AD-785 599

THE INFORMATION IN CONTINGENCY TABLES

Solomon Kullback

George Washington University

Prepared for:

Army Research Office - Durham

September 1974

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

Best
Available
Copy

Unclassified
Security Classification

AD-785 599

DOCUMENT CONTROL DATA - R & D	
(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)	
1. ORIGINATING ACTIVITY (Corporate author) The George Washington University Department of Statistics	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP NA
3. REPORT TITLE THE INFORMATION IN CONTINGENCY TABLES	
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Final	
5. AUTHOR(S) (First name, middle initial, last name) Solomon Kullback	
6. REPORT DATE September 1974	7a. TOTAL NO. OF PAGES 385 7b. NO. OF REFS 316
8a. CONTRACT OR GRANT NO. DAHC04-74-G-0164 b. PROJECT NO P-12226-M c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) K-1-74 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
10. DISTRIBUTION STATEMENT Approved for public release; distribution unlimited.	
11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY U. S. Army Research Office-Durham Box CM, Duke Station Durham, North Carolina 27706
13. ABSTRACT See attached.	

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unclassified
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Contingency Table Analysis						
Minimum Discrimination Information Estimation						
Log-linear Models						
Model Fitting						
Association						
Interaction						
Iterative Scaling Algorithms						
Logit Representation						
Exponential families						

Abstract

Through the use of the principle of minimum discrimination information estimation, leading to exponential families or multiplicative models or log-linear models it is shown, using illustrative examples exhibiting different aspects of contingency table analysis, that:

(1) Estimates of the cell entries under various hypotheses or models can be obtained;

(2) The adequacy or fit of the model, or the null hypothesis, can be tested;

(3) Main effect and interaction parameters can be estimated;

(4) The structure of the table can be studied in detail in terms of the various interrelationships among the classificatory variables;

(5) The procedures can be applied to test hypotheses about particular parameters and linear combinations of parameters that are of special interest;

(6) The procedures provide indication of outlier cells;

(7) Since the procedures and concepts are based on a general principle a unified treatment of multi-dimensional contingency tables is possible;

(8) The procedure provides estimates based on an observed or sample table, which satisfy certain external hypotheses as to underlying probability relations in the population table. These estimates also preserve the inherent properties of the observed data not affected by the hypothesis;

(9) In general, the m.d.i. estimate are best asymptotically normal;

(10) The minimum discrimination information test statistics are asymptotically distributed as chi-squared with appropriate degrees of freedom;

(11) Convergent iterative computer algorithms are available for the analyses.

THE INFORMATION IN CONTINGENCY TABLES

FINAL TECHNICAL REPORT

SOLOMON KULLBACK

SEPTEMBER 1974

U.S. ARMY RESEARCH OFFICE - DURHAM

GRANT NUMBER DAHCO 4-74-G-0164

THE GEORGE WASHINGTON UNIVERSITY

DEPARTMENT OF STATISTICS

K-1-74

id

THE FINDINGS IN THIS REPORT ARE NOT TO BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, UNLESS SO DESIGNATED BY OTHER AUTHORIZED DOCUMENTS.

Foreword

This report has been made possible by the support of the U.S. Army Research Office - Durham, North Carolina for which I express my appreciation. It is also the product of the interaction among many people, including my students, colleagues, collaborators, interested statisticians, referees and editors. The support of The George Washington University in providing an academic environment in which the teaching and research, to develop, expand and use the results presented herein, was made possible, stimulated and encouraged, must be, and is, gratefully acknowledged. In particular, the support and collaboration of Professor Henry Solomon, Professor Herbert Solomon, Associate Professor D. V. Gokhale, Associate Professor C. T. Ireland, Dr. H. H. Ku, Dr. Marian R. Fisher, and Mr. John C. Keegel have contributed greatly to any merits this report possesses, its demerits are my responsibility. The many examples were analyzed on the basis of computations using the facilities of the Computer Center of The George Washington University. The research program which underlies this report began under AFOSR Grant No. 932-65, continued under Grants AFOSR-68-1513, AFOSR-72-2348 and Contract No. N00014-67-A-0214-0015 under the joint sponsorship of the Army, Navy and Air Force. To Mrs. Glenda Howell for her typing and all the others who have contributed my sincere thanks.

