUMMARY REPORTS

The summary reports which follow are the computer printouts from the third stage of the computations. One report was printed for each N considered.

Unfortunately, the first line of each printout is ambiguous. This line should read "SUMMARY FOR i 2 X 3 SETS OF MARGINAL TOTALS WITH N = j."

All possible outcomes (tables) were generated for each of these sets of marginal totals. Thus, for each set of marginal totals, the complete distribution of exact cumulative probabilities was obtained. The total cumulative probability for each of these distributions should be 1.0. The accuracy check line shows the minimum and maximum total cumulative probability computed for the distributions summarized in the report. This indicates that none of the exact distributions deviated from the theoretical sum of 1.0 to the number of digits printed.

The exact probability is represented by the symbol E, the corresponding chi-square probability by the symbol C. MEAN represents the mean absolute percentage error of chi-square.

SUMMARY FOR 1 2X3 TABLES WITH N = 6
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	0	0	0	0.00
.005<P≤.010	0	0	0	0.00
.010<P≤.050	0	0	0	0.00
.050<P≤.100	0	0	0	0.00
.100<P≤.150	0	0	0	0.00
.150<P≤.200	0	0	0	0.00
.200<P	7	7	0	66.38
TOTAL	7	7	0	66.38

CHI-SQUARE PROBABILITIES — EXACT PROBABILITIES

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)							
.005<P≤.010	(2)							
.010<P≤.050	(3)							
.050<P≤.100	(4)							
.100<P≤.150	(5)							6
.150<P≤.200	(6)							
.200<P	(7)							1

	NO.	PERCENT
C<E	6	85.71
C=E	1	14.29
C>E	0	0.00

SUMMARY FOR 2 2X3 TABLES WITH N = 8
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	0	0	0	0.00
.005<P≤.010	0	0	0	0.00
.010<P≤.050	2	2	0	35.90
.050<P≤.100	2	2	0	12.87
.100<P≤.150	0	0	0	0.00
.150<P≤.200	0	0	0	0.00
.200<P	15	15	0	50.36
TOTAL	19	19	0	44.89

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)							
.005<P≤.010	(2)							
.010<P≤.050	(3)			2	2			
.050<P≤.100	(4)							4
.100<P≤.150	(5)							2
.150<P≤.200	(6)							
.200<P	(7)							9

	NO.	PERCENT
C<E	8	42.11
C=E	11	57.89
C>E	0	0.00

SUMMARY FOR 2 2X3 TABLES WITH N = 9
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	0	0	0	0.00
.005<P≤.010	0	0	0	0.00
.010<P≤.050	3	3	0	68.89
.050<P≤.100	0	0	0	0.00
.100<P≤.150	6	6	0	70.00
.150<P≤.200	0	0	0	0.00
.200<P	13	13	0	51.74
TOTAL	22	22	0	59.06

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)							
.005<P≤.010	(2)							
.010<P≤.050	(3)			3		6		
.050<P≤.100	(4)							
.100<P≤.150	(5)							
.150<P≤.200	(6)							3
.200<P	(7)							10

	NO.	PERCENT
C<E	9	40.91
C=E	13	59.09
C>E	0	0.00

SUMMARY FOR 5 2X3 TABLES WITH N = 10
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	1	0	1	41.50
.005<P≤.010	2	2	0	15.10
.010<P≤.050	10	8	2	30.41
.050<P≤.100	8	8	0	56.66
.100<P≤.150	1	1	0	5.18
.150<P≤.200	4	4	0	50.50
.200<P	35	34	1	41.85
TOTAL	61	57	4	41.00

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)							
.005<P≤.010	(2)	1	2					
.010<P≤.050	(3)			10	8			
.050<P≤.100	(4)						4	
.100<P≤.150	(5)					1		10
.150<P≤.200	(6)							
.200<P	(7)							25

	NO.	PERCENT
C<E	22	36.07
C=E	38	62.30
C>E	1	1.64

SUMMARY FOR 4 2X3 TABLES WITH N = 11
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	1	0	1	88.81
.005<P≤.010	2	2	0	32.57
.010<P≤.050	10	6	4	19.98
.050<P≤.100	8	8	0	43.11
.100<P≤.150	5	5	0	28.05
.150<P≤.200	4	4	0	58.21
.200<P	28	28	0	36.17
TOTAL	58	53	5	35.94

