8 1

METHODS FOR MULTIDIMENSIONAL CONTINGENCY TABLE ANALYSIS

by

GERALDINE HILKER

B.A., Kearney State College, 1973

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1975

Approved by:

K. E. Kemp
Major Professor

TABLE OF CONTENTS

1. Introduction

The purpose of this paper is to synthesize the diverse information in the literature of multidimensional contingency tables and to compare the techniques of analysis, computational complexity and types of hypotheses.

Multidimensional contingency table analysis was largely ignored until Bartlett's paper appeared in 1935. Since then numerous papers have been written attempting to analyze contingency tables in many fashions. Most of the methods are asymptotically equivalent (Berkson (1972)). The methods of maximum likelihood, minimum discrimination, and the general linear model are outlined in this paper. References are given for other methods as well. Estimation procedures are discussed and examples are given to test the hypothesis of no three-way interaction.

Contingency tables are characterized by two types of data: 1) factor - classification of the unit to which the experimental unit belongs and, 2) response - classification of what happened to the experimental unit. Hence, the data are the frequencies with which experimental units fall into the various combinations of factor-response categories.

The term 'interaction' has been used in different contexts in the statistical literature. Bhapkar and Koch (1968) list the four situations that arise in multidimensional contingency tables and the interpretation of the 'no interaction' hypothesis for each situation.

Model 1 - multi-response and no factor

a) concerned with the relationship between different responses

b) "No interaction" hypotheses pertain to whether some measure of association among the members of a certain set of responses depends upon the categories of the members of a subset of the remaining responses.

Model 2 - multi-response and one factor

a) concerned with the effect of the factor in the joint and marginal distribution of the responses and on measures of association among the responses.

b) "No interaction" hypotheses pertain to whether some measure of association among the members of a certain set of responses depends upon the categories of the members of a subset of the remaining responses and/or the categories of the factor.

Model 3 - multi-response and multi-factor

a) concerned with both the relationship among the responses and the way in which the factors combine.

b) "No interaction" hypothesis can take many forms. These forms question the pattern of association among responses and whether factor combinations affect the response distribution.

Model 4 - one response and multi-factor

a) concerned with the way that the factors influence the response (analogous to univatiate analysis of variance).

b) "No interaction" hypothesis pertains to the manner in which factors combine to determine the response distribution.

In all of the above models the marginals determined by adding factor frequencies are assumed to be fixed. Marginals determined

THIS BOOK CONTAINS NUMEROUS PAGES WITH THE PAGE NUMBERS CUT OFF

THIS IS AS RECEIVED FROM THE CUSTOMER

from response frequencies are variable. Also, the total sample size is fixed prior to conducting the experiment.

The procedure to be followed in contingency table analysis is: 1) specify the analysis method to be used, 2) formulate hypotheses of interest, 3) compute expected frequencies (if the method requires it) and 4) interpret the results.

2. Methods

Many methods have been proposed to analyse multidimensional contingency tables. The traditional Pearson's chi-square statistic

$$X_p^2 = \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

has been criticized because it doesn't produce any estimates of the effects that the variables have on each other. Three proposed methods of multidimension contingency table analysis are: 1) maximum likelihood; 2) minimum discrimination information; and 3) the general linear model. The background for these methods will now be presented. Later, examples will be given and other methods of analysis will be mentioned.

2.1 Maximum Likelihood

The maximum likelihood approach discussed here will be based on logits as given by Goodman (1970). Refer to Fryer (1966) for a discussion of logits. This type of analysis is appropriate for data of the Model 4 type and should be used primarily for stratified samples. For notational convenience a three-way contingency table ($I \times J \times K$) will be discussed.

Notation:

$i=1,2,,I$ $\qquad$ $j=1,2,,J$ $\qquad$ $k=1,2,,K$

n_{ijk} - observed frequency in cell (ijk)

$\sum_{i=1}^{I} n_{ijk} = n_{.jk}$ - marginal total of the j^{th} column in the k^{th} layer

$\sum_{j=1}^{J} n_{kjk} = n_{i.k}$ - marginal total of the i^{th} row in the k^{th} layer

$n_{ij.}$ is defined similarly

$\sum_{i=1}^{I} \sum_{j=1}^{J} n_{ijk} = n_{..k}$ - total of the k^{th} layer

$n_{i..}$ and $n_{.j.}$ are defined similarly

$\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} n_{ijk} = n_{...} = N$ - grand total

P_{ijk} - true cell probability, $P_{ijk} > 0$, $\sum\sum\sum P_{ijk} = 1$

p_{ijk} - observed cell probability

$v_{ijk} = \ln P_{ijk}$ are the logits, where ln refers to the natural logarithm

As in analysis of variance, v_{ijk} can be broken down as

$$v_{ijk} = u + w_i^A + w_j^B + w_k^C + w_{ij}^{AB} + w_{ik}^{AC} + w_{jk}^{BC} + w_{ijk}^{ABC}$$

where A, B, and C denote the three variables and $\sum w_i^A = \sum w_j^B = \sum w_k^C = \ldots \sum\sum\sum w_{ijk}^{ABC} = 0$.

