5-1966

Tests of Independence in Contingency Tables

Su-feng Wongbhan

Utah State University

TESTS OF INDEPENDENCE IN CONTINGENCY TABLES

by

Su-feng Wongbhan

A report submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Applied Statistics

Plan B

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1966

TABLE OF CONTENTS

Page

I. INTRODUCTION

This report is a survey of the literature for a combination of different tests for both two-way and multi-way tests of independence in contingency tables. The derivation of the commonly used chi-square statistic for the tests will be shown immediately below, then followed by the summary of the different tests which will be the contents of this report.

Consider a population classified according to the presence or absence of an attribute, A . The simplest kind of problem in interdependence arises when there are two attributes, A, B, and if we denote the absence of A by $\bar{A}$ and the absence of B by $\bar{B}$, the numbers falling into the four possible sub-groups may, in an obvious notation, be represented by

	B	not-B	Totals
A	(AB)	$(A\bar{B})$	(A)
not-A	$(\bar{A}B)$	$(\bar{A}\bar{B})$	$(\bar{A})$
Totals	(B)	$(\bar{B})$	n

. . . . . [1.1]

Write this 2 x 2 table (sometimes called a fourfold table) in the following form:

a	b	a + b
c	d	c + d
a + c	b + d	n

. . . . . [1.2]

If there is no association between A and B , that is to say, if the possession of A is irrelevant to the possession of B , there must be the same proportion of A's among the B's as among the not-B's. Thus, by definition, the attributes are "independent" in this set of n observations if

$$\frac{a}{a+c} = \frac{b}{b+d} = \frac{a+b}{n} .$$

Suppose that in the parent population the true probabilities corresponding to the frequencies a, b, c, d are p_{11}, p_{12}, p_{21}, p_{22} respectively. We write the probabilities with

p_{11}	p_{12}	$p_{1.}$
p_{21}	p_{22}	$p_{2.}$
$p_{.1}$	$p_{.2}$	1

$p_{1.} = p_{11} + p_{12}$, and so forth. We suppose the observations drawn with replacement from the population (or, equivalently that the parent population is infinite).

We also rewrite the table [1.2] in the notationally symmetrical form

n_{11}	n_{12}	$n_{1.}$
n_{21}	n_{22}	$n_{2.}$
$n_{.1}$	$n_{.2}$	n

The distribution of the sample frequencies is given by the multinomial whose general term is

$$L = \frac{n!}{n_{11}!\, n_{12}!\, n_{21}!\, n_{22}!}\, p_{11}^{n_{11}}\, p_{12}^{n_{12}}\, p_{21}^{n_{21}}\, p_{22}^{n_{22}} .$$

Now suppose we wish to test the hypothesis of independence in the 2 x 2 table, which is

$$H_o : p_{11}\, p_{22} = p_{12}\, p_{21} .$$

This hypothesis is composite, imposing one constraint, and having two degrees of freedom. We allow p_{11} and p_{12} to vary and express p_{21} and p_{22} by

$$p_{21} = \frac{p_{11}\,(1-p_{11} - p_{12})}{p_{11} + p_{12}}, \quad p_{22} = \frac{p_{12}\,(1 - p_{11} - p_{12})}{p_{11} + p_{12}} \qquad \ldots\ldots\ldots [1.3]$$

The logarithm of the Likelihood Function is therefore, neglecting constants,

$$\begin{aligned} \log L &= n_{11} \log p_{11} + n_{12} \log p_{12} + n_{21} \log p_{21} + n_{22} \log p_{22} \\ &= n_{11} \log p_{11} + n_{12} \log p_{12} + n_{21} \{\log p_{11} + \log (1-p_{11} \\ &\quad -p_{12}) - \log(p_{11} + p_{12}) \} + n_{22} \{\log p_{12} + \log (1-p_{11} \\ &\quad -p_{12}) - \log(p_{11} + p_{12}) \} \\ &= n_{.1} \log p_{11} - n_{.2} \log p_{12} + n_{2.} \{\log(1-p_{1.}) - \log p_{1.}\} . \end{aligned}$$

To estimate the parameters, we put

$$0 = \frac{\partial \log L}{\partial p_{11}} = \frac{n_{.1}}{p_{11}} - n_{2.}\left[\frac{1}{1-p_{1.}} + \frac{1}{p_{1.}}\right] = \frac{n_{.1}}{p_{11}} - \frac{n_2}{p_{1.}(1-p_{1.})},$$

$$0 = \frac{\partial \log L}{\partial p_{12}} = \frac{n_{.2}}{p_{12}} - \frac{n_2}{p_{1.}(1-p_{1.})},$$

giving for the ML estimators under H_o

$$\hat{p}_{11} = \frac{n_{1.}}{n}\,\frac{n_{.1}}{n}, \quad \hat{p}_{12} = \frac{n_{1.}}{n}\,\frac{n_{.2}}{n}. \quad \ldots \quad [1.3a]$$

[1.3] gives analogous expressions for $\hat{p}_{21}$ and $\hat{p}_{22}$. Thus we estimate the cell probabilities from the products of the proportional marginal frequencies.

Substituting these ML estimators into the Likelihood Function we have

$$L(n_{ij} \mid H_o, \hat{p}_{ij}) \propto (n_{1.}\, n_{.1})^{n_{11}} (n_{1.}\, n_{.2})^{n_{12}} (n_{2.}\, n_{.1})^{n_{21}} (n_{2.}\, n_{.2})^{n_{22}} / n^{2n} \quad \ldots \quad [1.4]$$

while the unconditional maximum of the LF is obtained by inserting the estimators [1.3a] to obtain

$$L(n_{ij} \mid \hat{p}_{ij}) \propto n_{11}^{n_{11}}\, n_{12}^{n_{12}}\, n_{21}^{n_{21}}\, n_{22}^{n_{22}} / n^n \quad \ldots \quad [1.5]$$

[1.4-5] gives for the Likelihood-Ratio test statistic

$$\ell = \left(\frac{n_{1\cdot}n_{\cdot 1}}{nn_{11}}\right)^{n_{11}} \left(\frac{n_{1\cdot}n_{\cdot 2}}{nn_{12}}\right)^{n_{12}} \left(\frac{n_{2\cdot}n_{\cdot 1}}{nn_{21}}\right)^{n_{21}}$$

$$\left(\frac{n_{2\cdot}n_{\cdot 2}}{nn_{22}}\right)^{n_{22}} .$$

Writing $n\hat{p}_{ij} = n_{i\cdot}n_{\cdot j}/n = m_{ij}$, this becomes

$$\ell = \left(\frac{m_{11}}{n_{11}}\right)^{n_{11}} \left(\frac{m_{12}}{n_{12}}\right)^{n_{12}} \left(\frac{m_{21}}{n_{21}}\right)^{n_{21}} \left(\frac{m_{22}}{n_{22}}\right)^{n_{22}} .$$

It can be shown that that $-2\log\ell$ is asymptotically distributed as χ^2 with one degree of freedom. Writing $D_{ij} = n_{ij} - m_{ij}$, and expanding as far as $D^2 (= D_{ij}^2$, all i, j), we have

$$-2\log\ell = 2 \sum_{i=1}^{2} \sum_{j=1}^{2} m_{ij}\left(1 + \frac{D_{ij}}{m_{ij}}\right)\left(\frac{D_{ij}}{m_{ij}} - \frac{D^2}{2m_{ij}^2}\right)$$

$$= D^2 \sum_i \sum_j \frac{1}{m_{ij}} \qquad \ldots\ldots\ldots \qquad [1.6]$$

[1.6] may be rewritten

$$-2\log\ell = \sum_i \sum_j \frac{(n_{ij} - m_{ij})^2}{m_{ij}} \equiv X^2 \qquad \ldots\ldots \qquad [1.7]$$

[1.7] is the general statistic for testing the independence in contingency tables. To apply [1.7] it should be assumed that the sample size is large. Note that [1.7] is just an approximate statistic. See reference (8) for the above derivation.

For two-way tables, Kendall and Stuart (8) give exact tests according three cases: both margins fixed, one margin fixed, and no margin fixed, for the 2 x 2 tables; Snedecor (19) devised a special method for R x 2 tables; Ostle (15) mentions a special approximate method and its continuity correction for 2 x 2 tables. These will be covered in Section 2 of this report.

The study of independence in a three-way contingency table was initiated in 1935 by Bartlett (2) who discussed the analysis of a 2 x 2 x 2 table and a 2 x 2 x 3 table. Bartlett, following a suggestion of Fisher (4), presented a test of the null hypothesis that three-factor interaction is zero. To apply this test to a 2 x 2 x 2 table the user must solve a cubic equation in one unknown. For a more complex three-way table, namely the 2 x 2 x t, the estimation problem was shown to involve the solution of (t-1) simultaneous third-degree equations in as many unknowns. Norton (14) demonstrated a rather neat iterative procedure for solving these systems of equations. More recently, Roy and Kastenbaum (17) extended the no three-factor interaction hypothesis to the general three-way contingency table, and in doing so found that the estimation of the parameters in this case involved the solution of (r-1)(c-1)(t-1) simultaneous third-degree equations in as many unknowns.