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	1	2					
.005<P≤.010	(2)							
.010<P≤.050	(3)			10	7			
.050<P≤.100	(4)				1	5	4	
.100<P≤.150	(5)							6
.150<P≤.200	(6)							2
.200<P	(7)							20

	NO.	PERCENT
C<E	26	44.83
C=E	32	55.17
C>E	0	0.00

SUMMARY FOR 14 2X3 TABLES WITH N = 12
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	8	0	8	36.05
.005<P≤.010	8	3	5	52.28
.010<P≤.050	36	31	5	38.10
.050<P≤.100	24	22	2	33.73
.100<P≤.150	15	15	0	53.68
.150<P≤.200	3	3	0	32.03
.200<P	113	113	0	40.20
TOTAL	207	187	20	40.25

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	6	3					
.005<P≤.010	(2)	2	5					
.010<P≤.050	(3)			36	17	8		
.050<P≤.100	(4)				6	7		12
.100<P≤.150	(5)				1		3	13
.150<P≤.200	(6)							9
.200<P	(7)							79

	NO.	PERCENT
C<E	72	34.78
C=E	132	63.77
C>E	3	1.45

SUMMARY FOR 9 2X3 TABLES WITH N = 13
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	17	4	13	72.34
.005<P≤.010	5	2	3	9.77
.010<P≤.050	28	22	6	22.79
.050<P≤.100	22	21	1	39.14
.100<P≤.150	11	10	1	37.94
.150<P≤.200	7	6	1	32.33
.200<P	70	70	0	32.14
TOTAL	160	135	25	35.45

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	7						
.005<P≤.010	(2)	10	5	4				
.010<P≤.050	(3)			24	14			
.050<P≤.100	(4)				7	10	4	
.100<P≤.150	(5)				1		1	12
.150<P≤.200	(6)					1	2	9
.200<P	(7)							49

	NO.	PERCENT
C<E	54	33.75
C=E	94	58.75
C>E	12	7.50

SUMMARY FOR 22 2X3 TABLES WITH N = 14
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	45	11	34	63.95
.005<P≤.010	13	9	4	36.93
.010<P≤.050	68	49	19	24.60
.050<P≤.100	47	44	3	35.43
.100<P≤.150	33	33	0	32.44
.150<P≤.200	25	25	0	37.76
.200<P	167	167	0	30.70
TOTAL	398	338	60	34.77

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	34	6					
.005<P≤.010	(2)	11	3	5				
.010<P≤.050	(3)		4	62	30			
.050<P≤.100	(4)			1	17	24	13	
.100<P≤.150	(5)					9	9	18
.150<P≤.200	(6)						3	16
.200<P	(7)							133

	NO.	PERCENT
C<E	121	30.40
C=E	261	65.58
C>E	16	4.02

SUMMARY FOR 25 2X3 TABLES WITH N = 15
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	51	12	39	93.45
.005<P≤.010	31	17	14	48.84
.010<P≤.050	83	57	26	24.28
.050<P≤.100	64	60	4	36.41
.100<P≤.150	32	32	0	32.77
.150<P≤.200	22	20	2	28.78
.200<P	209	208	1	31.20
TOTAL	492	406	86	38.27

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	47	11	3				
.005<P≤.010	(2)	4	11	1				
.010<P≤.050	(3)		9	79	39			
.050<P≤.100	(4)				25	26	4	9
.100<P≤.150	(5)					6	12	16
.150<P≤.200	(6)						5	27
.200<P	(7)						1	157

	NO.	PERCENT
C<E	148	30.08
C=E	330	67.07
C>E	14	2.85

SUMMARY FOR 38 2X3 TABLES WITH N = 16
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	101	16	85	97.08
.005<P≤.010	50	25	25	26.48
.010<P≤.050	137	85	52	22.93
.050<P≤.100	96	90	6	30.84
.100<P≤.150	61	56	5	31.67
.150<P≤.200	38	38	0	34.45
.200<P	319	316	3	27.61
TOTAL	802	626	176	36.51