Then the w's represent the possible effects of the three variables on v_{ijk}. The interpretation of the w's is identical to analysis of variance:

w_i^A, w_j^B, and w_k^C are main effects

w_{ij}^{AB}, w_{ik}^{AC} and w_{jk}^{BC} are 2-way interaction effects

w_{ijk}^{ABC} is the 3-way interaction effect

$$w_i^A = v_{i..} - v_{...}$$

$$w_j^B = v_{.j.} - v_{...}$$

$$\vdots$$

$$w_{ij}^{AB} = v_{ij.} - v_{i..} - v_{.j.} + v_{...}$$

$$\vdots$$

$$w_{ijk}^{ABC} = v_{ijk} - v_{ij.} - v_{i.k} - v_{.jk} + v_{i..} + v_{.j.} + v_{..k} - v_{...}$$

where the dot subscript on the v's denotes the averages over the corresponding index ($v_{i.} = \sum_{j=1}^{J} v_{ij}/J$). For a given hypothesized distribution each w estimate ($\hat{w}$) can be expressed as

$\hat{w} = \sum\sum\sum a_{ijk} \ln n_{ijk}$, where a_{ijk} are constants whose values depend on the null hypothesis being tested and $\sum\sum\sum a_{ijk} = 0$.

2.2 Minimum Discrimination Information

The principle of minimum discrimination information (MDI) estimation is similar to the forward type of regression analysis. It has been suggested by Margolin, et al (1974) that the MDI statistic should be avoided in small samples because it is less conservative than Pearson's chi-square statistic. In Ireland and Kullback (1968b) the following summary of the principle for a three-way table is given. The proofs appear in Ireland and Kullback (1968a).

Notation:

P-table - entries are P_{ijk} which are either observed entries or estimated i=1, 2,,I; j=1,2,,J; k-1,2,,K

$P_{ijk} > 0, \sum\sum\sum P_{ijk} = 1$ The P-table entries correspond to an hypothesized distribution for the data.

p-table - entries are p_{ijk} that satisfy certain conditions of interest, $\sum\sum\sum p_{ijk} = 1$ (observed probabilities)

p^*-table - entries are p^*_{ijk} which most closely resemble P_{ijk} by the principle of MDI (estimated probabilities)

Theorem 1

Let the P-contingency table be given. Consider all p-contingency tables of the same dimension such that the marginal probabilities are given and fixed. Then the minimum value of the discrimination information

$$I(p:P) = \sum\sum\sum p_{ijk} \ln(p_{ijk}/P_{ijk})$$

is attained for $p_{ijk} = p^*_{ijk} = a_i b_j c_k P_{ijk}$ where the a_i's, b_j's and c_k's are determined subject to the marginal probability restrictions. Then the minimum value is

$$I(p^*,P) = \sum_{i=1}^{I} p_{i..} \ln a_i + \sum_{j=1}^{J} p_{.j.} \ln b_j + \sum_{k=1}^{K} p_{..k} \ln c_k$$

Theorem 2

The p^* contingency table can be computed by an iterative procedure that satisfies the marginal restrictions. (This procedure will be discussed using frequencies rather than proportions. The initial value is $p^{(0)}_{ijk} = P_{ijk}$.)

Theorem 3

If P_{ijk} in Theorem 1 is $\hat{p}_{ijk} = n_{ijk}/N$ where n_{ijk} is the number of observations in cell (ijk) and $\sum\sum\sum n_{ijk} = N$ then the minimizing set $\hat{p}^*_{ijk}$ is a best asymptotically normal estimate and the MDI statistic for goodness of fit

$$2NI(\hat{p}^* : \hat{p}) = 2N \sum\sum\sum \hat{p}^*_{ijk} \ \ln(\hat{p}^*_{ijk}/\hat{p}_{ijk})$$

is asymptotically chi-square with I + J + K - 3 degrees of freedom.

Theorem 4

The following is true for p^* computed according to Theorem 2 where the p^*-table and p-table have common specified marginals

$$I(p:P) = I(p^*:P) + I(p:p^*)$$

2.3 General Linear Model

Grizzle, Starmer, and Koch (1969) introduced a linear model approach to contingency table analysis that does not require iterative procedures to estimate cell probabilities. Since then numerous articles have been written describing this as the unified approach to an ysis of mu tidimensional contingency tables. Grizzle, Starmer, and Koch (GSK) applied methods of linear regression and weighted least squares to test hypotheses using a minimum modified chi-square test statistic or the equivalent Wald's statistic.

Notation:

$i=1,2,,,s$ population factors

$j=1,2,,,r$ response categories

P_{ij} - true cell probability $i=1,2,,s$: $j=1,2,,r;\sum\sum\sum P_{ijk} = 1.$

n_{ij} - ob ved frequency

$\hat{P}_{ij} = n_{ij}/n_{i.}$ - unbiased estimate of P_{ij}

$\underline{P}_i' = (P_{i1}, P_{i2} iii P_{ir})$ $i=1,2,,s$

$\underline{P}' = (\underline{P}_1', \underline{P}_2',,,,\underline{P}_r')$

$\hat{\underline{p}}_i'$ and $\hat{\underline{p}}'$ have corresponding notation

$\mathrm{Var}(\hat{\underline{P}}_i) = \mathrm{Var}(\hat{\underline{p}}_i) = (\, D(\hat{\underline{p}}_i) - \hat{\underline{p}}_i\hat{\underline{p}}_i')/n_{i.}$

where $D(\hat{\underline{p}}_i)$ is a diagonal matrix with the elements of $\hat{\underline{p}}_i$ on the diagonal

$\mathrm{Var}(\hat{\underline{P}}) = \mathrm{Var}(\hat{\underline{p}})$ - block diagonal with $\mathrm{Var}(\hat{\underline{p}}_i)$ on the main diagonal