TABLE OF CONTENTS

1. Introduction

Data which result from experiments in the physical sciences and engineering are usually outcomes of controlled experiments, and expressible in quantitative terms. In many other fields however, the data are seldom results of controlled experiments. In addition, the observations usually can be expressed only in qualitative or categorical terms, a yes - no, alive - dead, agree - disagree, class A - class B - class C, etc. type of response.

For example, an individual may be classified by sex, by race, by profession, by smoking habit, by age, by incidence of coronary heart disease. If we take observations over a sample of many such individuals, the result will be a multidimensional contingency table with as many dimensions as there are classifications. Contingency tables are cross-classifications.of vectors of discrete random variables showing the number of subjects belonging to distinct categories of each of several qualitative or categorical classifications. The number of counts of individuals in a cell of this table represents that portion of the sample having the specific attributes within each of the classifications. A problem of interest, for example, might be to determine the factors that are associated with the presence or absence of coronary heart disease.

Data from many fields are often presented in this manner, that is, in a cross-tabulated form. Statistical analyses of these types of data has had a long history, as may be seen from the bibliography, but were mainly concerned with the simple kind, the two-way table. Analyses of multidimensional contingency tables have been investigated intensively only during the last decade or so.

Conclusions drawn from contingency tables may be only exploratory in nature. One of the difficulties can be the availability of meaningful and reliable data. The first problem one faces in the analysis of cross-classified data is the decision on the number of classifications to be included and the categories within each classification. Typical among the problems in the analysis is how to segregate the effect on the response of some of the background variables, individually or jointly, from that of the others that are of particular interest. The data analytic attitude is empirical rather than theoretical. A more empirical attitude is natural when detailed theoretical understanding is unavailable. Estimation of parameters in models should be considered less as attempts to discover underlying truths and more as data calibrating devices which make it easier to conceive of noisy data in terms of smooth distributions and relations. With a given data set, a variety of models may be tried on, and one selected on the ground of looks and fit. (See Dempster (1971).)

Consider, for example, an experiment performed to compare the effectiveness of safety release devices for refrigerators in relation to children's safety. Children between two to five years of age are induced to crawl into refrigerators equipped with six different types of release devices. If a child can open the door of the refrigerator, from inside, within a certain time period, the response is classified as a success, otherwise a failure. The background variables studied included age, sex, weight, socio-economic status of parents. The experimental variable was one of six devices. (A partial analysis of this data may be found in

Kullback et al. 1962b, p. 581) Some balancing of the background variables was achieved.

In other instances none of the factors are subject to experimental control, and whatever available data could be collected is reported. The analysis of this type of data, though it may only be seeking preliminary information can be important in fields of health and safety. The uncontrolled experimental data are sometimes the only realistic data available when these data deal with life, death, health, and safety, and some of these factors and responses are only expressible in qualitative terms, in the present state of art.

It is expected that the number of problems calling for the techniques of the analysis of multidimensional contingency tables will increase. Experience at the George Washington University with such a growing demand confirms this. The examination and interpretation of data from social phenomena, housing, psychology, education, environmental problems, health, safety, manpower, business, experimental testing of devices, military research and development, etc., are potential source areas.