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	86	12					
.005<P≤.010	(2)	15	31	19				
.010<P≤.050	(3)		7	117	57			
.050<P≤.100	(4)			1	39	47	13	
.100<P≤.150	(5)					13	25	24
.150<P≤.200	(6)					1		40
.200<P	(7)							255

	NO.	PERCENT
C<E	237	29.55
C=E	541	67.46
C>E	24	2.99

SUMMARY FOR 37 2X3 TABLES WITH N = 17
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	141	20	121	112.65
.005<P≤.010	39	8	31	29.00
.010<P≤.050	168	105	63	22.76
.050<P≤.100	97	86	11	27.85
.100<P≤.150	57	54	3	30.39
.150<P≤.200	52	51	1	29.35
.200<P	317	314	3	25.90
TOTAL	871	638	233	40.19

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	119						
.005<P≤.010	(2)	22	31	20				
.010<P≤.050	(3)		8	144	41			
.050<P≤.100	(4)			4	55	45	11	
.100<P≤.150	(5)				1	11	30	19
.150<P≤.200	(6)					1	10	40
.200<P	(7)						1	258

	NO.	PERCENT
C<E	206	23.65
C=E	628	72.10
C>E	37	4.25

SUMMARY FOR 66 2X3 TABLES WITH N = 18
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	278	59	219	104.24
.005<P≤.010	84	28	56	30.23
.010<P≤.050	294	201	93	25.49
.050<P≤.100	176	159	17	28.72
.100<P≤.150	93	90	3	29.55
.150<P≤.200	84	77	7	29.09
.200<P	590	572	18	27.03
TOTAL	1599	1186	413	40.78

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	255	11	2				
.005<P≤.010	(2)	21	42	27				
.010<P≤.050	(3)	2	31	261	72	2		
.050<P≤.100	(4)			4	101	57	18	
.100<P≤.150	(5)				3	33	52	46
.150<P≤.200	(6)					1	12	74
.200<P	(7)						2	470

	NO.	PERCENT
C<E	361	22.58
C=E	1174	73.42
C>E	64	4.00

SUMMARY FOR 62 2X3 TABLES WITH N = 19
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	316	50	266	130.37
.005<P≤.010	101	33	68	32.51
.010<P≤.050	306	186	120	23.72
.050<P≤.100	157	139	18	24.63
.100<P≤.150	120	111	9	24.92
.150<P≤.200	77	72	5	26.34
.200<P	563	549	14	22.85
TOTAL	1640	1140	500	44.81

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	282	15					
.005<P≤.010	(2)	34	64	20				
.010<P≤.050	(3)		22	280	66			
.050<P≤.100	(4)			6	91	70	12	
.100<P≤.150	(5)					49	48	30
.150<P≤.200	(6)					1	16	64
.200<P	(7)						1	469

	NO.	PERCENT
C<E	325	19.82
C=E	1251	76.28
C>E	64	3.90

SUMMARY FOR 97 2X3 TABLES WITH N = 20
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	557	92	465	126.60
.005<P≤.010	152	71	81	34.14
.010<P≤.050	488	321	167	22.34
.050<P≤.100	260	218	42	23.67
.100<P≤.150	171	161	10	26.72
.150<P≤.200	131	126	5	28.17
.200<P	900	873	27	22.89
TOTAL	2659	1862	797	45.74

CHI-SQUARE PROBABILITIES / EXACT PROBABILITIES

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	495	24					
.005<P≤.010	(2)	61	88	32				
.010<P≤.050	(3)	1	40	440	82			
.050<P≤.100	(4)			16	170	112	18	
.100<P≤.150	(5)				8	55	88	34
.150<P≤.200	(6)					4	24	116
.200<P	(7)						1	750

	NO.	PERCENT
C<E	506	19.03
C=E	2022	76.04
C>E	131	4.93

SUMMARY FOR 99 2X3 TABLES WITH N = 21
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	671	116	555	145.70
.005<P≤.010	181	61	120	32.40
.010<P≤.050	531	318	213	23.63
.050<P≤.100	287	238	49	23.70
.100<P≤.150	192	179	13	25.65
.150<P≤.200	129	121	8	23.23
.200<P	968	941	27	22.58
TOTAL	2959	1974	985	51.63