$f(P)$ - any function of P such that the continuous first and second derivatives exist with respect to the P_{ij}'s

$f(\hat{p}) = E(f(P))$

$\underline{F}' = (f_1(\hat{\underline{p}}), f_2(\hat{\underline{p}}),,,,f_t(\hat{\underline{p}})) \quad t \leq (r-1)s$

$\underline{H} = (dF(\underline{x})/d\underline{x} \mid \underline{x} = \hat{\underline{p}})$

$\underline{V} = \mathrm{var}(\underline{F}) = \underline{H}\ \mathrm{Var}(\hat{\underline{p}})\ \underline{H}'$

Define a linear model $F(\underline{P}) = \underline{XB}$ where $\underline{X}$ is a known $(t \times u)$ design matrix and $\underline{B}$ is a $(u \times 1)$ vector of unknown parameters. The weighted least squares estimate of B (b) is

$$\underline{b} = (\underline{X}'\ \underline{V}^{-1}\ \underline{X})^{-1}\ \underline{X}'\underline{V}^{-1}\underline{F}$$

The Wald test statistic for goodness of fit is

$$\chi^2_\omega = (\underline{F} - \underline{Xb})'\underline{V}^{-1}(\underline{F} - \underline{Xb})$$

with $(t-u)$ degrees of freedom.

3. Formulation of Hypotheses

Forming meaningful and interpretable hypotheses in a multidimensional contingency table is no easy task. The interpretation of the "no interaction" hypotheses for the four types of models was explained in the Introduction. Hypotheses of interest pertaining to the three methods of analysis will be presented.

3.1 Maximum Likelihood

A maximum likelihood hypothesis is formulated from a hierarchial model. An hierarchial model is defined such that, if any w-term is set to zero all the higher w-terms involving that particular one are also set to zero, and if that w-term is not zero, all the lower terms must be present.

Example: If in the model

$$v_{ijk} = u + w_i^A + w_j^B + w_k^C + w_{ij}^{AB} + w_{ik}^{AC} + w_{jk}^{BC} + w_{ijk}^{ABC},$$

w_{ik}^{AC} is set to zero then w_{ijk}^{ABC} must also be zero. If w_{ik}^{AC} is not zero then w_i^A and w_k^C must be in the model.

This must be taken into consideration when building hypotheses. The following is a list of possible hypotheses for the three-way table from Goodman (1970).

	Parameters set to zero	Degrees of freedom	# of hypotheses of this kind	Basic Set
1.	w_{ijk}^{ABC}	(I-1)(J-1)(K-1)	1	w_{ij}^{AB} w_{ik}^{AC} w_{jk}^{BC}
2.	w_{ijk}^{ABC} w_{jk}^{BC}	I(J-1)(K-1)	3	w_{ij}^{AB} w_{jk}^{BC}
3.	w_{ijk}^{ABC} w_{jk}^{BC} w_{ik}^{AC}	(IJ-1)(K-1)	3	w_{ij}^{AB} w_k^C
4.	w_{ijk}^{ABC} w_{jk}^{BC} w_{ik}^{AC} w_k^C	IJ(K-1)	3	w_{ij}^{AB}
5.	w_{ijk}^{ABC} w_{jk}^{BC} w_{ik}^{AC} w_{ij}^{AB}	IJK-I-J-K+2	1	w_i^A w_j^B w_k^C
6.	w_{ijk}^{ABC} w_{jk}^{BC} w_{ik}^{AC} w_{ij}^{AB} w_k^C	IJK-I-J+1	3	w_i^A w_j^B
7.	w_{ijk}^{ABC} w_{jk}^{BC} w_{ik}^{AC} w_{ij}^{AB} w_k^C w_j^B	I(JK-1)	3	w_i^A
8.	w_{ijk}^{ABC} w_{jk}^{BC} w_{ik}^{AC} w_{ij}^{AB} w_k^C w_j^B w_i^A	IJK-1	1	

Hypothesis 2 states that variables B and C are independent given the category of variable A. Hypothesis 5 states that the three variables are mutually independent. In hypothesis 6, two items are hypothesized: 1) variables A and B are independent and 2) the categories of variable C are equiprobable given the joint category of A and B.

The degrees of freedom can be calculated in two ways: 1) the number of w's in the basic sets that are set to zero by the null hypothesis or 2) (IJK-1) minus the number of w's in the basic sets that are not set to zero by the hypothesis.

For information to partition a hypothesis (partition into hypotheses about marginal tables, independence, conditional independence, etc.) refer to Goodman (1971b). Goodman (1971c) presents a stepwise procedure to obtain estimates either with quantitative or qualitative variables and exhibits methods for analysis of Model 1 and Model 4 type data. In Goodman (1971c) the methods developed by Goodman are applied to a survey analysis.

3.2 Minimum Discrimination Information

By Theorem 4 an analysis of information table can be constructed where both the information components and the degrees of freedom are additive. Ku, et al (1971) showed that if n is the observed frequency and n_b^* is estimated, then under a set of marginal constraints that include the marginals used in estimating n_a^*

$$\begin{aligned} 2I\,(n:n_a^*) &= 2\textstyle\sum n \ln (n/n_a^*) \\ &= 2I\,(n_b^* : n_a^*) + 2I\,(n:n_b^*) \\ &= 2\textstyle\sum n_b^* \ln(n_b^*/n_a^*) + 2\textstyle\sum n \ln (n/n_b^*) \end{aligned}$$

This additive property allows the total information to be analyzed using additive components similar to analysis of variance. Information statistics such as $2I\ (n{:}n_a^*)$ that compare an observed table with an expected table are measures of interaction. Information statistics comparing two expected tables ($2I\ (n_b^*{:}n_a^*)$) are called measures of effect, that is, the effect of adding additional constraints in estimating n_b^* not contained in the set of constraints for n_a^*.