Critics of methods for contingency table analysis have maintained that most of the procedures used, at least in the past, were only of a global chi-squared test nature. However, for a recent example of this see Patil (1974). Through the use of the principle of minimum discrimination information (m.d.i.) estimation, leading to exponential families or multiplicative models or log-linear models we shall show, using illustrative examples exhibiting different aspects, that:

(1) Estimates of the cell entries under various hypotheses or models can be obtained;

(2) The adequacy or fit of the model, or the null hypothesis, can be tested;

(3) Main effect and interaction parameters can be estimated;

(4) The structure of the table can be studied in detail in terms of the various interrelationships among the classificatory variables;

(5) The procedures can be applied to test hypotheses about particular parameters and linear combinations of parameters that are of special interest;

(6) The procedures provide indication of outlier cells. These may cause a model not to fit overall, yet fit the other cells excluding the outliers;

(7) Since the procedures and concepts are based on a general principle a unified treatment of multidimensional contingency tables is possible. Sequences of generalizations step by step to higher order dimensional contingency tables are not necessary as has been the case with other ad hoc procedures (see for example, Patil (1974), Sugiura and Otake (1974));

(8) The procedure provides estimates based on an observed or sample table, which satisfy certain external hypotheses as to underlying probability relations in the population table. These estimates also preserve the inherent properties of the observed data not affected by the hypothesis;

(9) In general, the m.d.i. estimates are best asymptotically normal (BAN) and in the many applications of fitting models to a table based on observed sets of marginal values the m.d.i. estimates in particular are maximum-likelihood estimates;

(10) The test statistics are minimum discrimination information (m.d.i.) statistics which are asymptotically distributed as chi-squared with appropriate degrees of freedom. In the case of fitting models to a table based on observed sets of marginal values the m.d.i. statistics are log-likelihood ratio statistics. The m.d.i. statistics are additive, as are the associated degrees of freedom, so that the total under an hypothesis can be analyzed into components each under a sub-hypothesis. The analysis is analogous to analysis of variance and regression analysis techniques, using a design matrix, a set of regression parameters, and explanatory variables.

(11) In models fitting estimates to an observed table based on sets of observed marginal values as explanatory variables, some estimates can be ›ssed explicitly as products of marginal values. However this is not generally true, and expected cell frequencies (functions of marginal values), can be computed by an iterative proportional fitting procedure (Ku et al. (1971)), and the use of a computer to perform the iterations becomes necessary. For the foregoing cases which we shall term internal, and problems involving tests of external hypotheses on underlying populations a number of iterative computer programs are available. They provide as output, design matrices, the observed cell entries and the cell estimates as well as their logarithms, parameter estimates, outlier values, m.d.i. statistics and their corresponding significance levels, and covariance matrices of parameter estimates, to assist in and simplify the numerical aspects of the inference. In this respect it is of interest to

cite the following quotation from a book review by D. J. Finney in Journal Royal Statistical Society, Series A (General) Vol. 136 (1973), Part 3, p. 461, "No mention is made of the extent to which computers have destroyed the need to assess statistical methods in terms of arithmetical simplicity: indeed the emphasis on avoiding lengthy, but easily programmed, iterative calculations is remarkable."

Classical problems in the historical development of the analysis of contingency tables concerned themselves primarily with such questions as the independence or conditional independence of the classificatory variables, or homogeneity or conditional homogeneity of the classificatory variables over time or space, for example, similar to such tests in multivariate analysis as independence, multiple correlation, partial correlation, canonical correlation, etc. Such classical problems turn out to be special cases of the techniques we shall discuss. (See for example Kullback et al. 1962a, 1962b.) These techniques result in analyses which are essentially regression type analyses. As such they enable us to determine the relationship of one or more "dependent" qualitative or categorical variables of interest on a set of "independent" classificatory variables, as well as the relative effects of changes in the "independent" variables on the "dependent variables." The object of the analyses is the study of the interaction between and among the classifications. The term interaction is used here in a general sense to cover both dependence and association (see for example, Bartlett (1935), Simpson (1951), Roy and Kastenbaum (1956), Ku et al. (1971)). It may be noted here that in a seminar on a study of the historical development of the concept of interaction in the analysis of multidimensional contingency tables, the following series of papers, among the many that could be selected, was found to be very instructive: Bartlett (1935), Lancaster (1951), Simpson (1951), Roy and Kastenbaum (1956), Darroch (1962), Lewis (1962), Plackett (1962, 1969), Birch (1963, 1964, 1965), Goodman (1963b, 1970, 1971), Good (1963), Kastenbaum (1965), Mantel (1966), Berkson (1968, 1972), Bhapkar and Koch (1968a, 1968b), Ku and Kullback (1968), Dempster (1971), Ku, Varner and Kullback (1971). It was pointed out by Darroch (1962), "That 'interaction' in contingency tables enjoys only a few of the fortuitously simple properties of interactions in the analysis of variance." (See Kullback, 1973.)