More recently, two quite simple methods of testing the null hypothesis H_0 of zero three factor interaction have been proposed for the 2 x 2 x t table. A test based upon an analysis of the log-frequencies has been presented by Plackett (16) using a criterion suggested earlier by Woolf (21), and a test based on an analysis of the observed frequencies has been presented by Goodman (5). The test statistics for both these tests can be computed explicitly and simply. (No iterative procedures are needed). When H_0 is ture, the asymptotic distribution of these test statistics is the chi-square distribution with t-1 degrees of freedom.

The test proposed by Plackett (16) for the hypothesis of zero second-order interaction in an r x c x t congency table is, in some respects, the simplest valid test of this hypothesis yet presented in the statistical literature. Based on a transformation matrix, the rows of which are orthogonal to each other, he proposed a test which requires the inversion of t-1 matrices, each having (r-1)(c-1) rows and (r-1)(c-1) columns. A modification of Plackett's method presented by Goodman (6) will require the inversion of only one matrix of size (r-1)(c-1) and t matrices each of size (u-1), where u = [min r, c]. Hence the Goodman's test will be easier to apply than the other valid tests of this null hypothesis. Section 3 will cover the above three-way analysis.

The extensions to higher-order interactions are merely outlined by Lewis (13) and Darroch (3), as they are not likely to be of interest. This will be mentioned in Section 4.

For the partition of chi-square, Lancaster (12) showed that a contingency table can be split up exactly into single degree of freedom when expected frequencies are estimated from the marginal totals. Kimball (10) gave short-cut formulas to simplify Lancaster's formulas. The explanation of these for the case of 2 x 3 tables will be in Section 2. 6. The generalization can be found in references (12) and (10).

In Section 5, the interactions in contingency tables are compared with interactions in the analysis of variance by Darroch (3).

2. TWO-WAY TABLES

2.1 The General R x C Tables (8)

Suppose n randomly selected items are classified according to two different criteria. The results could be presented as in [2.1]. Where n_{ij} represents the number of items belonging to the (ij)th cell of the r x c table, and

$$n_{i\cdot} = \sum_{j=1}^{c} n_{ij} , \quad n_{\cdot j} = \sum_{i=1}^{r} n_{ij} , \quad n = \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} .$$

n_{11}	n_{12}	. . .	n_{1c}	$n_{1\cdot}$		
n_{21}	n_{22}	. . .	n_{2c}	$n_{2\cdot}$		
.	.		.	.		
.	.		.	.		[2.1]
.	.		.	.		
n_{r1}	n_{r2}	. . .	n_{rc}	$n_{r\cdot}$		
$n_{\cdot 1}$	$n_{\cdot 2}$	. . .	$n_{\cdot c}$	n		

In [2.1], if the two variables were independent, the frequency in the ith row and the jth column would be $n_{i\cdot} n_{\cdot j}/n$. The deviation from independence in that particular cell of the table is, therefore, measured by

$$D_{ij} = n_{ij} - n_{i\cdot} n_{\cdot j}/n = n_{ij} - m_{ij} ,$$

where

$$m_{ij} = n_{i\cdot} n_{\cdot j}/n .$$

We may define a coefficient of association* in terms of the so-called square contingency $\sum_{i,j} D_{ij}^2/(n_{i\cdot} n_{\cdot j})$, and shall write**

$$X^2 = \sum_{i,j} \frac{D_{ij}^2}{n_{i\cdot} n_{\cdot j}/n} = \sum_{i,j} \frac{(n_{ij} - m_{ij})^2}{m_{ij}} \equiv n \left[\sum_{i,j} \frac{n_{ij}^2}{n_{i\cdot} n_{\cdot j}} - 1 \right] \qquad \ldots\ldots\ldots \quad [2.2]$$

On the hypothesis of independence, χ^2 is asymptotically distributed in the χ^2 form, if the sample size is sufficiently large. The degrees of freedom are given by $(rc-1) - (r-1) - (c-1) = (r-1)(c-1)$, the number of classes minus 1 minus the number of parameters fitted.

---

*"Association" here means the interdependence of the variables in contingency tables. It is a terminology used in the textbook, "The Advanced Theory of Statistics" by Kendall and Stuart (1961).

**Following recent practice, we write X for the test statistic and reserve the symbol χ^2 for the distributional form.

2.2 Special Approximate Methods for 2 x 2 Tables (8)(15)

For the 2 x 2 table in the form [1.2], Yule (23) defines a coefficient of association, V, by the equation

$$V = \frac{(ad-bc)}{\left((a+b)(c+d)(a+c)(b+d) \right)^{1/2}} .$$

It can be shown that

$$X^2 \equiv nV^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \qquad [2.3]$$

is approximately distributed in the χ^2 form with one degree of freedom.

This is a short-cut method of computing chi-square for testing the hypothesis of independence and will give the same numerical value of chi-square that would be obtained by using formula [2.2].

2.3 Continuity Correction in the Large-Sample Test (8)(15)

As always, when using a continuous distribution to approximate a discrete one, a continuity correction improves the large-sample test based on [2.2]. In this case, the continuity correction, first suggested by Yates (22), requires that [2.3] should have the term (ad-bc) in its numerator replaced by $|ad-bc| - 0.5\,n$, which is the same as increasing (if $ad>bc$) b and c by 0.5, and reducing a and d by 0.5. Thus the corrected test statistic is

$$X^2 = \frac{n(|ad-bc| - 0.5\,n)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$= \sum_{i=1}^{2} \sum_{j=1}^{2} (|n_{ij} - m_{ij}| - 0.5)^2/m_{ij} \,.$$

This correction should not be applied to $r \times c$ tables in which $r > 2$ and $c > 2$.

2.4 The Exact Method for 2×2 Tables (8)(15)

For the table

n_{11}	n_{12}	$n_{1.}$
n_{21}	n_{22}	$n_{2.}$
$n_{.1}$	$n_{.2}$	n

[2.5]

consider two fractions, $\hat{p}_1 = n_{.1}/n_{1.}$ and $\hat{p}_2 = n_{21}/n_{2.}$ which are estimates of p_1 and p_2 , the parameters of two binomial populations.

On the hypothesis, which we write

$$H_o : p_1 = p_2$$

the probability of observing the table [2.5] when "all marginal frequencies are fixed" is

$$P_{1\cdot} = P[n_{ij} \mid n, n_{1.}, n_{.1}] = P[n_{ij} \mid n, n_{1.}]/P[n_{.1}/n]$$

$$= \binom{n_{1.}}{n_{11}}\binom{n_{2.}}{n_{21}} \Big/ \binom{n}{n_{.1}} = \frac{n_{1.}!\, n_{2.}!\, n_{.1}!\, n_{.2}!}{n!\, n_{11}!\, n_{12}!\, n_{21}!\, n_{22}!} \qquad \ldots\ldots\ldots \quad [2.6]$$

To obtain the final probability to be used in assessing the validity of $H_o : p_1 = p_2$, it is necessary to add to P_1 the probabilities of more divergent fractions than those observed. Assuming $\hat{p}_1 < \hat{p}_2$ (and the table can always be arranged to make this so), the next more divergent situation would be the one in which n_{11} and n_{22} are each decreased by unity, and n_{12} and n_{21} are each increased by unity. For this array, we calculate

$$P_2 = \frac{n_{1.}!\, n_{.1}!\, n_{2.}!\, n_{.2}!}{n!\ (n_{11}-1)(n_{12}+1)(n_{21}+1)(n_{22}-1)}$$

The cell entries are again changed, following the same rule as before, and P_3 is calculated. Continue in this manner until $P_{n_{11}}$ is calculated. Then, if

$$P = \sum_{i=1}^{n_{11}+1} P_i$$

is less than or equal to α, the hypothesis $H_o : p_1 = p_2$ should be rejected.

The exact test based on the probabilities [2.6] actually gives UMPU (uniformly most powerful unbiassed) tests for the other two cases, i.e., one set of marginal frequencies fixed. This result was first given by Tocher (20).

2.5 Special Method for 2 x C Tables: The Binomial Homogeneity Test (8)(19)

A particular case of the r x c table which is of special interest is the 2 x c table, where we are comparing c samples in respect of the possession of non-possession of an attribute. The general formula [2.2] for chi-squre reduces here to

$$X^2 = \sum_{i=1}^{2} \sum_{j=1}^{c} \frac{(n_{ij} - n_{i\cdot}\, n_{\cdot j}/n)^2}{n_{i\cdot}\, n_{\cdot j}/n} \qquad [2.7]$$

If we write

$$\hat{p} = n_{1\cdot}/n$$

for the maximum likelihood estimate from the table of the probability of observing a "success" (i.e., an entry in the first row of the table), [2.7] may be expressed as

$$X^2 = \sum_{j=1}^{c} \left[\frac{(n_{1j} - n_{\cdot j}\,\hat{p})^2}{n_{\cdot j}\,\hat{p}} + \frac{\{(n_{\cdot j} - n_{1j}) - n_{\cdot j}(1-\hat{p})\}^2}{n_{\cdot j}(1-\hat{p})} \right]$$

$$X^2 = \sum_{j=1}^{c} \frac{(n_{1j} - n_{\cdot j}\hat{p})^2}{n_{\cdot j}\hat{p}(1-\hat{p})} , \qquad [2.8]$$

distributed asymptotically as χ^2 with c-1 degrees of creedom. The test of the homogeneity of the c binomial samples based on [2.8] is thus seen essentially to be based on the sum of squares of c independent binomial variables each measured from its expectation (estimated on the hypothesis of homogeneity) and divided by its estimated standard error $\{n_{\cdot j}\hat{p}(1-\hat{p})\}^{1/2}$. There are c-1 degrees of freedom because we estimate the expectation linearly from the data--if it were given independently of the observations of p, we would replace $\hat{p}$ by p in [2.8] and have the full c degrees of freedom for .