To clarify this, assume that the null hypothesis is H_o: $p_{ijk} = p_{i..}p_{.j.}p_{..k}$ (independence among the three-way cell entries in a three-way table). The comparison is between an observed table and an expected table and, therefore, the information statistic measures interaction. H_o usually doesn't specify $p_{i..}$, $p_{.j.}$, and $p_{..k}$. Thus, H_o can be partitioned into additive components specifying each of the above marginals.

H_r: $p_{i..} = r$ $\quad i=1,2,,I$ $\quad$ (row)

H_c: $p_{.j.} = s$ $\quad j=1,2,,J$ $\quad$ (column)

H_d: $p_{..k} = t$ $\quad k=1,2,,K$ $\quad$ (depth)

where r,s, and t are given constraints. A table to give the information due to interaction can be constructed as follows:

Component due to	Information	Degrees of Freedom
Rows	$2I(p_{i..}{:}r)$	I-1
Columns	$2I(p_{.j.}{:}s)$	J-1
Depth	$2I(p_{..k}{:}t)$	K-1
Independence	$2I(p_{ijk}{:}rst)$	IJK-I-J-K+2
Total	$2I(p^*{:}p)$	IJK-1

If H_0 is true then $2I(p_{ijk}:rst)$ is asymptotically chi-square with IJK-I-J-K+2 degrees of freedom.

The null hypothesis H_0: $p_{ijk}=p_{i..}p_{.jk}$ tests whether or not the row classification is independent of column and depth classifications since

$$p_{ij.} = p_{ijk} = p_{i..}p_{.jk} = p_{i..}p_{.j.}$$

$$p_{i.k} = p_{ijk} = p_{i..}p_{.jk} = p_{i..}p_{..k}.$$

Again, $p_{i..}$ and $p_{.jk}$ are usually not specified in H_0 and $2I\ (p^*:p)$ can be partitioned into additive components. These components specify values of $p_{i..}$, $p_{.jk}$ and $p_{i..}p_{.jk}$. Kullback (1959) gives details for other hypotheses to test.

3.3 General Linear Model

Assume that the model is

$$F(\underline{P}) = \underline{XB}$$

where $F(\underline{P})$ is a (txl) vector of functions of $\underline{P}$, $\underline{X}$ is a (txu) known design matrix and $\underline{B}$ is a (uxl) vector of unknown parameters. Given that the above model fits the data, tests can be performed concerning parameters and/or combinations of parameters by hypotheses of the form

$$H_0 : \underline{CB} = 0,$$

where $\underline{C}$ is a (dxu) matrix of constants. This may be accomplished by the usual methods of weighted multiple regression.

The test statistic is

$$SS(\underline{CB} = 0) = \underline{b}'\underline{C}'(\underline{C}(\underline{X}'\underline{V}^{-1}\underline{X})^{-1}\underline{C}')^{-1}\underline{Cb},$$

where $\underline{V}$ is the estimated covariance matrix of $F(\underline{P})$. This test statistic is compared to the tabled chi-square value with d degrees of freedom.

GSK discuss examples with one and more populations. Forthofer, et al (1971b) extend the work of GSK to more complex problems involving association, rank correlation and "ridits" (ordered data). In Johnson and Koch (1970) qualitative data obtained by a stratified random sample is analysed by the GSK methods. Johnson and Koch (1971); Koch, et al (1972); and Koch and Tolley (1975) analyse problems using the weighted least squares approach. In recent statistical literature the analysis of contingency tables by log-linear models has been developed. Bishop, et al (1975) discuss this thoroughly. Ku and Kullback (1974) apply the principle of minimum discrimination to the log-linear model.

4. Estimation of Cell Frequencies

There are basically three methods of obtaining estimates:

1) direct procedure;

2) maximum likelihood equation solving;

3) iterative proportional fitting algorithm.

When direct estimates exist (refer to Bishop (1975), pages 73-83) the maximum likelihood estimate can be obtained from the familar formula

$$\frac{(\text{row total})(\text{column total})}{(\text{grand total})}$$

with a two-way table and the hypothesis of no interactions.

The second procedure involves solving a system of simultaneous equations which increase in difficulty as the dimensions increase and, therefore, will not be illustrated.

Maximum likelihood estimates can be obtained for any hierarchial model by iterative fitting under the conditions imposed by the null hypothesis.

The iterative scheme successively adjusts the marginals proportionally. If direct estimates exist, they will be obtained after one cycle. For more details refer to Bishop (1975), pages 83-97.