Following this general introduction we shall consider further aspects of contingency tables in greater expository detail. We then present an introduction to minimum discrimination information estimation, the log-linear representation, associated design matrices and parameters, without detailed mathematical proofs. This will enable the reader then to study the many illustrative examples that follow and present various aspects of the possible analyses. The mathematical statistical proofs etc. are to be found at the end of the presentation.

2. Contingency Tables

1. Description

There are two ways in which statistical data are collected. In one form, actual measurements are recorded for each individual in the sample; in the other, the individuals are classified as belonging to different categories. On many occasions classifications are used to reduce original data on direct measurements. A well-known example is that of "frequency-distributions". Data collected in the form of measurements may later be grouped and presented as a frequency distribution. An important advantage of grouping is that it results in a considerable reduction of data. On the other hand, it is not usually possible to convert grouped or classified data back into the original form.

A contingency table is a form of presentation of grouped data. In the simplest case, a group of N items may be classified into just two groups, according to, say, presence or absence of a certain characteristic. For a fixed (given) characteristic the different groups of classification are called categories. For example, a group of N individuals may be classified according to hair-color (characteristic), the categories being black, brown, blonde and "other". The categories may be qualitative as above, or may be quantitative, as for example in the classification by weight in pounds consisting of five categories: 40-80, 80-120, 120-160, 160-200, 200-240. When there is only one characteristic according to which data are classified we get a one-way-table. If there are two ways of classification, say according to Rows and Columns, the Row-classification having r categories and the Column-classification having c categories, the table is called a two-way table or a r x c table. The latter notation gives the number of categories in each classification. Carrying this notation further, a r x c x d table will have three characteristics of classification, the first having r categories, the second having c and the third d.

2. Examples:

Example 1: The following is a one-way table with one classification-characteristic (Geographic Area) and four categories. It gives the distribution of students by Geographic Area.

East	North	West	South	Total
4201	4552	2840	5130	16723

Example 2: Consider the distribution of 20 balls in six cells

Cell	1	2	3	4	5	6	Total
Occupancy	2	4	4	5	1	4	20

It may be recalled at this point that in many situations such a distribution of N balls in k cells is adequately described by the multinomial distribution. We may therefore expect that the multinomial distribution will have an important role to play in the analysis of contingency tables.

Example 3: The distribution of students by Geographic area (as in Ex. 1) and sex gives rise to the following 2 x 4 contingency table.

Sex	Geographic Area				Totals
	East	North	West	South	
Male	2201	2350	1400	3100	9051
Female	2000	2202	1440	2030	7672
Totals	4201	4552	2840	5130	16723

Note that this is called a 2 x 4 table since the Row-classification (sex) has 2 categories. If the geographic areas were written in rows and the sex were to correspond to columns we would get a 4 x 2 table. We will follow this convention throughout.

Observe that for a two-way table there are two sets of marginal totals. In the above table the totals on the right can be looked upon as a one-way table with sex as a characteristic and two categories, male and female. At the bottom of the above table, we see the one-way table of Ex. 1. This shows that any two-way table is associated with two one-way tables given by the marginal totals of each characteristic.

Example 4.: The data below are octane determinations on independent samples of gasoline obtained in two regions of the northeastern United States in the summer of 1953. (Brownlee, Statistical Theory and Methodology, J. Wiley, 1965, p. 306).