2.6 Partition of χ^2 in the Case of the 2 x 3 Tables (10)(12)

Lancaster (12) proved that a 2 x 3 table can be reduced exactly to two 2 x 2 tables as follows:

Let q_{ij} be the probability of an observation falling into the class in the ith row and the jth column where i=1, 2; j=1, 2, 3.

Define $$q_{i\cdot} = \sum_{j=1}^{3} q_{ij} , \quad q_{\cdot j} = \sum_{i=1}^{2} q_{ij} .$$

The null hypothesis is that there is no association between the probability that an observation should fall in any row and in any column, i. e.

$$q_{ij} = q_{i.}\, q_{.j} \quad .$$

If there is no association we may write

$$P(n_{ij}/n,\ q_{ij}) = n! \prod_{i,j} \left(\frac{q_{ij}^{n_{ij}}}{n_{ij}!} \right) \qquad (i=1.2;\ j=1,2,3)$$

$$= \frac{n!}{n_{1.}!\, n_{2.}!}\, q_{1.}^{n_{1.}}\, q_{2.}^{n_{2.}} \times \frac{n!}{n_{.1}!\, n_{.2}!\, n_{.3}!}\, q_{.1}^{n_{.1}}\, q_{.2}^{n_{.2}}\, q_{.3}^{n_{.3}}$$

$$\times \frac{n_{1.}!\, n_{2.}!\, n_{.1}!\, n_{.2}!\, n_{.3}!}{n! \prod_{i,j} n_{ij}!}$$

Thus the row and column totals are sufficient statistics for $q_{i.}$ and $q_{.j}$ respectively. Further, we have

$$P(n_{ij}/n,\ q_{ij}) = P(n_{i.}/n,\ q_{i.})\ P(n_{.j}/n,\ q_{.j})\ P(n_{ij}/n_{i.}\ ,\ n_{.j}) \quad .$$

Hence

$$P(n_{ij}/n_{i.}\ ,\ n_{.j}) = \frac{n_{1.}!\, n_{2.}!\, n_{.1}!\, n_{.2}!\, n_{.3}!}{n! \prod_{i,j} n_{ij}!}$$

and summing over all n_{ij} , $\sum P(n_{ij}/n_{i.}\ ,\ n_{.j}) = 1$.

By a rearrangement,

$$P(n_{ij}/n_{i.}\ ,\ n_{.j})\ \frac{R_{12}!\ R_{22}!\ n_{.1}!\ n_{.2}!}{n_{11}!\ n_{12}!\ n_{21}!\ n_{22}!\ T_{22}!}\ \frac{T_{22}!\ n_{1.}!\ n_{2.}!\ n_{3.}!}{R_{12}!\ R_{22}!\ n_{13}!\ n_{23}!\ n!}\ ,$$

the two terms on the right-hand side beings the probabilities corresponding to the 2 x 2 tables

n_{11}	n_{12}	R_{12}		R_{12}	n_{13}	$n_{1.}$
n_{21}	n_{22}	R_{22}		R_{22}	n_{23}	$n_{2.}$
$n_{.1}$	$n_{.2}$	T_{22}		T_{22}	$n_{.3}$	n

We see above that we have two independent values χ_1 , χ_2 and $\chi_T^2 = \chi_1^2 + \chi_2^2$ with 2 degrees of freedom.

Kimball (10) gave a short-cut formula for the exact partition of χ^2 in the 2 x 3 tables as follows:

Let the observed frequencies and marginal totals of a 2 x 3 table be

n_{11}	n_{12}	n_{13}	$n_{1.}$
n_{21}	n_{22}	n_{23}	$n_{2.}$
$n_{.1}$	$n_{.2}$	$n_{.3}$	n

Then

$$\chi_1^2 = \frac{n^2 (n_{11} n_{12} - n_{12} n_{21})^2}{n_{1.} n_{2.} n_{.1} n_{.2} (n_{.1} + n_{.2})}$$

$$\chi_2^2 = \frac{n [n_{23} (n_{11} + n_{12}) - n_{13} (n_{21} + n_{22})]^2}{n_{1.} n_{2.} n_{.3} (n_{.1} + n_{.2})}$$

and

$$\chi_T^2 = \chi_1^2 + \chi_2^2$$

The partition of χ^2 for the general r x c tables or three-way tables, see references (10) and (12) .

3. THREE-WAY TABLES

3.1 Tests of No Second-Order Interaction

3.1.1 2 x 2 x 2 Tables

(a) Bartlett's original test. A 2 x 2 x 2 table may be in this form:

A_1				A_2				
B_1		B_2		B_1		B_2		Totals
C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	
p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	1
m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	n
n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n

The three classifications are designated by A, B, and C, respectively, while p_i, m_i, and n_i, are the probabilities, expected values and observed values, respectively, corresponding to the respective C_i subclasses.

The first-order interaction BC for A_1 is defined as

$$\frac{p_1/p_2}{p_3/p_4}$$

and for A_2 as

$$\frac{p_5/p_6}{p_7/p_8}$$

The null hypothesis tested for a 2 x 2 contingency table is

$$\frac{p_1/p_2}{p_3/p_4} = 1 \ ,$$

that is, the interaction is unity.

The null hypothesis for testing the existence of an ABC interaction is that the BC interaction is the same for both A_1 and A_2, or symbolically that

$$\frac{p_1/p_2}{p_3/p_4} = \frac{p_5/p_6}{p_7/p_8}$$

which reduces to

$$p_1 \, p_4 \, p_6 \, p_7 = p_2 \, p_3 \, p_5 \, p_8 \ , \qquad \ldots\ldots\ldots \qquad [3.1]$$

which is equivalent to

$$\frac{p_1 \, p_4 \, p_6 \, p_7}{p_2 \, p_3 \, p_5 \, p_8} = 1$$

By the method of maximum likelihood, we obtain the following estimators of m_i's: (see reference (1)).

$$\begin{aligned} \hat{m}_1 &= n_1 + x, & \hat{m}_2 &= n_2 - x, \\ \hat{m}_4 &= n_4 + x, & \hat{m}_3 &= n_3 - x, \\ \hat{m}_6 &= n_6 + x, & \hat{m}_5 &= n_5 - x, \\ \hat{m}_7 &= n_7 + x, & \hat{m}_8 &= n_8 - x, \end{aligned} \qquad [3.2]$$

for some number x .

Multiploying both sides of the equality [2.1] by n , we obtain

$$m_1 m_4 m_6 m_7 = m_2 m_3 m_5 m_8 . \qquad [3.3]$$

We obtain a value for x from the following equation by substitution.

$$(n_1 + x)(n_4 + x)(n_6 + x)(n_7 + x) = (n_2 - x)(n_3 - x)(n_5 - x)(n_8 - x) .$$

From x and n_i s, values for the $\hat{m}_i$ may be obtained. A test of the null hypothesis may now be performed by using the criterion

$$X^2 = \sum_{i=1}^{8} \frac{(n_i - m_i)^2}{m_i} \qquad [3.4]$$

where the associated degree of freedom is 1 .

Lewis made an adjustment for continuity and [3.4] becomes

$$X^2 = (|x| - .5)^2 \left(\frac{1}{n_1 + x} + \frac{1}{n_2 - x} + . \; . \; . + \frac{1}{n_8 - x}\right)$$

if $|x| > .5$.

(b) Lancaster's test (11)(13). In contrast with Bartlett's formulation, Lancaster's definition of second-order interaction is lengthy and difficult to verbalize. For the $2 \times 2 \times 2$ case in which $p_{i..}$, etc. are estimated from the data, and only the total n is fixed in advance, while for Bartlett's test the marginal totals $n_{.jk}$ are fixed in advance.

It can be shown that Lancaster's second-order interaction is zero if

$$n_1' (qq'q'') + n_4 (pp'q'') + n_6 (pq'p'') + n_7 (qp'p'') = n_2 (pq'q'')$$
$$+ n_3 (qp'q'') + n_5 (qq'p'') + n_8 (pp'p'')$$

where

$$p, p' \text{ and } p'' = p_{1..} , p_{.1} , \text{ and } p_{..1}$$

and

$$q, q' \text{ and } q'' = p_{2..} , p_{.2.} , \text{ and } p_{..2} .$$

This expression is substantially different from Bartlett's expression (see reference (18)).

(c) Goodman's test (7). Let p_{ijk} be the probability that an observation will fall in the ith row, jth column, kth layer of a three-way table, and let n_{ijk} denote the corresponding frequency in a sample of total size n, and θ_{ijk} denote the conditional probability that an observation will fall in the ith row and the jth column, given that it is in the kth layer.