Iterative algorithm for a three-way (I x J x K) table:

1. Start with an initial value - usually $n_{ijk}^{(0)} = 1$ or $N/(I \times J \times K)$. For restrictions on initial values see Bishop (1975), pages 92-95.
2. Compute the marginals $n_{ij.}^{(0)}$
3. Adjust $n_{ijk}^{(0)}$ by $n_{ij.}/n_{ij.}^{(0)}$ (observed/expected)
   Adjusted entries are $n_{ijk}^{(1)}$
4. Compute the marginals $n_{i.k}^{(1)}$
5. Adjust $n_{ijk}^{(1)}$ by $n_{i.k}/n_{i.k}^{(1)}$ Adjusted entries are $n_{ijk}^{(2)}$
6. Compute the marginals $n_{.jk}^{(2)}$
7. Adjust $n_{ijk}^{(2)}$ by $n_{.jk}/n_{.jk}^{(2)}$ Adjusted entries are $n_{ijk}^{(3)}$
8. Continue steps (2) through (7) using $n_{ijk}^{(3)}$ as the starting entry.
9. Continue the process until successive cycles differ by no more than some specified range (.01, .001, etc.).

The MDI method uses the statistic

$$\chi_I^2 = 2N\sum\sum\sum p_{ijk}^{*} \ln(p_{ijk}^{*} / p_{ijk})$$
$$= 2\sum\sum\sum n_{ijk}^{*} \ln(n_{ijk}^{*} / n_{ijk})$$

where p_{ijk}^{*} (n_{ijk}^{*}) are the estimates obtained by the iterative procedure and $p_{ijk}(n_{ijk})$ are observed values.

For the data given in Table 1 below, the Pearson, maximum likelihood, and MDI χ^2 values are .851, .853, and .856, respectively, for testing the hypotheses of three-way interaction. The tabled chi-square value with (I-1)(J-1)(K-1) = 2 degrees of freedom and a .95 probability level is 5.99. Therefore, the hypothesis of no three-way interaction is accepted. However, one cannot assume that the three classifications are independent. Perhaps an hypothesis to test the association between birth order and number of losses would be of interest.

TABLE 1

Data on the Number of Mothers with Previous Infant Losses, Grizzle (1961)

Birth Order		Number of mothers with Losses	No Losses
2	Problem	20	82
	Control	10	54
3-4	Problem	26	41
	Control	16	30
5^+	Problem	27	22
	Control	14	23

i=1(problem), 2(control)

j=1(losses), 2(no losses)

k=1(2), 2(3-4), 3(5^+)

Sets of observed marginals:

$n_{ij.}$	
73	145
40	107

$n_{i.k}$		
102	67	49
64	46	37

$n_{.jk}$		
30	42	41
136	71	45

TABLE 2

Iterative Values, Ku, et al (1971)

cell	n_{ijk}	$n_{ijk}^{(0)}$	$n_{ijk}^{(1)}$	$n_{ijk}^{(2)}$	$n_{ijk}^{(3)}$	....	$n_{ijk}^{(14)}$	$n_{ijk}^{(15)}$
111	20	30.416	24.333	34.156	19.869		20.503	20.503
112	26	30.416	24.333	22.435	26.959		27.213	27.213
113	27	30.416	24.333	16.408	25.410		25.284	25.284
121	10	30.416	48.333	67.844	80.633		81.497	81.497
122	16	30.416	48.333	40.564	40.540		39.787	39.787
123	14	30.416	48.333	32.592	24.639		23.716	23.716
211	82	30.416	13.333	17.415	10.130		9.497	9.497
212	41	30.416	13.333	12.517	15.041		14.787	14.787
213	22	30.416	13.333	10.067	15.590		15.716	15.716
221	54	30.416	35.667	46.585	55.367		54.503	54.503
222	30	30.416	35.667	33.483	30.369		31.213	31.213
223	23	30.416	35.667	26.932	20.361		21.284	21.284

An example will be given to clarify the procedure using the data of Table 1. The following illustrates the first cycle to fit the hypotheses of no three-way interaction.

1. $n_{ijk}^{(0)} = 365/(2x2x3) = 30.416 \quad i=1,2;\ j=1,2:\ k=1,2,3$

2. $n_{ij.}^{(0)} = 30.416 + 30.416 + 30.416 = 91.248 \quad i=1,2;\ j=1,2$

3. $n_{111}^{(1)} = n_{111}^{(0)} \times (n_{11.}/n_{11.}^{(0)}) = 30.416 \times (73/91.248) = 24.333$

$n_{112}^{(1)} = n_{112}^{(0)} \times (n_{11.}/n_{11.}^{(0)}) = 30.416 \times (73/91.248) = 24.333$

$\vdots$

$n_{223}^{(1)} = n_{223}^{(0)} \times (n_{22.}/n_{22.}^{(0)}) = 30.416 \times (107/91.248) = 35.667$

4. $n_{1.1}^{(1)} = n_{111}^{(1)} + n_{121}^{(1)} = 24.333 + 48.333 = 72.667$

$\vdots$

$n_{2.3}^{(1)} = n_{213}^{(1)} + n_{223}^{(1)} = 13.333 + 35.667 = 49$

5. $n_{111}^{(2)} = n_{111}^{(1)} \times (n_{1.1}/n_{1.1}^{(1)}) = 24.333 \times (102/72.667) = 34.156$

$\vdots$

$n_{223}^{(2)} = n_{223}^{(1)} \times (n_{2.3}/n_{2.3}^{(1)}) = 35.667 \times (37/49) = 26.932$

6. $n_{.11}^{(2)} = n_{111}^{(2)} + n_{211}^{(2)} = 34.156 + 17.415 = 51.571$

$\vdots$

$n_{.23}^{(2)} = n_{123}^{(2)} + n_{223}^{(2)} = 32.592 + 26.932 = 59.524$

7. $n_{111}^{(3)} = n_{111}^{(2)} \times (n_{.11}/n_{.11}^{(2)}) = 34.156 \times (30/51.571) = 19.869$

$\vdots$

$n_{223}^{(3)} = n_{223}^{(2)} \times (n_{.23}/n_{.23}^{(2)}) = 26.932 \times (45/59.524) = 20.361$

The first cycle of the iterative proportional fitting algorithm is completed. The cycles are continued until the change between successive cycles is less than a specified tolerance. Five cycles are required to obtain a change no greater than 0.001.