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	621	27					
.005<P≤.010	(2)	48	109	34				
.010<P≤.050	(3)	2	45	484	91			
.050<P≤.100	(4)			13	190	120	17	1
.100<P≤.150	(5)				6	69	79	51
.150<P≤.200	(6)					3	31	109
.200<P	(7)						2	807

	NO.	PERCENT
C<E	529	17.88
C=E	2311	78.10
C>E	119	4.02

SUMMARY FOR 133 2X3 TABLES WITH N = 22
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	1007	158	849	149.43
.005<P≤.010	247	80	167	28.48
.010<P≤.050	758	463	295	22.21
.050<P≤.100	388	316	72	21.65
.100<P≤.150	252	230	22	23.87
.150<P≤.200	223	208	15	24.96
.200<P	1309	1266	43	21.35
TOTAL	4184	2721	1463	53.12

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	918	25	4				
.005<P≤.010	(2)	88	154	61				
.010<P≤.050	(3)	1	68	669	117			
.050<P≤.100	(4)			24	267	159	23	2
.100<P≤.150	(5)				4	89	138	35
.150<P≤.200	(6)					4	58	152
.200<P	(7)						4	1120

	NO.	PERCENT
C<E	716	17.11
C=E	3275	78.27
C>E	193	4.61

SUMMARY FOR 143 2X3 TABLES WITH N = 23
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	1225	186	1039	168.25
.005<P≤.010	287	112	175	32.29
.010<P≤.050	819	475	344	22.72
.050<P≤.100	466	374	92	20.96
.100<P≤.150	287	255	32	22.54
.150<P≤.200	208	194	14	21.39
.200<P	1441	1386	55	19.69
TOTAL	4733	2982	1751	59.80

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	1099	34	2				
.005<P≤.010	(2)	123	178	47				
.010<P≤.050	(3)	3	75	745	142	2		
.050<P≤.100	(4)			25	314	140	11	4
.100<P≤.150	(5)				10	137	126	43
.150<P≤.200	(6)					8	67	162
.200<P	(7)						4	1232

	NO.	PERCENT
C<E	713	15.06
C=E	3772	79.70
C>E	248	5.24

SUMMARY FOR 203 2X3 TABLES WITH N = 24
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	1875	303	1572	165.52
.005<P≤.010	388	158	230	33.03
.010<P≤.050	1206	708	498	22.92
.050<P≤.100	663	536	127	22.71
.100<P≤.150	407	370	37	22.57
.150<P≤.200	318	285	33	22.92
.200<P	2105	1998	107	20.64
TOTAL	6962	4358	2604	61.16

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	1730	60					
.005<P≤.010	(2)	144	220	78				
.010<P≤.050	(3)	1	108	1094	220			
.050<P≤.100	(4)			34	423	232	27	
.100<P≤.150	(5)				20	169	196	80
.150<P≤.200	(6)					5	85	206
.200<P	(7)					1	10	1819

	NO.	PERCENT
C<E	1099	15.79
C=E	5540	79.57
C>E	323	4.64

SUMMARY FOR 194 2X3 TABLES WITH N = 25
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	2057	321	1736	193.20
.005<P≤.010	428	151	277	30.84
.010<P≤.050	1253	693	560	22.46
.050<P≤.100	649	513	136	19.84
.100<P≤.150	426	386	40	21.00
.150<P≤.200	309	279	30	20.69
.200<P	2074	1991	83	18.46
TOTAL	7196	4334	2862	70.21

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	1899	55	6				
.005<P≤.010	(2)	156	273	61				
.010<P≤.050	(3)	2	100	1133	184			
.050<P≤.100	(4)			53	448	219	10	
.100<P≤.150	(5)				17	199	191	46
.150<P≤.200	(6)					8	101	230
.200<P	(7)						7	1798