For the 2 x 2 x 2 table the hypothesis of zero three-factor interaction is given by

$$H_o : \frac{p_{111}\, p_{221}}{p_{121}\, p_{211}} = \frac{p_{112}\, p_{222}}{p_{122}\, p_{212}}$$

which can be rewritten as

$$H_o : \frac{\theta_{111}\, \theta_{221}}{\theta_{121}\, \theta_{211}} = \frac{\theta_{112}\, \theta_{222}}{\theta_{122}\, \theta_{212}}$$

Writing $(\theta_{11k}\, \theta_{22k})/(\theta_{12k}\, \theta_{21k}) = \Delta_k$, we note that Δ_k is a measure of the two-factor interaction (i. e., the row-column interaction) in the 2 x 2 table corresponding to the kth layer, and the null hypothesis H_o specifies that this two-factor interaction is the same for each layer; i. e.,

$$H_o : \Delta_1 = \Delta_2 .$$

Writing $n_{..k} = \sum_{i,j} n_{ijk}$, the conditional distribution of the kth set of random variables $\{n_{11k}, n_{12k}, n_{21k}, n_{22k}\}$ given $n_{..k}$

is the usual multinomial distribution associated with the double dichotomy having parameters $\{\theta_{11k}, \theta_{12k}, \theta_{21k}, \theta_{22k}\}$. The maximum likelihood estimator of Δ_k is $d_k = (n_{11k}\, n_{22k})/(n_{12k}\, n_{21k})$, and its variance can be estimated consistently by

$$v_k = d_k^2\, u_k ,$$

where $u_k = \sum_{i,j} n_{ijk}^{-1}$ (see reference (8)). Having computed d_k and v_k, it is possible to test various hypotheses concerning the Δ_k.

To test the null hypothesis H_0 that $\Delta_1 = \Delta_2$, we compute the statistic

$$X^2 = (d_1 - d_2)^2/(v_1 + v_2) , \qquad [3.5]$$

which will be distributed asymptotically (when $n \to \infty$) as chi-square with one degree of freedom when H_0 is true. [3.5] was suggested by Goodman in reference (5).

Denoting the natural logarithm of x by $\log x$. The maximum likelihood estimator of $\log \Delta_k$ is $\log d_k$ and its variance can be estimated consistently by u_k. Denoting $\log \Delta_k$ by T_k and $\log d_k = \log n_{11k} + \log n_{22k} - \log n_{12k} - \log n_{21k}$ by g_k, we can test the null hypothesis H_0 that $T_1 = T_2$ by calculating the test statistic

$$Y^2 = (g_1 - g_2)^2/(u_1 + u_2) ,$$

which will be distributed asymptotically as chi-square with one degree of freedom when H_0 is true.

The null hypothesis H_0 that $T_1 = T_2$ is, of course, equivalent to the null hypothesis that $\Delta_1 = \Delta_2$, and the statistic Y^2 will be asymptotically equivalent when H_0 is true to the statistic X^2 (see reference (6)).

The test based upon Y^2 analyzes the log-frequencies, whereas the test based upon X^2 analyzes the frequencies themselves.

3.1.2 2 x 2 x T Tables

(a) Norton's test (14). By Bartlett's method, the 2 x 2 x 2 table in the following example has but a single degree of freedom and a single quantity, x, by which each observed value departs, positively or negatively, from the corresponding expected value.

	Males		Females	
	Alive	Dead	Alive	Dead
Controls	a	b	e	f
Experimentals	c	d	g	h

The departure, x, of the eight observed numbers from those which are proportional and have the same marginal values, may be found by solving

$$(a-x)(d-x)(f-x)(g-x) = (b+x)(c+x)(e+x)(h+x) . \qquad [3.6]$$

After a sufficiently accurate value of x has been found, it may be applied to the observed values to find the expected values, and chi-square may then be calculated as usual. This procedure has been discussed in Section 3.1.1 (a).

Bartlett mentions the table of type $2 \times 2 \times 3$, observing that it may be treated by comparing two levels, which may then be combined (if homogeneous) and tested against the third level, reducing the problem to that of two $2 \times 2 \times 2$ tables. Since this is not completely general, he also gave a pair of simultaneous equations which yield the departures for such a table, which has two degrees of freedom. Equation [3.6] may be written

$$\frac{(a-x)(d-x)}{(b+x)(c+x)} = \frac{(e+x)(h+x)}{(f-x)(g-x)} . \qquad [3.7]$$

The left member of [3.7] contains all the elements of a 2×2 table, and if that 2×2 table alone were being tested for homogeneity (i.e., proportionality), the value of x would satisfy the equation

$$\frac{(a-x)(d-x)}{(b+x)(c+x)} = 1.$$

A $2 \times 2 \times 2$ table may be regarded as the association of two 2×2 tables, and the test of homogeneity of the $2 \times 2 \times 2$ table is a test not of homogeneity of each of the two 2×2 tables, but of the agreement between those two tables as to the nature and degree of their respective departures from homogeneity. Therefore x is chosen to satisfy [3.7], whatever may be the value of the two members of [3.7] when x is so chosen.

Turning to the case of the table of the form $2 \times 2 \times 3$, which has four additional observed quantities, s, t, u and v, and another degree of freedom and hence another departure y, the generalization of [3.7] is

$$\frac{(a-x)(d-x)}{(b+x)(c+x)} = \frac{(e-y)(h-y)}{(f+y)(g+y)} = \frac{(s+x+y)(v+x+y)}{(t-x-y)(u-x-y)} ,$$

which is equivalent to the equation given by Bartlett. Letting $-z = x + y$, this is

$$\frac{(a-x)(d-x)}{(b+x)(c+x)} = \frac{(e-y)(h-y)}{(f+y)(g+y)} = \frac{(s-z)(v-z)}{(t+z)(u+z)} \qquad [3.8]$$

and

$$x + y + z = 0 .$$

A new notation will now be introduced to facilitate the representation of the generalization of [3.8] to the case of a table of the form $2^N \times t$. Let each observed number in the table be represented by a symbol of the form $M_1 M_2 M_3 \ldots M_{N+1}$, where M_u is either one or two for $u = 1, 2, 3, \ldots, N$, and M_{N+1} is $1, 2, 3, \ldots$ or t. For example, in the $2 \times 2 \times 2$ table given above, "a" might be designated by 111. Let x_i be the departure appropriate to the ith level of the table, i running from 1 to t. In any one member of the generalization of [3.8], corresponding to a particular level i among the t, those observed numbers may be placed (conventionally) in the numerator for which the representation $M_1 M_2 \ldots M_{N+1}$ contains an even number of ones in the first N positions, and those containing an odd number are then placed in the denominator. Representing the ith member of the generalization of [3.8] by w_i, for a $2 \times 2 \times t$ table,

$$w_i = \frac{(11i - x_i)(22i - x_i)}{(12i - x_i)(21i - x_i)} . \qquad [3.9]$$

For higher table ($2^3 \times t$, $2^4 \times t$, and so on) there are simply more terms in both numerator and denominator of w_i. It is necessary to find t quantities x_i such that $w_i = w_j$ for all i and j and such that

$$\sum_i x_i = 0 . \qquad [3.10]$$

Then the x_i may be applied to the observed numbers of their respective levels to find the expected numbers, and chi-square may be calculated as usual.

It is the arithmetic solution of the t simultaneous equations [3.9] together with [3.10] which is the problem, and which may be accomplished rather easily, using the following procedure of successive approximation. Equation [3.9] may be written

$$w_i = \frac{(11i)(22i)}{(12i)(21i)} \; \frac{(1-x_i/11i)(1-x_i/22i)}{(1+x_i/12i)(1+x_i/21i)} \, . \qquad [3.11]$$

Assuming the validity of expanding the terms in the denominator which involve x_i, the approximation to terms of order x_i is

$$w_i = \frac{(11i)(22i)}{(12i)(21i)} \left(1 - \frac{x_i}{11i} - \frac{x_i}{12i} - \frac{x_i}{21i} - \frac{x_i}{22i}\right) . \qquad [3.12]$$

Letting

$$\frac{1}{p_i} = \frac{(11i)(22i)}{(12i)(21i)} ,$$

and

$$\frac{1}{s_i} = \frac{1}{11i} + \frac{1}{12i} + \frac{1}{21i} + \frac{1}{22i} ,$$

equation [3.12] becomes

$$w_i = \frac{1}{p_i} \left(1 - \frac{x_i}{s_i}\right) .$$

Equating the approximate values of w_i and w_j given by [3.13] and solving for x_j ,

$$x_j = s_j \left[1 - \frac{p_j}{p_i} \left(1 - \frac{x_i}{s_i} \right) \right] .$$

Since we require

$$\sum_i (x_i) = 0 ,$$

$$p_j \sum_j (s_j) - \sum_j (s_j \, p_j) + \frac{x_i}{s_i} \sum_j (s_j \, p_j) = 0 .$$

Solving for x_i ,

$$x_i = s_i \, (1 - hp_i) , \qquad [3.14]$$

where

$$h = \frac{\sum_i (s_i)}{\sum_i (s_i \, p_i)}$$

Equation [3.14] provides approximate values of the x_i , and the calculations require only the preparation of the s_i , p_i and h . To get additional corrections, the x_i must be added to the corresponding observed values, and these adjusted values (which are first approximations to the expected values, on the hypothesis of homogeneity) are used to repeat the calculations of [3.14]. This process is continued until the x_i are determined with satisfactory accuracy.