5. Interpretation of Results

Once an hypothesis has been formulated and the computations are performed, the results remain to be interpreted. A no interaction hypothesis takes on different meanings as discussed in the Introduction, depending on the model being used.

Consider a test for no three-way interaction in a three dimensional contingency table with A, B, and C as the three classifications. It is not necessary to begin the analysis with this particular hypothesis. Regardless of an acceptance or rejection of the hypothesis of no three-way interaction, further hypotheses concerning other relationships should be tested. If the hypothesis of no three-way interaction is accepted, one should not assume that all three classifications are independent. Perhaps two classifications interact, but are independent of the third. Then three more hypotheses may be of interest: 1) the association between classification A and classification B is the same for all categories of C; 2) the association between classification A and classification C is the same for all categories of B; 3) the association between classification B and C is the same for all categories of A. These hypotheses test whether or not one classification is independent of the other two. If one of the above hypotheses is rejected, then an hypothesis testing whether or not one classification is independent of one other classification (rather than two) can be made.

With the linear model approach, parameters are added to the hypothesized model until the model fits the data. Then various hypotheses concerning constraints on the model parameters may be of interest, such as hypotheses of no linear effects on each of the classifications. Once a model is fitted, it is desirable to have the test of the null hypothesis on each separate effect in the model individually significant. Koch and Reinfurt (1970) suggest deleting the nonsignificant effects from the model but warn that this may affect subsequent tests of significance. For details the reader should see Koch and Reinfurt (1970).

6. Examples

The data of Table 1 was taken from Baltimore schools. Information recorded was: 1) whether the child has been referred by teachers as presenting behavior problems or was in the control group, 2) whether or not the mother suffered any infant losses prior to the birth of this child and 3) the birth order of the child in the study. The hypothesis of no three way interaction will be discussed for the methods of Pearson, maximum likelihood, and minimum discrimination information.

This problem was discussed by Grizzle (1961). He used a different iterative scheme than described here, but the results are equivalent. Grizzle used the Pearson chi-square test statistic

$$X_p^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{(n_{ijk} - \hat{n}_{ijk})^2}{\hat{n}_{ijk}}$$

where $\hat{n}_{ijk}$ is any maximum likelihood estimate.

An hypothesis of no three-way interaction for the maximum likelihood approach would be H_0: $w_{ijk}^{ABC} = 0$. The usual iterative procedure can be applied to obtain the cell estimates with the suggested initial value being $n_{ijk}^{(0)} = 1$. However, the estimates converge to the estimates using N/(I x J x K) as the initial value. The test statistic is

$$\chi_L^2 = 2N\sum\sum\sum p_{ijk} \ln(p_{ijk}/\hat{p}_{ijk}),$$

where ln refers to the natural logarithm and $\hat{p}_{ijk}$ is the maximum likelihood estimate obtained from the general iterative procedure or (depending on the hypothesis to be tested) the easier formulas presented in Goodman (1971a). Using frequencies rather than probabilities the formula is

$$\chi_L^2 = 2\sum\sum\sum n_{ijk} \ln(n_{ijk}/\hat{n}_{ijk}),$$

where n_{ijk} is the maximum likelihood estimate for the frequency.

TABLE 3
Data of Kastenbaum and
Lamphiear (1959)

Population	Litter size	Treatment	Number of deaths 0	1	2
1	7	A	58	11	5
2		B	75	19	7
3	8	A	49	14	10
4		B	58	17	8
5	9	A	33	18	15
6		B	45	22	10
7	10	A	15	13	15
8		B	39	22	18
9	11	A	4	12	17
10		B	5	15	8

The data in Table 3 were collected in a laboratory where the experiment was on litters of mice. The mice of different litter sizes were treated with one of two treatments and the number of deaths per litter before weaning was observed.

GSK used this example to illustrate the linear model approach to multidimensional contingency table analysis.

$r = 3$ responses

$s = 10$ populations

P_{i0}, P_{i1}, and P_{i2} are the probabilities of observing zero, one, and two or more deaths, respectively. Define logarithmic functions, $(F(P))$,

$$m_{i0} = \ln (P_{i0}/P_{i2})$$

$$m_{i1} = \ln (P_{i1}/P_{i2}),$$

where the functions are considered to be additive functions of the overall mean, treatment effect and litter size effect.
As seen below

$$m_{10} = u_0 + a_0 + b_{10},$$

where u_j = overall effect $j=0,1$

a_j - A treatment effect $j=0,1$

b_{1j} - the first litter size effect $j=0,1$

The other b_{ij}'s are defined similarly with the effect of the size of the last litter being the negative of all the others and the effect of B treatment being the negative of the A treatment.