	NO.	PERCENT
C<E	1002	13.92
C=E	5851	81.31
C>E	343	4.77

SUMMARY FOR 255 2X3 TABLES WITH N = 26
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	2950	371	2579	197.27
.005<P≤.010	589	209	380	32.33
.010<P≤.050	1705	965	740	21.59
.050<P≤.100	890	694	196	19.58
.100<P≤.150	551	489	62	20.33
.150<P≤.200	406	369	37	21.15
.200<P	2817	2668	149	18.57
TOTAL	9908	5765	4143	73.41

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	2733	75	3				
.005<P≤.010	(2)	211	348	93				
.010<P≤.050	(3)	6	166	1549	222			
.050<P≤.100	(4)			60	639	274	34	2
.100<P≤.150	(5)				29	264	246	68
.150<P≤.200	(6)					12	116	333
.200<P	(7)					1	10	2414

	NO.	PERCENT
C<E	1350	13.63
C=E	8063	81.38
C>E	495	5.00

SUMMARY FOR 278 2X3 TABLES WITH N = 27
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	3482	551	2931	213.57
.005<P≤.010	661	245	416	34.46
.010<P≤.050	1905	1060	845	22.87
.050<P≤.100	1001	786	215	19.64
.100<P≤.150	621	537	84	19.91
.150<P≤.200	453	399	54	19.57
.200<P	3106	2956	150	17.92
TOTAL	11229	6534	4695	80.73

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	3250	91	1				
.005<P≤.010	(2)	226	397	100				
.010<P≤.050	(3)	6	173	1737	271	4		
.050<P≤.100	(4)			67	699	312	16	
.100<P≤.150	(5)				31	284	286	75
.150<P≤.200	(6)					21	137	317
.200<P	(7)						14	2714

	NO.	PERCENT
C<E	1473	13.12
C=E	9218	82.09
C>E	538	4.79

SUMMARY FOR 345 2X3 TABLES WITH N = 28
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	4684	624	4060	223.41
.005<P≤.010	853	316	537	31.06
.010<P≤.050	2466	1368	1098	22.01
.050<P≤.100	1260	961	299	18.74
.100<P≤.150	794	668	126	19.72
.150<P≤.200	608	546	62	20.02
.200<P	3920	3685	235	17.37
TOTAL	14585	8168	6417	85.48

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	4357	106					
.005<P≤.010	(2)	316	535	168				
.010<P≤.050	(3)	11	212	2220	294	2		
.050<P≤.100	(4)			78	919	373	28	
.100<P≤.150	(5)				47	391	354	72
.150<P≤.200	(6)					26	210	405
.200<P	(7)					2	16	3443

	NO.	PERCENT
C<E	1802	12.36
C=E	12075	82.79
C>E	708	4.85

SUMMARY FOR 355 2X3 TABLES WITH N = 29
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	5287	719	4568	247.80
.005<P≤.010	942	322	620	34.15
.010<P≤.050	2638	1419	1219	22.43
.050<P≤.100	1318	984	334	17.68
.100<P≤.150	890	765	125	18.45
.150<P≤.200	618	543	75	18.60
.200<P	4131	3875	256	16.29
TOTAL	15824	8627	7197	96.05

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	4964	108	2				
.005<P≤.010	(2)	315	572	148				
.010<P≤.050	(3)	8	262	2366	281			
.050<P≤.100	(4)			122	999	375	12	2
.100<P≤.150	(5)				38	493	367	62
.150<P≤.200	(6)					21	221	382
.200<P	(7)					1	18	3685

	NO.	PERCENT
C<E	1739	10.99
C=E	13300	84.05
C>E	785	4.96

SUMMARY FOR 456 2X3 TABLES WITH N = 30

ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	7248	1062	6186	251.56
.005<P≤.010	1216	442	774	31.97
.010<P≤.050	3444	1874	1570	21.90
.050<P≤.100	1753	1316	437	18.89
.100<P≤.150	1130	944	186	19.11
.150<P≤.200	804	715	89	19.07
.200<P	5449	5091	358	16.96
TOTAL	21044	11444	9600	99.80