(b) Goodman's test (7). The analysis of the 2 x 2 x 2 table in Section 3.1.1 (c) will now be generalized to cover the analysis of the 2 x 2 x t table. The hypothesis of zero three-factor inter-action in the 2 x 2 x t table is given by

$$H_0 : \frac{p_{111}\, p_{221}}{p_{121}\, p_{211}} = \frac{p_{11k}\, p_{22k}}{p_{12k}\, p_{21k}} \quad \text{for } k = 2, 3, \ldots, t,$$

which can be rewritten as

$$H_0 : \Delta_1 = \Delta_k, \quad \text{for } k = 2, 3, \ldots, t,$$

where

$$\Delta_k = (p_{112}\, p_{22k})/(p_{12k}\, p_{21k}) = (\theta_{11k}\, \theta_{22k})/(\theta_{12k}\, \theta_{21k})$$

is a measure of two-factor interaction in the 2 x 2 population tableau corresponding to the kth layer. The hypothesis H_0 states that the two-factor interaction in the 2 x 2 population tableau corresponding to the kth layer is the same for $k = 1, 2, \ldots, t$. Estimating Δ_k by $d_k = (n_{11k}\, n_{22k})/(n_{12k}\, n_{21k})$ in the kth layer, a test of H_0 can be based upon the statistic

$$X^2 = \sum_{k=1}^{t} (d_k - d)^2/v_k = \sum_{k=1}^{t} d_k^2 w_k - \left[\sum_{k=1}^{t} d_k w_k\right]^2 / \sum_{k=1}^{t} w_k ,$$

where

$$v_k = d_k^2 u_k , \quad u_k = \sum_{i,j} n_{ijk}^{-1} , \quad w_k = v_k^{-1} ,$$

$$d = \sum_{k=1}^{t} d_k w_k \Big/ \sum_{k=1}^{t} w_k .$$

When H_0 is true, the statistic X^2 will be distributed asymptotically as chi-square with $t-1$ degrees of freedom. Note that X^2 can be rewritten more simply as

$$X^2 = \sum_{k=1}^{t} m_k - \left[\sum_{k=1}^{t} b_k m_k\right]^2 \Big/ \sum_{k=1}^{t} b_k^2 m_k ,$$

where $m_k = u_k^{-1}$ and $b_k = d_k^{-1} = (n_{12k} n_{21k})/(n_{11k} n_{22k})$. This statistic can also be written as

$$X^2 = \sum_{k=1}^{t} (c_k - 1)^2 / u_k ,$$

where $c_k = d/d_k$, which indicates that X^2 tests whether the c_k ($k = 1, 2, \ldots, t$) are significantly different from one.

(c) Woolf's test (21). Let the frequencies in the kth 2 x 2 table be denoted by n_{1k}, n_{2k}, n_{3k}, n_{4k}, where n_{1k}, n_{2k} occupy the first row and n_{1k}, n_{3k} the first column. Compute $z_k = \log n^*_{1k} - \log n_{2k} - \log n_{3k} + \log n_{4k}$ and u_k from

$$1/u_k = 1/n_{1k} + 1/n_{2k} + 1/n_{3k} + 1/n_{4k}.$$

If there is zero second-order interaction, then

$$X^2 = \sum u_k z_k^2 - (\sum u_k z_k)^2 / \sum u_k$$

is asymptotically distributed as chi-square with t-1 degrees of freedom. This analysis is computed more easily than Norton's.

3.1.3 R x C x T Tables

(a) Kastenbaum's and Lamphiear's test (9). Roy and Kastenbaum (17) defined no second-order interaction hypothesis for an r x c x t table by the following set of (r-1)(c-1)(t-1) conditions on the probabilities:

$$H : \frac{p_{rct}\, p_{ijt}}{p_{ict}\, p_{rjt}} = \frac{p_{rck}\, p_{ijk}}{p_{ick}\, p_{rjk}}, \qquad [3.15]$$

where i = 1, 2, ..., r-1; j = 1, 2, ..., c-1; k = 1, 2, ..., t-1.

*Natural logarithm.

Under the null hypothesis, estimates of the parameters may be achieved by first solving the following simultaneous systems of third-degree equations for all x_{ijk} :

$$\frac{\left(n_{rct} - \sum_{i=1}^{r-1}\sum_{j=1}^{c-1} x_{ijt}\right)\left(n_{ijt} - x_{ijt}\right)}{\left(n_{ict} + \sum_{j=1}^{c-1} x_{ijt}\right)\left(n_{rjt} + \sum_{i=1}^{r-1} x\right)} = \frac{\left(n_{rck} - \sum_{i=1}^{r-1}\sum_{j=1}^{c-1} x_{ijk}\right)\left(n_{ijk} - x_{ijk}\right)}{\left(n_{ick} + \sum_{j=1}^{c-1} x_{ijk}\right)\left(n_{rjk} + \sum_{i=1}^{r-1} x_{ijk}\right)} \qquad [3.16]$$

where $i = 1, 2, \ldots, r-1$; $j = 1, 2, \ldots, c-1$; $k = 1, 2, \ldots, t$, and where

$$x_{ijt} = -\sum_{k=1}^{t-1} x_{ijk}, \quad \text{or} \quad \sum_{k=1}^{t} x_{ijk} = 0 .$$

To solve equations [3.16], first let all $x_{ijk} = 0$, except for one set: x_{11k} (say) .

Then the general term of [3.16] reduces to

$$a_{11k} = \frac{(n_{rck} - x_{11k})(n_{11k} - x_{11k})}{(n_{1ck} + x_{11k})(n_{r1k} + x_{11k})}$$

for $k = 1, 2, \ldots, t$. It follows, by direct application of Norton's procedure that

$$x_{11k}^{(1)} \cong c_{11k}^{(1)} \left[1 - h_{11}^{(1)} b_{11k}^{(1)} \right] , \qquad [3.17]$$

where

(i) the superscript (w) refers to the wth correction,

(ii) $$\sum_{w=1}^{m} x_{ijk}^{(w)} = x_{ijk} ,$$

(iii) m is the number of iterations necessary to make $x_{ijk}^{(m)} = 0$ with desired accuracy ,

(iv) $$b_{11k} = \frac{n_{1ck}\, n_{r1k}}{n_{rck}\, n_{11k}} ,$$

(v) $$\frac{1}{c_{11k}} = \frac{1}{n_{rck}} + \frac{1}{n_{11k}} + \frac{1}{n_{1ck}} + \frac{1}{n_{r1k}} ,$$

(vi) $$h_{11} = \sum_{k=1}^{t} c_{11k} \Big/ \sum_{k=1}^{t} b_{11k}\, c_{11k}$$

Equation [3.17] provides first approximations for the x_{11k} . These values of $x_{11k}^{(1)}$ are now either added to or subtracted from the observed cell frequencies with which they are associated according as Equation [3.16] specifies. After this set of corrections has been applied, the iteration continues with the next set of equations involving x_{12k} (say). Again, set all $x_{ijk} = 0$ except for the x_{12k}, solve for $x_{12k}^{(1)}$, apply these corrections to the appropriate cell frequencies, and continue this procedure for all $x_{ijk}^{(1)}$ [$i = 1, 2, \ldots, r-1$; $j = 1, 2, \ldots, c-1$]. When all the first corrections have been determined and applied, the iteration begins again with $x_{11k}^{(2)}$, and continues until all the $x_{ijk}^{(m)} = 0$ with desired accuracy. Note that the x_{ijk} must be numerically smaller than their respective n_{ijk}. The calculation of the test statistic is as follows:

Let

$$u_{rct} = \sum_{i=1}^{r-1} \sum_{j=1}^{c-1} \sum_{k=1}^{t-1} x_{ijk}, \quad u_{rck} = \sum_{i=1}^{r-1} \sum_{j=1}^{c-1} x,$$

$$u_{ict} = \sum_{j=1}^{c-1} \sum_{k=1}^{t-1} x_{ijk}, \quad u_{ick} = \sum_{j=1}^{c-1} x_{ijk},$$

$$u_{rjt} = \sum_{i=1}^{r-1} \sum_{k=1}^{t-1} x_{ijk}, \quad u_{rjk} = \sum_{i=1}^{r-1} x_{ijk},$$

$$u_{ijt} = \sum_{k=1}^{t-1} x_{ijk}, \quad u_{ijk} = x_{ijk}.$$

Then the statistic

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{t} u_{ijk}^2 / (n_{ijk} + s_{ijk} u_{ijk}) \qquad [3.18]$$

is distributed approximately as chi-square with $(r-1)(c-1)(t-1)$ degrees of freedom, where $s_{ijk} = +1$, if $ijk = rct$ or if any two subscripts differ from rct; and $s_{ijk} = -1$, if only one or all three subscripts differ from rct. Equation [3.18] provides a test statistic for testing H_0 in [3.15].