Region A:	84.0	83.5	84.0	85.0	83.1	83.5	81.7
	85.4	84.1	83.0	85.8	84.0	84.2	82.2
	83.6	84.9					
Region D:	80.2	82.9	84.6	84.2	82.8	83.0	82.9
	83.4	83.1	83.5	83.6	86.7	82.6	82.4
	83.4	82.7	82.9	83.7	81.5	81.9	81.7
	82.5						

The problem of interest was whether the variability in the octane numbers could be regarded as the same for the two regions. Since the number of sample-values for region A and D are small (16 and 22 respectively) the data can be conveniently analyzed in the given form. For the sake of illustration, suppose that we classify the octane readings into three categories; below 83.5 as "poor", between 83.5 and 84.5 as "normal" and above 84.5 as "better", we will get the following 2 x 3 table:

Region	Gasoline quality			Totals
	Poor	Normal	Better	
A	4	8	4	16
D	16	5	1	22
Totals	20	13	5	38

This illustrates how to prepare contingency tables from actual measurement-data. But the example brings out another important point. The contingency table, in fact, represents two frequency distributions, one from Region A and the other for Region D laid side by side. This table is different from the ones we came across earlier in that we did not start the classification with a total of 38 values, to be classified according to Region and Quality; rather we had a priori a set of 16 values for Region A and 22 values for Region D. (Further the sampling for the two regions was done independently). In other words, the set of marginal totals (on the one-way table) for Region was fixed before the experiments. Later on we will have ample opportunities to see the effect of such restrictions on the analyses. At present, it is enough to know that tables as above may be regarded as contingency tables with fixed (restricted) marginal totals.

3. Problems associated with contingency tables

In the analysis of contingency tables we are usually interested in the relationship between one classification and one or more of the other classifications. Thus in the example 4. on comparison of octane ratings we would like to compare the variability of the values for classifications given by Regions A and D. As another example, consider a three-way r x c x d contingency table in which the row-classification represents the response of an experiment on animals, the column classification types of treatment and the depth classification sex. The following hypotheses may be of interest.

1. Response is independent of treatment irrespective of sex.

2. Response is independent of the different combinations of treatment and sex (as against the possibility that a particular treatment is more "effective" in terms of the response, for a particular sex).

3. Given sex, response is independent of treatment.

We shall see in subsequent chapters how these hypotheses can be formulated mathematically. Of course, not all contingency tables can be interpreted in such a straightforward manner. In some instances, all three classifications can be considered as responses; then we may be interested in the independence or association among these responses. In other cases, a classification may be controlled, experimentally or naturally, like three specified levels of fertilizer applied or sex, in which case the classification is termed as a factor. For convenience, we shall group all the concepts of association, dependence, etc. under the general term of <u>interaction</u>. No interaction between treatment and sex appears to be a more acceptable phrase than independence between treatment and sex, since the term independence is usually reserved to express the relationship between random variables. We may also say that the interaction between response and treatment does not interact with sex, meaning the degree of association between response and treatment is the same for both sexes. This concept gives rise to the idea of second-order interaction. There are a number of different approaches to the mathematical formulation and interpretation of the concept of "no interaction". One such approach, through the concept of "generalized independence" is powerful and general enough to include all hypotheses of "no interaction" (formulated in a specific manner) <u>and</u> many other hypotheses about homogeneity, symmetry, etc. that we come across in analyzing contingency tables. Before this concept is introduced, we shall need the necessary symbolism and notation.

4. Notation and preliminaries:

We have seen that the entries in the "cells" of a contingency table are frequencies of occurence. We will denote these frequencies generically by the letter x, with or without subscripts. These frequencies are a result of classification of a fixed number of individuals according to a certain probability distribution. Hence the observed frequencies x can be looked upon as realizations of a random variable X.

The cell of a contingency table and the observed frequency in that cell are symbolically associated in the following manner. In the example 1, we have a one-way table representing the distribution of 16723 students by geographic area. We denote the occurrence in the table by x(i) with the notation

Characteristic	Index	1	2	3	4
Geographic area	i	East	North	West	South

Thus x(3), for example, equals 2840. The total 16723 of all x(i) for i=1,2,3 and 4, will be denoted by x(.).