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	6808	155	12				
.005<P≤.010	(2)	426	719	184				
.010<P≤.050	(3)	14	342	3129	439	2		
.050<P≤.100	(4)			119	1249	525	54	
.100<P≤.150	(5)				64	561	424	113
.150<P≤.200	(6)				1	38	300	565
.200<P	(7)					4	26	4771

	NO.	PERCENT
C<E	2473	11.75
C=E	17537	83.33
C>E	1034	4.91

SUMMARY FOR 878 2X3 TABLES WITH N = 36
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	21066	2905	18161	378.85
.005<P≤.010	2932	1032	1900	32.40
.010<P≤.050	7923	4146	3777	21.60
.050<P≤.100	3981	2864	1117	16.76
.100<P≤.150	2477	2012	465	16.21
.150<P≤.200	1832	1584	248	16.95
.200<P	11765	10736	1029	14.50
TOTAL	51976	25279	26697	164.61

CHI-SQUARE PROBABILITIES — EXACT PROBABILITIES

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	20021	371	15				
.005<P≤.010	(2)	1003	1749	406				
.010<P≤.050	(3)	42	812	7176	793	10		
.050<P≤.100	(4)			326	3016	933	55	
.100<P≤.150	(5)				172	1435	917	139
.150<P≤.200	(6)					98	797	1023
.200<P	(7)					1	63	10603

	NO.	PERCENT
C<E	4662	8.97
C=E	44797	86.19
C>E	2517	4.84

SUMMARY FOR 1516 2X3 TABLES WITH N = 42
ALL EXPECTED CELL FREQUENCIES WERE ≥ 1

ACCURACY CHECK: 1.00000000 ≤ SUM P(E) ≤ 1.00000000

RANGE E	NO.	C<E	C≥E	MEAN
.000≤P≤.005	50706	7265	43441	585.31
.005<P≤.010	5991	2150	3841	32.33
.010<P≤.050	15885	7986	7899	21.34
.050<P≤.100	7832	5374	2458	15.64
.100<P≤.150	4843	3782	1061	14.86
.150<P≤.200	3540	2926	614	14.76
.200<P	22528	20138	2390	12.97
TOTAL	111325	49621	61704	276.22

CHI-SQUARE PROBABILITIES		EXACT PROBABILITIES (1)	(2)	(3)	(4)	(5)	(6)	(7)
.000≤P≤.005	(1)	48520	678	30				
.005<P≤.010	(2)	2112	3752	825				
.010<P≤.050	(3)	74	1561	14352	1352			
.050<P≤.100	(4)			678	6114	1646	50	
.100<P≤.150	(5)				365	2970	1559	171
.150<P≤.200	(6)				1	218	1766	1828
.200<P	(7)					9	165	20529

	NO.	PERCENT
C<E	8139	7.31
C=E	98003	88.03
C>E	5183	4.66

APPENDIX C

SAMPLE PRINTOUTS OF STAGE 2 COMPUTATIONS

2X3 TABLE WITH N = 10
ROW TOTALS = 5 5
COL TOTALS = 5 3 2

EXPECTED FREQUENCIES: 2.500 1.500 1.000
2.500 1.500 1.000

	ROW 1			ROW 2			EXACTP	SUMP	XSQ	PCHI
1	0	3	2	5	0	0	3.968E-03	7.937E-03	1.000E+01	6.738E-03
2	5	0	0	0	3	2	3.968E-03	7.937E-03	1.000E+01	6.738E-03
3	4	0	1	1	3	1	3.968E-02	1.667E-01	4.800E+00	9.072E-02
4	1	3	1	4	0	1	3.968E-02	1.667E-01	4.800E+00	9.072E-02
5	3	0	2	2	3	0	3.968E-02	1.667E-01	5.200E+00	7.427E-02
6	2	3	0	3	0	2	3.968E-02	1.667E-01	5.200E+00	7.427E-02
7	1	2	2	4	1	0	5.952E-02	2.857E-01	4.133E+00	1.266E-01
8	4	1	0	1	2	2	5.952E-02	2.857E-01	4.133E+00	1.266E-01
9	2	1	2	3	2	0	1.190E-01	5.238E-01	2.533E+00	2.818E-01
10	3	2	0	2	1	2	1.190E-01	5.238E-01	2.533E+00	2.818E-01
11	2	2	1	3	1	1	2.381E-01	1.000E+00	5.333E-01	7.659E-01
12	3	1	1	2	2	1	2.381E-01	1.000E+00	5.333E-01	7.659E-01