(b) Plackett's test (16). Plackett made an extension of Woolf's method for the case of $2 \times 2 \times t$ to the case of $r \times c \times t$.

Suppose that the frequencies in the table, say n_v, n_w, ..., arise from sampling a single multinomial distribution, with probabilities p_v, p_w, ..., and total frequency n. Then

$$\text{var}(\log n_v) \sim \frac{1}{np_v} - \frac{1}{n}, \quad \text{cov}(\log n_v, \log n_w) \sim -\frac{1}{n}.$$

The analysis is conducted in terms of contrasts of the $\{\log n_v\}$, which estimate the corresponding contrasts of $\{\log p_v\}$. We can regard these variables as uncorrelated and the variance of $\log n_v$ as $1/(np_v)$. Let R be a matrix of order $(r-1) \times r$, the rows of which are orthogonal to each other and to the unit vector. Similarly define C. From the kth layer of the table, we form $(r-1)(c-1)$ linear combinations

$\{z_{ki}\}$ of the logarithms of the frequencies, the coefficients in z_{ki} being taken from the ith row in the direct product of R and C. The asymptotic distribution of the vector $z_k = \{z_{ki}\}$ is multivariate normal with a dispersion matrix V_k which is estimated by replacing $1/(np_v)$ by $1/n_v$. On the hypothesis that the second-order interaction is zero, the vectors $z_1, z_2, \ldots, z_t$ have the same centroid, in which case

$$X^2 = \sum z_k' V_k^{-1} z_k - \left(\sum z_k' V_k^{-1}\right)\left(\sum V_k^{-1}\right)^{-1}\left(\sum V_k^{-1} z_k\right)$$

is asymptotically distributed as chi-squared with $(r-1)(c-1)(t-1)$ degrees of freedom. The evaluation of chi-squared by this method thus involves the inversion of t matrices of side $(r-1)(c-1)$, so that the classifications are preferably labelled with $r \leq c \leq t$.

(c) A modification of Plackett's test by Goodman (6).

Writing $e_{ijk} = \log n_{ijk}$, Goodman forms $(r-1)(c-1)$ linear combinations $\{y_{ijk}\}$ of the e_{ijk} by taking $y_{ijk} = e_{ijk} - e_{ick} - e_{rjk} + e_{rck}$ for $i = 1, 2, \ldots, r-1$, and $j = 1, 2, \ldots, c-1$. In this case, it is easy to see that the rows of the transformation matrices corresponding to R and to C, yielding the y_{ijk} rather than the z_{dk}, will not be orthogonal to each other. Denoting the column vector $\{y_{i1k}, y_{i2k}, \ldots, y_{i,c-1,k}\}$ by y_{ik}, and the column vector $\{y_{1k}, y_{2k}, \ldots, y_{r-1,k}\}$ by y_k, we note that y_k has $(r-1)(c-1)$ elements

$$y_{ijk} \quad (i=1, 2, \ldots, r-1; \; j=1, 2, \ldots, c-1) ;$$

viz. the r-1 column vectors

$$y_{ik} \quad (i=1, 2, \ldots, r-1)$$

each containing c-1 elements

$$y_{ijk} \quad (j=1, 2, \ldots, c-1) .$$

The asymptotic distribution of y_k is multivariate normal with a dispersion matrix $Q^{(k)}$ having $(r-1)(c-1)$ rows and $(r-1)(c-1)$ columns. The dispersion matrix $Q^{(k)}$ can be estimated consistently by $\hat{Q}^{(k)}$ obtained by noting that in an analysis of the y_{ijk}, which are contrasts of the e_{ijk}, the e_{ijk} can be regarded as uncorrelated with variance estimated consistently by n_{ijk}^{-1}. Applying multivariate normal theory, we find that when H_o is true the statistic

$$X^2 = \sum_{k=1}^{t} (y_k - \hat{y})' \, M^{(k)} \, (y_k - \hat{y})$$

$$= \sum_{k=1}^{t} y_k' \, M^{(k)} \, y_k - v' D v$$

is distributed asymptotically as chi-squared with $(t-1)(r-1)(c-1)$ degrees of freedom, provided that we define $M^{(k)}$ as the inverse of $\hat{Q}^{(k)}$, D as the inverse of

$$\sum_{k=1}^{t} M^{(k)} = M ,$$

$$\hat{y} = Dv , \quad \text{and}$$

$$v = \sum_{k=1}^{t} M^{(k)} y_k .$$

The statistic χ^2 can be used to test the null hypothesis H_o that the second-order interaction is zero.

To apply the test described above, it will actually not be necessary to invert the matrix $\hat{Q}^{(k)}$ $(k = 1, 2, \ldots, t)$, a matrix of side $(r-1)(c-1)$, in order to calculate $M^{(k)}$, while to apply Plackett's method it will be required to invert the t matrices $\hat{V}^{(k)}$ $(k = 1, 2, \ldots, t)$, each of side $(r-1)(c-1)$. The calculation of $M^{(k)}$ can be simplified because of the special form of the estimated dispersion matrix $\hat{Q}^{(k)}$. For an explicit expression for this estimated dispersion matrix, see reference (6) .

3.2 Tests of Mutual Independence* (13)

3.2.1 Random Sampling Throughout Table

Take an example of the following $2 \times 2 \times 2$ table:

---

*Section 3.2 is ready to be applied to the general $r \times c \times t$ tables.

	C_1		C_2		
	B_1	B_2	B_1	B_2	$n_{i\cdot\cdot}$
A_1	$n_{1\cdot\cdot}$	n_{121}	n_{112}	n_{122}	$n_{1\cdot\cdot}$
A_2	$n_{2\cdot\cdot}$	n_{221}	n_{212}	n_{222}	$n_{2\cdot\cdot}$
$n_{\cdot jk}$	$n_{\cdot 11}$	$n_{\cdot 21}$	$n_{\cdot 12}$	$n_{\cdot 22}$	
$n_{\cdot\cdot k}$	$n_{\cdot\cdot 1}$		$n_{\cdot\cdot 2}$		

[3.19]

where A_1, B_1, and C_1 denote successful recall, while A_2, B_2, and C_2 denote failure to recall. The A, B, and C classifications representing the rows, columns and layers, respectively.

If $p_{1\cdot\cdot}$, $p_{\cdot 1\cdot}$, and $p_{\cdot\cdot 1}$ are the probabilities of recalling A, B, and C respectively, it is of interest to enquire whether the corresponding events are independent, or whether the recall of one item appears to facilitate (or inhibit) the recall of others. In the (most usual) case where the probabilities are estimated from the data:

$$\hat{p}_{1\cdot\cdot} = n_{1\cdot\cdot}/n$$

$$\hat{p}_{\cdot 1\cdot} = n_{\cdot 1\cdot}/n$$

$$\hat{p}_{\cdot\cdot 1} = n_{\cdot\cdot 1}/n ,$$

so that the corresponding estimated probabilities for failure to recall are

$$\hat{p}_{2\cdot\cdot} = 1 - \hat{p}_{1\cdot\cdot}, \hat{p}_{\cdot 2\cdot} = 1 - \hat{p}_{\cdot 1\cdot}, \hat{p}_{\cdot\cdot 2} = 1 - \hat{p}_{\cdot\cdot 1} .$$

When the 3 recall probabilities are independent, the expected frequency for each cell may be computed directly by multiplication. Thus, the expected frequency for the (122) cell = $n\hat{p}_{122} = n(\hat{p}_{1\cdot\cdot})(\hat{p}_{\cdot 2\cdot})(\hat{p}_{\cdot\cdot 2})$.

The test for complete mutual independence is now achieved by setting up the null hypothesis:

$$H_o : p_{ijk} = p_{i\cdot\cdot}\, p_{\cdot j\cdot}\, p_{\cdot\cdot k}$$

and testing (against the general alternative $H_A \neq H_o$) by the formula:

$$X^2 = \sum_{ijk} \left(n_{ijk} - \frac{n_{i\cdot\cdot}\, n_{\cdot j\cdot}\, n_{\cdot\cdot k}}{n^2} \right)^2 \Big/ \left(\frac{n_{i\cdot\cdot}\, n_{\cdot j\cdot}\, n_{\cdot\cdot k}}{n^2} \right) .$$

The degrees of freedom would be (rct-r-c-t-2) = 4 .

3.2.2 <u>Random Sampling Within Each (R x C) Section</u>

In [3.19] the final form of the 8-cell distribution of frequencies was left wholly to change, since only the grand total n of the table was specified in advance by the sampling procedure. This is not always the case. Sometimes (e.g., when random sampling might produce disproportionately low frequencies in some sections of the r x c x t table) the experimenter might decide to specify not only the sample size n, but also the size of the sub-total $n_{..k}$ of each of the $(r \times c)$ layers.

Kullback, Mood, and Roy modified the null hypothesis as follows:

If the $n_{..k}$ totals are fixed in advance, it might be argued that there are in effect t distinct tables of size $r \times c$. In these circumstances p_{ijk} denotes the probability that an observation falls in the (ij)th cell of the kth two-way table. Moreover, if each two-way table is considered separately:

$$p_{ijk} = n_{ijk}/n_{..k}$$

and

$$p_{..k} = 1 .$$

Hence the hypothesis:

$H_o : p_{ijk} = p_{i..} \, p_{.j.} \, p_{..k}$ would be modified to

$H_o : p_{ijk} = p_{i..} \, p_{.j.}$ for $k = 1, 2$.