That is, $\sum_{i=1}^{4} x(i) = x(.) = 16723$. For the two-way table of Ex 3, we denote the frequencies in the table by $x(ij)$ with the notation

Characteristic	Index	1	2	3	4
Sex	i	Male	Female		
Geographic area	j	East	North	West	South

Then $x(2,3) = 1440$, $x(1,4) = 3100$ and so on. To denote marginal totals we will use the dot notation as before. The row marginals are

$$\sum_{j=1}^{4} x(1j) = x(1.) = 9051, \sum_{j=1}^{4} x(2j) = x(2.) = 7672$$

The column marginals are

$$\sum_{i=1}^{2} x(i1) = x(.1) = 4201, \ldots.., \sum_{i=1}^{2} x(i4) = x(.4) = 5130$$

The grand total is denoted by $x(..)$ so that $x(1.) + x(2.) = x(..) = x(.1) + x(.2) + x(.3) + x(.4) = 16723 = N$

Now consider the following three-way table:

Propagation of plum root stocks from root-cuttings

Response (Mortality)	At once		Spring		Totals
	Long	Short	Long	Short	
Alive	156	107	84	31	378
Dead	84	133	156	209	582
Totals	240	240	240	240	960

The frequencies in the cells are denoted by x(ijk) with the notation

Characteristic	Index	1	2
Mortality	i	Alive	Dead
Time of planting	j	At once	Spring
Length of cutting	k	Long	Short

The marginals are as follows:

One-way marginals: $\sum_j \sum_k x(ijk) = x(i..),\ i=1,2$

$\sum_i \sum_k x(ijk) = x(.j.),\ j=1,2$

$\sum_i \sum_j x(ijk) = x(..k),\ k=1,2$

Two-way marginals: $\sum_i x(ijk) = x(.jk),\ j=1,2,\ k=1,2$

$\sum_j x(ijk) = x(i.k),\ i=1,2\quad k=1,2$

$\sum_k x(ijk) = x(ij.),\ i=1,2\quad j=1,2$

Note that $\sum_i x(ij.) = x(.j.)$, $\sum_j x(ij.) = x(i..)$, $\sum_i x(i..) = x(...)$ etc.

For the above table, x(1..) = 378, x(2..) = 582 and x(...) = 960. It should be observed that x(.jk) = 240 for all the four combinations of j and k. This restriction is imposed by the method of experimentation; for each combination of the planting time and cutting length exactly 240 root-stocks were used and their mortality observed. This is another case of fixed marginals, similar to the one encountered in Ex. 4.

The notation for cell frequencies and for marginal totals can be extended in an obvious manner to four-way, five-way and higher order tables.

Let us now recall that in a contingency table a number of individuals are classified into cells. In other words for a given cell, an individual is classified in the cell with a certain probability. In a four-way table, for example, each cell will be denoted by (i,j,k,ℓ) for some values of the indices i, j, k and ℓ. The probability that an individual will be classified in this cell will be denoted by $p(ijk\ell)$. Just as we defined the marginal totals for the cell frequencies $x(ijk\ell)$ we may define marginal totals for probabilities. For example,

$$p(i\ldots) = \sum_j \sum_k \sum_\ell p(ijk\ell)$$
$$p(.j.\ell) = \sum_i \sum_k p(ijk\ell)$$
etc.

For a two-way table the cell probabilities will be denoted by $p(ij)$, for a three-way table by $p(ijk)$ and so on. But we would like to develop the theory of all contingency tables in a unified manner. For this purpose it is necessary to use a symbol, ω, say, which will generically denote cells like (ij) in a two-way table, $(ijk\ell)$ in a four-way table and so on. For example, in a 2x3x5 table, the symbol $x(\omega)$ will replace $x(ijk)$, being one of the 2x3x5 = 30 cells. The symbol ω here corresponds to the triplet (ijk) and takes "values" (1,1,1), (1,1,2)...(1,1,5), (1,2,1).....(2,3,5).