By the above definition of the m's the following is true:

$$
\begin{bmatrix} m_{10} \\ m_{20} \\ m_{30} \\ m_{40} \\ m_{50} \\ m_{60} \\ m_{70} \\ m_{80} \\ m_{90} \\ m_{100} \\ m_{11} \\ m_{21} \\ m_{31} \\ m_{41} \\ m_{51} \\ m_{61} \\ m_{71} \\ m_{81} \\ m_{91} \\ m_{101} \end{bmatrix}
=
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0 & & & & & & \\
1 & -1 & 1 & 0 & 0 & 0 & & & & & & \\
1 & 1 & 0 & 1 & 0 & 0 & & & & & & \\
1 & -1 & 0 & 1 & 0 & 0 & & & & & & \\
1 & 1 & 0 & 0 & 1 & 0 & & & \underline{0} & & & \\
1 & -1 & 0 & 0 & 1 & 0 & & & & & & \\
1 & 1 & 0 & 0 & 0 & 1 & & & & & & \\
1 & -1 & 0 & 0 & 0 & 1 & & & & & & \\
1 & 1 & -1 & -1 & -1 & -1 & & & & & & \\
1 & -1 & -1 & -1 & -1 & -1 & & & & & & \\
& & & & & & 1 & 1 & 1 & 0 & 0 & 0 \\
& & & & & & 1 & -1 & 1 & 0 & 0 & 0 \\
& & & & & & 1 & 1 & 0 & 1 & 0 & 0 \\
& & & & & & 1 & -1 & 0 & 1 & 0 & 0 \\
& & & \underline{0} & & & 1 & 1 & 0 & 0 & 1 & 0 \\
& & & & & & 1 & -1 & 0 & 0 & 1 & 0 \\
& & & & & & 1 & 1 & 0 & 0 & 0 & 1 \\
& & & & & & 1 & -1 & 0 & 0 & 0 & 1 \\
& & & & & & 1 & 1 & -1 & -1 & -1 & -1 \\
& & & & & & 1 & -1 & -1 & -1 & -1 & -1
\end{bmatrix}
\begin{bmatrix} u_0 \\ a_0 \\ b_{10} \\ b_{20} \\ b_{30} \\ b_{40} \\ u_1 \\ a_1 \\ b_{11} \\ b_{21} \\ b_{31} \\ b_{41} \end{bmatrix}
$$

The general linear model is

$$F(\underline{P}) = \underline{XG}$$

where $F(\underline{P})$ is the vector of m's above and $\underline{G}$ is a matrix composed of elements that are the logarithm of the product of the identity matrix times $\underline{P}$.

$$\underline{X} = \begin{vmatrix} 1 & 0 & -1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & & 0 & 0 & 0 \\ & \vdots & & & & & & & \vdots & \\ 0 & 0 & 0 & 0 & 0 & 0 & & 1 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & & 0 & 0 & 0 \\ & \vdots & & & & & & & \vdots & \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{vmatrix}$$

Note that the first element is

$$m_{10} = \ln(P_{10}) - \ln(P_{12})$$

$$= \ln(P_{10}/P_{12}),$$

which is identical to the definition given earlier. Now, substitute the $\underline{p}$ estimates for $\underline{P}$ into the matrices and perform a weighted multiple regression analysis.

$$\underline{b} = (\underline{X}'\underline{V}^{-1}\underline{X})^{-1}\underline{X}'\underline{V}^{-1}\underline{F}$$

with $\underline{V}$ and $\underline{F}$ as described in the notation given for the linear model approach.

The test statistic is

$$\chi^2_W = SS(F(\underline{P}) = \underline{XG}) = \underline{F}'\underline{V}^{-1}\underline{F} - \underline{b}'\ (\underline{X}'\underline{V}^{-1}\underline{X})\underline{b},$$

= sum of squares of deviation from the model,

= 3.1269 with t-u = 20 - 12 = 8 degrees of freedom.

Berkson (1968) used the minimum logit chi-square and the results were equivalent. Berkson (1972) used iterative maximum likelihood estimates and obtained a chi-square value of 3.159 with the maximum likelihood method and a value of 3.175 with the MDI.

The tabled chi-square value at the .95 probability level is 15.51. Therefore, the hypothesized model fits the data. Now other tests of hypotheses can be made such as the test for no effect of treatment simultaneously on m_{io} and m_{i1}. This hypothesis is $H_0:\underline{CB} = 0$, where

$$\underline{C} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} .$$

7. Summary

In addition to the contingency table analysis problems that have already been given, there are problems of small frequency counts and ordered data. The only possible ways of handling categories with very small frequencies is to delete them from the analysis or pool them with other categories. However, the interaction between the remaining variables will be affected. This problem is discussed in Bishop (1971) and (1975). Problems also arise when there is a natural ordering of the categories and the assumption of normality can not be made This is discussed in William (1972) and Weisburg (1972).

In conclusion, there are many problems with multidimensional contingency table analysis. Recently statisticians have become more concerned with analysis of such tables. Literature on contingency tables is plentiful, but confusing, because different methods must be used to handle specific problems.

REFERENCES

1. Bartlett, M.S. (1935). Contingency table interactions. Journal of the Royal Statistical Society, Supplement 2, 248-252.

2. Berkson, Joseph (1968). Application of minimum logit estimate to a problem of Grizzle with a notation on the problem of 'no interaction'. Biometrics, 24, 78-85.

3. Berkson, Joseph (1972). Minimum discrimination information, the 'no interaction' problem, and the logistic function. *Biometrics*, 28, 443-467.