3.3 Tests of Partial Independence (13)

3.3.1 One Classification Independent of All Others

If the chi-quare test for mutual independence gives a significant result, it should not be assumed that all 3 classifications interact. It might be the case that just two of the classifications interact, and that the third is completely independent. This gives rise to 3 easily testable hypotheses, since any of the 3 classifications could be the independent one.

To test whether the row classification of [3.19] is independent of the others, it is only necessary to see whether the 4 ratios n_{111}/n_{211}, n_{121}/n_{221}, n_{112}/n_{212}, and n_{122}/n_{222} may be regarded as chance deviations from the over-all marginal ratio $n_{1..}/n_{2..}$.

The null hypothesis for this test is:

$$H_0 : p_{ijk} = p_{i..}\, p_{.jk}$$

so that

$$X^2 = \sum_{ijk} \left(n_{ijk} - \frac{n_{i..}\, n_{.jk}}{n} \right) \Big/ \frac{n_{i..}\, n_{.jk}}{n}$$

but with only 3 degrees of freedom on this occasion, since there are 3 pairs of frequencies that are free to vary, and just 1 degree of freedom within each pair. In the general r x c x t case, the test of independence of the row classification would have (ct-1)(r-1) degrees of freedom, and similar extensions can be made for higher-order tables. For instance, the test for complete independence of any one classification in the 2^k case would have $2^{k-1}-1$ degrees of freedom.

Since the $p_{.jk}$ values are estimated from the border totals $n_{.jk}$, the test proceeds as if the $n_{.jk}$ totals were fixed in advance. The chi-square computation is therefore not in any way changed if the $n_{.jk}$ totals are fixed, but the power function of the test will have a different structure, and the null hypothesis should be expressed in the form:

$$H_o : p_{ijk} = p_{i..} \; n_{.jk}/n$$

to indicate that only $p_{i..}$ are left undetermined by the sampling procedure.

3.3.2 <u>One Classification Independent of One Other</u> (13)

If the row classification is completely independent of the other two, it follows that the row x column interaction is zero, and that the row x layer interaction is also zero. This raises the general possibility that in any multi-way table, certain classifications might be independent of some, but not all, of the others.

To test for zero row x column interaction, the irrelevant "layer" classification is eliminated by summation in order to give a straightforward two-way table of size r x c , which has cell frequencies of $n_{ij\cdot}$ and marginal frequencies of $n_{i\cdot\cdot}$ (for rows) and $n_{\cdot j\cdot}$ (for columns) . The null hypothesis of zero row x column interaction is therefore:

$$H_o : p_{ij\cdot} = p_{i\cdot\cdot} \, p_{\cdot j\cdot}$$

which is tested in the normal way (for two-way tables) and has (r-1) (c-1) degrees of freedom.

Similarly, the null hypothesis of zero row x layer interaction is:

$$H_o : p_{i\cdot k} = p_{i\cdot\cdot} \, p_{\cdot\cdot k}$$

with (r-1)(t-1) degrees of freedom.

3.4 Tests of Conditional Independence (13)

In some three-way tables it is of interest to test the hypothesis that given, for example, any layer, the row and column classifications are independent.

This hypothesis of conditional independence is particularly relevant if the layer totals $n_{\cdot\cdot k}$ are fixed in advance, since it is likely that attention would then be mainly directed toward the r x c interaction within each layer. In fact, the fixing of the $n_{\cdot\cdot k}$ totals usually implies that there are t independent two-way tables of size $r \times c$, so the chi-squares of these t tables may be summed to provide an over-all test of conditional independence.

For the $r \times c \times t$ case in which there is random sampling throughout, the null hypothesis is:

$$H_o : p_{ijk} = p_{i\cdot k}\, p_{\cdot jk} / p_{\cdot\cdot k} .$$

If the $n_{\cdot\cdot k}$ totals are fixed, the null hypothesis is amended by substituting $n_{\cdot\cdot k}/n$ for $p_{\cdot\cdot k}$, and the power function of the test is different. However, the chi-square computation remains the same and is :

$$X^2 = \sum_{ijk} (n_{ijk} - n_{i\cdot k}\, n_{\cdot jk}/n_{\cdot\cdot k})^2 / (n_{i\cdot k}\, n_{\cdot jk}/n_{\cdot\cdot k})$$

with $t(r-1)(c-1)$ degrees of freedom.

Since the above hypothesis involves $p_{i\cdot k}$ and $p_{\cdot jk}$ it is relevant to point out that if

$$p_{i\cdot k} = p_{i\cdot\cdot}\, p_{\cdot\cdot k} \quad \text{(i. e., there is no } r \times t \text{ interaction)}$$

or if

$$p_{ijk} = p_{\cdot j\cdot}\, p_{\cdot\cdot k} \quad \text{(i. e., there is no } c \times t \text{ interaction)}$$

then the original hypothesis becomes:

$$p_{ijk} = p_{i..}\, p_{.jk} \qquad \text{(complete independence of rows)}$$

or

$$p_{ijk} = p_{.j.}\, p_{i.k} \qquad \text{(complete independence of columns),}$$

both of which imply:

$$p_{ij.} = p_{i..}\, p_{.j.} \qquad \text{(absence of } r \times c \text{ interaction) .}$$

It is therefore apparent that if the row and column classifications are independent for each layer, it does not follow that there is over-all independence between rows and columns, for the original hypothesis

$$p_{ijk} = p_{i.k}\, p_{.jk}/p_{..k} \quad \text{does not, by itself, imply}$$

$$p_{ij.} = p_{i..}\, p_{.j.} \quad .$$

Hence some extra condition, such as the absence of $r \times t$ or $c \times t$ interaction, must be superimposed before the row and column classifications become both conditionally and unconditionally independent. If both these extra conditions are simultaneously imposed, then the original hypothesis becomes (after substituting $p_{i.k} = p_{i..}\, p_{..k}$ and $p_{.jk} = p_{.j.}\, p_{..k}$) :

$$H_o : p_{ijk} = p_{i..}\, p_{.j.}\, p_{..k}$$

which is the condition of complete mutual independence.

It is now apparent that the hypothesis of complete independence can be exactly broken down into 3 component hypotheses which are mutually exclusive in logic, and which are concerned with (a) conditional independence of rows and columns, (b) absence of r x t interaction, and (c) absence of c x t interaction. If the chi-squares associated with these three component hypotheses are designated χ_1^2, χ_2^2, and χ_3^2 respectively, and if the chi-square for complete independence is designated χ^2, it is therefore reasonable to ask whether:

$$\chi_1^2 + \chi_2^2 + \chi_3^2 = \chi^2$$

A sketchy and incomplete answer to this inquiry has been provided by Roy (17) who states that, unlike what happens in standard analysis of variance procedures, there is no additivity in the usual algebraic sense, but additivity does occur in probability and asymptotically as $n \to \infty$.

3.5 Homogeneity of Two-Way Tables (13)

With suitable hypotheses and restrictions an r x c x t table may be thought of as being a set of r two-way tables of size c x t.

Conversely, r independent samples of a c x t table may be treated as an r x c x t three-way table. In such cases it is often interesting to test whether the c x t tables are homogeneous. (This is especially so if the $n_{i..}$ total of each c x t table is fixed in advance.)

The two-way tables will be homogeneous if the probabilities associated with corresponding cells are homogeneous, i.e., if

$$p_{1jk}/p_{1..} = p_{2jk}/p_{2..} = \dots = p_{rjk}/p_{r..} = p_{.jk} \ .$$

The null hypothesis (for random sampling throughout) is therefore:

$$H_0 : p_{ijk} = p_{i..} \ p_{ijk}$$

which is precisely the same as the hypothesis for the complete independence of the row classification.

Since

$p_{ijk} = p_{i..} \ p_{.jk}$ implies (by summing over k)

$p_{ij.} = p_{i..} \ p_{.j.}$ and also implies

$p_{ijk} = p_{ij.} \ p_{.jk}/p_{.j.}$

and since the latter two equalities imply the former, it follows that the (c x t) tables are homogeneous if and only if:

(a) the row and column classifications are independent

$(p_{ij\cdot} = p_{i\cdot\cdot}\ p_{\cdot j\cdot})$,

(b) the row and layer classifications are independent, given

the column classification $(p_{ijk} = p_{ij\cdot}\ p_{\cdot jk}/p_{\cdot j\cdot})$.

4. HIGHER-ORDER TABLES (13)(3)

There is little practical interest in the analysis of higher-order interactions in contingency tables with more than three factors and we shall therefore limit our discussion to a brief consideration of five-way tables.

For a five-way table, there might be no completely independent classification, but one pair of classifications (which interact with each other) might nevertheless be independent of the remaining three (which also interact mutually). The null hypothesis

$$H_0 : p_{ghijk} = p_{gh\cdots}\, p_{\cdot\cdot ijk}$$

would represent such a case, and would yield a set of expected frequencies to which the chi-square test could be applied as before.