2X3 TABLE WITH N = 15
ROW TOTALS = 8 7
COL TOTALS = 5 5 5

EXPECTED FREQUENCIES: 2.667 2.667 2.667
2.333 2.333 2.333

	ROW 1			ROW 2			EXACTP	SUMP	XSQ	PCHI
1	5	3	0	0	2	5	1.554E-03	9.324E-03	1.018E+01	6.162E-03
2	3	5	0	2	0	5	1.554E-03	9.324E-03	1.018E+01	6.162E-03
3	3	0	5	2	5	0	1.554E-03	9.324E-03	1.018E+01	6.162E-03
4	5	0	3	0	5	2	1.554E-03	9.324E-03	1.018E+01	6.162E-03
5	0	3	5	5	2	0	1.554E-03	9.324E-03	1.018E+01	6.162E-03
6	0	5	3	5	0	2	1.554E-03	9.324E-03	1.018E+01	6.162E-03
7	4	0	4	1	5	1	3.885E-03	2.098E-02	8.571E+00	1.376E-02
8	0	4	4	5	1	1	3.885E-03	2.098E-02	8.571E+00	1.376E-02
9	4	4	0	1	1	5	3.885E-03	2.098E-02	8.571E+00	1.376E-02
10	2	1	5	3	4	0	7.770E-03	6.760E-02	6.964E+00	3.074E-02
11	5	2	1	0	3	4	7.770E-03	6.760E-02	6.964E+00	3.074E-02
12	1	2	5	4	3	0	7.770E-03	6.760E-02	6.964E+00	3.074E-02
13	5	1	2	0	4	3	7.770E-03	6.760E-02	6.964E+00	3.074E-02
14	2	5	1	3	0	4	7.770E-03	6.760E-02	6.964E+00	3.074E-02
15	1	5	2	4	0	3	7.770E-03	6.760E-02	6.964E+00	3.074E-02
16	1	3	4	4	2	1	3.885E-02	3.007E-01	3.750E+00	1.534E-01
17	4	3	1	1	2	4	3.885E-02	3.007E-01	3.750E+00	1.534E-01
18	1	4	3	4	1	2	3.885E-02	3.007E-01	3.750E+00	1.534E-01
19	3	4	1	2	1	4	3.885E-02	3.007E-01	3.750E+00	1.534E-01
20	3	1	4	2	4	1	3.885E-02	3.007E-01	3.750E+00	1.534E-01
21	4	1	3	1	4	2	3.885E-02	3.007E-01	3.750E+00	1.534E-01
22	2	2	4	3	3	1	7.770E-02	5.338E-01	2.143E+00	3.425E-01
23	4	2	2	1	3	3	7.770E-02	5.338E-01	2.143E+00	3.425E-01
24	2	4	2	3	1	3	7.770E-02	5.338E-01	2.143E+00	3.425E-01
25	2	3	3	3	2	2	1.554E-01	1.000E+00	5.357E-01	7.650E-01
26	3	2	3	2	3	2	1.554E-01	1.000E+00	5.357E-01	7.650E-01
27	3	3	2	2	2	3	1.554E-01	1.000E+00	5.357E-01	7.650E-01

BIBLIOGRAPHY

Abramowitz, M. and Stengun, I. 1964. Handbook of Mathematical Functions, National Bureau of Standards Applied Mathematics Series 55. Equations 26.4.4 and 26.4.5.

Armsen, P. 1955. Tables for significance tests of 2 X 2 tables. Biometrika, 42:494-511.

Barnard, G. A. 1947. Significance tests for 2 X 2 tables. Biometrika, 34:123-138.

Burford, R. L. 1968. Statistics: A Computer Approach. Columbus, Ohio: Charles E. Merrill Publishing Company.

Cochran, W. G. 1952. The X^2 test of goodness of fit. Annals of Mathematical Statistics, 23:315-345.