4. Bishop, Yvonne M.M. (1971). Effects of collapsing multidimensional contingency tables. Biometrics, 27, 545-562.

5. Bishop, Yvonne M.M.; Fienburg, Stephen E. and Holland, Paul W. (1975). Discrete Multivariate Analysis: Theory and Practice. Cambridge, Massachusetts MIT Press.

6. Bhapkar, V.P. and Koch, Gary G. (1968). Hypothesis of 'no interaction' in multi-dimensional contingency tables. Technometrics, 10, 107-122.

7. Bhapkar, V.P. and Koch, Gary, G. (1968). On the hypothesis of 'no interaction' in contingency tables. Biometrics, 24, 567-594.

8. Cochran, William G. (1954). Some methods for strengthening the common χ tests. Biometrics, 10, 417-451.

9. Fleiss, Joseph I. (1973). Statistical Methods for Rates and Proportions. John Wiley and Sons, New York.

10. Forthofer, Ronald N. and Koch, Gary G. (1973). An analysis for compounded functions of categorical data. Biometrics, 29, 143-157.

11. Fryer, H.C. (1966). Concepts and Methods of Experimental Statistics. *Allyn and Bacon, Inc., Boston.*

12. Goodman, Leo A. (1970). The multivariate analysis of qualitative data: interactions among multiple classifications. Journal of the American Statistical Association, 65, 226-256.

13. Goodman, Leo A. (1971a). The analysis of multidimensional contingency tables: stepwise procedure and direct estimation methods for building models for multiple classifications. Technometrics, 13, 33-61.

14. Goodman, Leo A. (1971b). Partitioning of chi-square, analysis of marginal contingency tables and estimation of expected frequencies in multi-dimensional contingency tables. Journal of the American Statistical Association, 66, 339-344.

15. Goodman, Leo A. (1971c). A general model for the analysis of surveys. American Journal of Sociology, 77, 1035-1085.

16. Goodman, Leo A. (1971d). The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classifications. Technometrics, 13, 33-61.

17. Grizzle, J.E. (1961). A new method of testing hypotheses and estimating parameters for the logistic model. Biometrics, 17, 372-385.

18. Ireland, C.T. and Kullback, S. (1968a). Contingency tables with given marginals. Biometrika, 55, 179-188.

19. Ireland, C.T. and Kullback, S. (1968b). Minimum discrimination information estimation. Biometrics, 24, 707-713.

20. Johnson, William D. and Koch, Gary G. (1971). A note on the weighted least squares analysis of the Ries-Smith contingency table data. Technometrics, 13, 438-447.

21. Johnson, William D. and Koch, Gary G. (1970). Analysis of qualitative data: linear functions. Health Services Research, Winter, 358-369.

22. Kastenbaum, M.A. and Lamphiear, D.E. (1959). Calculation of chi-square to test the no three-factor interaction hypothesis. Biometrics, 15, 107-115.

23. Koch, Gary G. and Reinfurt, Donald W. (1970). The analysis of complex contingency table data from general experimental design and sample surveys. Institute of Statistics Mimeo Series No. 716, University of North Carolina.

24. Koch, Gary G.; Johnson, William D. and Tolley, H. Dennis. (1972). A linear models approach to the analysis of survival and extent of disease in multidimensional contingency tables. Journal of the American Statistical Association, 67, 783-789.

25. Koch, Gary G. and Tolley, H. Dennis. (1975). A generalized modified χ^2 analysis of categorical bacteria survival data from a complex dilution experiment. Biometrics, 31, 59-92.

26. Ku, Harry and Kullback, Solomon. (1974). Loglinear models in contingency table analysis. The American Statistical Association, 67, 783-789.

27. Ku, Harry H.; Varner, Ruth N. and Kullback, S. (1971). On the analysis of multidimensional contingency tables. Journal of the American Statistical Association, 66, 55-64.

28. Kullback, Solomn. (1959). Information Theory and Statistics. John Wiley and Sons, Inc., New York.

29. Lewis, B.N. On the analysis of interaction in multidimensional contingency tables. Journal of the Royal Statistical Society, Ser. A 125, 88-117.

30. Margolin, Barry H. and Light, Richard J. (1974). Analysis of variance for categorical data. Journal of the American Statistical Association, 69, 755-764.

31. Weisburg, Herbert. (1972). Upper and lower probability inferences from ordered multinomial data. Journal of the American Statistical Association, 67, 884-890.

32. William, O.D. and Grizzle, J.E. (1972). Analysis of contingency tables having ordered response categories. Journal of the American Statistical Association, 67, 55-63.

ACKNOWLEDGEMENTS

I wish to thank my committee members, Dr. Milliken, Dr. Dayton, and especially my major professor, Dr. Kemp, for their suggestions in preparation of this paper.

METHODS FOR MULTIDIMENSIONAL
CONTINGENCY TABLE ANALYSIS

by

GERALDINE HILKER

B.A., Kearney State College, 1973

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY

Manhattan, Kansas

The analysis of multidimensional contingency tables is often confusing. Basically, there are three methods of analysis. These are: 1) maximum likelihood; 2) minimum discrimination information; and 3) the general linear model.

The form of the hypotheses of interest and the procedure to estimate cell frequencies varies with each method. In general, an iterative procedure can be used for the maximum likelihood and the minimum discrimination information approaches. The general linear models approach does not require prior estimates of the cell frequencies. However, knowledge of linear models is necessary.

The above three methods suggest test statistics which are asymptotically equivalent.