Similarly, it is possible to envisage more elaborate forms of homogeneity of conditional independence, but once again these do not involve any new conceptual problems and do not, therefore, merit special consideration.

5. ANALOGY WITH THE ANALYSIS OF VARIANCE (3)

Bartlett defined the condition for no second-order interaction in a 2 x 2 x 2 table as

$$\frac{p_{111}\, p_{221}}{p_{121}\, p_{211}} = \frac{p_{112}\, p_{222}}{p_{122}\, p_{212}} \qquad [5.1]$$

Instead of [5.1], consider the following symmetrical set of (dependent) conditions

$$\frac{p_{ijk}\, p_{i'j'k}}{p_{i'jk}\, p_{ij'k}} = \frac{p_{ijk'}\, p_{i'j'k'}}{p_{i'jk'}\, p_{ij'k'}} \quad (\text{all } i \neq i',\ j \neq j',\ k \neq k') . \qquad [5.2]$$

In the analysis of variance, interactions of any order are defined recursively; more precisely, the Nth order interaction can be defined as the difference between two (N-1)th order interactions. Consider a three-factor experiment in which the three factors A, B, C have r, c, t levels respectively. Let μ_{ijk} denote the mean value of the variable under investigation for the cell $A_i \cap B_j \cap C_k$, and let $\mu_{\cdot jk}, \mu_{i\cdot k}, \mu_{ij\cdot}, \mu_{i\cdot\cdot}, \mu_{\cdot j\cdot}, \mu_{\cdot\cdot k}, \mu_{\cdot\cdot\cdot}$ denote the usual averages of these means. The zeroth-order interaction, usually called the main effect, of A_i is defined as $\mu_{i\cdot\cdot} - \mu_{\cdot\cdot\cdot}$. The first-order interaction of A_i and B_j is $k(\mu_{ij\cdot} - \mu_{\cdot j\cdot}) - (\mu_{i\cdot\cdot} - \mu_{\cdot\cdot\cdot})$, that is the difference between the zeroth-order interaction of A_i within B_j and the (marginal) zeroth-order interaction of A_i. The second-order interaction of A_i, B_j and C_k is

$$(\mu_{ijk} - \mu_{\cdot jk} - \mu_{i\cdot k} + \mu_{\cdot\cdot k}) - (\mu_{ij\cdot} - \mu_{\cdot j\cdot} - \mu_{i\cdot\cdot} + \mu_{\cdot\cdot\cdot}),$$

the difference between the first-order interaction of A_i and B_j within C_k and the marginal first-order interaction of A_i and B_j. Thus, if there is no second-order interaction in the entire experiment,

$$\mu_{ijk} = \mu_{\cdot jk} + \mu_{i\cdot k} + \mu_{ij\cdot} - \mu_{i\cdot\cdot} - \mu_{\cdot j\cdot} - \mu_{\cdot\cdot k} + \mu_{\cdot\cdot\cdot} \qquad [5.3]$$

for all i, j, k.

Equations [5.2] and [5.3] are thus the respective definitions of no second-order interaction and, as they stand, the resemblance between them is not very strong. Apart from the difference that in [5.2] the probabilities are multiplied and divided whereas in [5.3] the means are added and subtracted, the main difference is that, whereas [5.3] involves single levels A_i, B_j, C_k, [5.2] involves pairs of levels A_i, $A_{i'}$, B_j, $B_{j'}$, C_k, $C_{k'}$. However, it is possible to reformulate [5.3] to bring the definitions closer together.

Let us call $\mu_{i\cdot\cdot} - \mu_{i'\cdot\cdot}$ the relative effect or relative zeroth-order interaction of A_i, $A_{i'}$. Then the relative first-order interaction between A_i, $A_{i'}$ and B_j, $B_{j'}$ is the difference between the zeroth-order interaction of A_i, $A_{i'}$ specific to B_j and the zeroth-order interaction of A_i, $A_{i'}$ specific to $B_{j'}$, namely

$$\mu_{ij.} - \mu_{i'j.} - \mu_{ij'.} - \mu_{i'j'.} .$$

The relative second-order interaction between A_i, $A_{i'}$, B_j, $B_{j'}$ and C_k, $C_{k'}$ is built up in a similar way and is zero for all levels when

$$\mu_{ijk} - \mu_{i'jk} - \mu_{ij'k} + \mu_{i'j'k} = \mu_{ijk'} - \mu_{i'jk'} - \mu_{ij'k'} + \mu_{i'j'k'} , \quad [5.4]$$

for all $i \neq i'$, $j \neq j'$, $k \neq k'$. Condition [5.4] is equivalent to [5.3] and it is easily seen that relative interactions are formed from differences of "absolute" interactions while absolute interactions are averages of relative interactions.

The analogy between [5.2] and [5.4] is obvious. Equation [5.2] can therefore be arrived at by calling $P(A_i)/P(A_{i'})$ the relative zeroth-order interaction of A_i, $A_{i'}$. Next,

$$\{P(A_i|B_j)/P(A_{i'}|B_j)\}/\{P(A_i|B_{j'})/P(A_{i'}|B_{j'})\} = (p_{ij.}\ p_{i'j'.})/(p_{i'j.}\ p_{ij'.})$$

is the relative first-order interaction of A_i, $A_{i'}$ and B_j, $B_{j'}$. If there is no first-order interaction, that is, if this ratio is one for all $i \neq i'$, $j \neq j'$, then classifications A and B are independent.

Continuing, the relative second-order interaction of A_i, $A_{i'}$, B_j, $B_{j'}$ and C_k, $C_{k'}$ is

$$\left(\frac{p_{ijk}\, p_{i'j'k}}{p_{i'jk}\, p_{ij'k}}\right) \Big/ \left(\frac{p_{ijk'}\, p_{i'j'k'}}{p_{i'jk'}\, p_{ij'k'}}\right)$$

and, when there is no interaction, [5.2] holds.

REFERENCES

(1) Anderson, R. L. and T. A. Bancroft. 1952. Statistical theory in research. McGraw-Hill Book Company, Inc., New York.

(2) Bartlett, M. S. 1935. Contingency table interactions. Journal of the Royal Statistical Society, Supplement 2, 248-52.

(3) Darrock, J. N. 1962. Interactions in multi-factor contingency tables. Journal of the Royal Statistical Society, Series B, 24: 251-63.

(4) Fisher, R. A. 1922. On the interpretation of chi-square from contingency tables, and the calculation of P . Journal of the Royal Statistical Society, 85:87-94.

(5) Goodman, L. A. 1963a. On methods for comparing contingency tables. Journal of the Royal Statistical Society, Series A, 126: 94-108.

(6) Goodman, L. A. 1963b. On Plackett's test for contingency table interactions. Journal of the Royal Statistical Society, Series B, 25:179-88.

(7) Goodman, L. A. 1964. Simple methods for analyzing three-factor interaction in contingency tables. Journal of the American Statistical Association, 59:319-52.

(8) Kendall, M. G. and A. Stuart. 1961. The advanced theory of statistics. Volume 2. Charles Griffin and Company, London.

(9) Kastenbaum, M. A. and D. E. Lamphiear. 1959. Calculation of chi-square to test the no three-factor interaction hypothesis. Biometrics, 15:107-15.

(10) Kimball, A. W. 1954. Short-cut formulas for the exact partition of chi-square in contingency tables. Biometrics, 10: 452-58.

(11) Lancaster, H. O. 1951. Complex contingency tables treated by the partition of chi-square. Journal of the Royal Statistical Society, Series B, 13:242-9.

(12) Lancaster, H. O. 1949. The derivation and partition of in certain discrete distributions. Biometrika, 36:117-29.

(13) Lewis, B. N. 1962. On the analysis of interaction in multi-dimensional contingency tables. Journal of the Royal Statistical Society, Series A, 125:88-117.

(14) Norton, H. W. 1945. Calculation of chi-square from complex contingency tables. Journal of the American Statistical Association, 40:251-9.

(15) Ostle, B. 1954. Statistics in Research. Iowa State University Library, Ames.

(16) Plackett, R. L. 1962. A note on interactions in contingency tables. Journal of the Royal Statistical Society, Series B, 24:162-6.

(17) Roy, S. N. and M. A. Kastenbaum. 1956. On the hypothesis of no interaction in a multi-way contingency table. Annals of Mathematical Statistics, 27:749-57.

(18) Snedecor, G. W. 1958. Chi-squares of Bartlett, Mood, and Lancaster in a 2 contingency table. Biometrics, 14:560-2.

(19) Snedecor, G. W. 1946. Statistical methods. Fourth Edition. Iowa State University Library, Ames.

(20) Tocher, K. D. 1950. Extension of the Neyman-Pearson theory of tests to discontinuous variates. Biometrika, 37:130-44.

(21) Woolf, B. 1955. On estimating the relation between blood group and disease. Annals of Human Genetics, London, 19:251-3.

(22) Yates, F. 1934. Contingency tables involving small numbers and the chi-square test. Journal of the Royal Statistical Society, Supplement, 1:217.

(23) Yule, G. U. 1912. On the methods of measuring association between two attributes. Journal of the Royal Statistical Society, 75:579.