Cochran, W. G. 1954. Some methods for strengthening the common X^2 tests. Biometrics, 10:417-451.

Cramer, H. 1946. Mathematical Methods for Statistics. Princeton: Princeton University Press.

El Shanawany, M. R. 1936. An illustration of the accuracy of the chi-square approximation. Biometrika, 28:179-187.

Ferguson, G. A. 1966. Statistical Analysis in Psychology and Education. New York: McGraw Hill.

Finney, D. J. 1948. The Fisher-Yates test of significance in 2 X 2 contingency tables. Biometrika, 34:145-156.

Freeman, G. H. and Halton, J. H. 1951. Note on an exact treatment of contingency, goodness of fit, and other problems of significance. Biometrika, 38:141-149.

Latscha, R. 1953. Tests of significance in a 2 X 2 contingency table: extension of Finney's table. Biometrika, 40:74-86.

Leslie, P. H. 1955. A simple method of calculating the exact probability in 2 X 2 contingency tables with small marginal totals. Biometrika, 42:522-523.

Lewontin, R. C. and Felsenstein, J. 1965. The robustness of homogeneity tests in 2 X N tables. Biometrics, 21:19-33.

Leyton, M. K. 1968. Rapid calculation of exact probabilities for 2 X 3 contingency tables. Biometrics, 24:714-717.

Maxwell, A. E. 1961. Analysing Qualitative Data. London: Methuen and Co. Ltd.

Nass, C. A. G. 1959. The X^2 test for small expectations in contingency tables, with special reference to accidents and absenteeism. Biometrika, 46:365-385.

Neyman, J. and Pearson, E. S. 1931. Further notes on the chi-square distribution. Biometrika, 22:298-305.

Pearson, E. S. 1947. The choice of statistical tests illustrated on the interpretation of data classified in a 2 X 2 table. Biometrika, 34:139-145.

Pearson, K. 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine Series, 50:157-175.

Roscoe, J. T. and Byars, J. A. 1970. Some notes on the chi-square approximation of the multinomial and selected alternatives. Paper presented at the annual convention of the American Research Association, Minneapolis, March, 1970.

Slakter, M. J. 1966. Comparative validity of the chi-square and two modified chi-square goodness-of-fit tests for small but equal expected frequencies. Biometrika, 53:619-622.

Stuart, F. 1970. Fortran Programming. New York: John Wiley & Sons, Inc.

Tate, M. W. 1965. Statistics in Education and Psychology. New York: The Macmillan Company.

Tate, M. W. and Clelland, R. C. 1957. Nonparametric and Shortcut Statistics. Danville, Ill.: Interstate Printers and Publishers.

Tate, M. W. and Hyer, L. A. 1969. Significance values for an exact multinomial test and accuracy of the chi-square approximation. U. S. Department of Health, Education, and Welfare grant No. OEG-0-8-080023-3696(010), project No. 8-B-023

Wise, M. E. 1963. Multinomial probabilities and the chi-square and X^2 distributions. Biometrika, 50:145-154.

VITA

David L. March was born on November 26, 1942, in Harrisburg, Pennsylvania, the son of Clair R. and Mary E. March. He received his elementary and secondary education in the Camp Hill (Pennsylvania) School District.

In 1960, Mr. March entered Lehigh University as a major in General Science and Mathematics. He was graduated in 1964 with a B.S. degree. Mr. March remained at Lehigh to enter a secondary education teaching intern program. As part of this program, he taught ninth grade general mathematics for one semester at Harrison-Morton Junior High School in Allentown, Pennsylvania.

After receiving his M.A. in Education from Lehigh in 1965, Mr. March was employed by the Industrial Management Institute of Reading (Pennsylvania) as an instructor in mathematics, physics, and strength of materials. He was promoted to assistant director of IMI in the spring of 1967.

In 1968, Mr. March began a twenty-four month fellowship in the Educational Research Training Program at Lehigh University. He also worked, on a part-time basis, for the Lehigh University Computing Center where he is presently employed full-time.

Mr. March is a member of the American Educational Research Association, the American Statistical Association, and the Association for Computing Machinery.