Let us now go back to some problems associated with the analysis of contingency tables discussed in 3, and see how we can formulate them symbolically, with the help of the notation developed. We considered a rxcx2 table in which the row-classification represents response in an experiment on animals, the column classification represents types of treatment and the depth classification represents sex. The cell probabilities are p(ijk).

1. Response is independent of treatment irrespective of sex.

Since the sex of the animal is immaterial in the statement of the hypothesis, we consider marginal totals of probabilities of the form p(ij.). Now, since the response is postulated to be independent of treatment we further have

$$p(ij.) = p(i..)\ p(.j.) \quad i=1,\ldots,r,\ j=1,\ldots c.$$

2. Response is independent of the different combinations of treatment and sex.

The probability corresponding to a particular combination of treatment and sex is given by the (marginal) total p(.jk).The hypothesis is formulated, therefore, as

$$p(ijk) = p(i..)\ p(.jk) \quad \begin{array}{l} i=1...r \\ j=1...c \\ k=1,2 \end{array}$$

3. Given sex, response is independent of treatment. Let the conditional probability of being classified in the cell (ijk), given that the individual is classified in the k-th depth classification (sex), be denoted by $p(ij|k)$. Also, the marginal conditional probability of classification in the i-th category irrespective of the column classification is $p(i.k)/p(..k)$ and a similar marginal probability for the j-th category of the column classification, given k, is $p(.jk)/p(..k)$. The hypothesis then states that

$$p(ij|k) = \frac{p(i.k)p(.jk)}{p^2(..k)} \quad k=1,\ 2\ ,\ i=1...r,\ j=1...c.$$

But $p(ij|k) = p(ijk)/p(..k)$, so that the above relations can be restated as

$$p(ijk) = \frac{p(i.k)p(.jk)}{p(..k)} \quad k=1,2,\ j=1...r,\ j=1...c.$$

Observe that $\sum_i \sum_j p(ij|k) = 1$, since given that an individual fell into the k-th category, it must be classified in one of the (i,j) cells corresponding to the fixed k.

This imposes the restriction that

$$\sum_i \sum_j p(ij|k) = 1 = \sum_i \sum_j \frac{p(ijk)}{p(..k)}, \; k = 1,2$$

i.e.

$$\sum_i \sum_j p(ijk) = p(..k), k=1,2.$$

Note that the second hypothesis (of independence) led us to the formulation $p(ijk) = p(i..)\,p(.jk)$ and the third hypothesis (of conditional independence) led to $p(ijk) = p(i.k)p(.jk)/p(..k)$. The cell-probabilities in each case are expressed as products of marginal probabilities. From another point of view, we can say that the trivariate function $p(ijk)$ is expressed as a product of (simpler) univariate and bivariate functions, of the form $p(.jk)$ and $p(i..)$, for example. When the cell probabilities are thus expressible as products of functions of a smaller subset of arguments, we say that the probabilities obey generalized independence. By generalized independence is meant that the cell probability of a multidimensional contingency table may be expressed as the product of factors which are functions of various marginals (Ireland and Kullback, 1968; Ku and Kullback, 1968; Ku et al., 1971). The common notions of independence, conditional independence, homogeneity, or conditional homogeneity in contingency tables are all special cases of generalized independence. This is a consequence of the fact that in accordance with the minimum discrimination information theorem, the m.d.i. estimates are formulated as members of an exponential family, which may also be expressed as a multiplicative model or a logarithmic linear additive model (Kullback, 1959; Ireland and Kullback, 1968; Ku et al., 1971). Note that we do not assume such a model to start with, as others have, but derive this model by the principle of minimum discrimination information estimation (Birch, 1963; Bishop, 1967, 1969; Goodman, 1970; Mantel, 1966).

5. Estimates

We shall denote estimates of the cell entries under various hypotheses or models by $x_\alpha^*(\omega)$, where values of the subscript α will range over the hypotheses or models.

For two-way 2x2 tables the primary question of interest is whether the row and column variables are independent. An example of such a table is shown in Table 1.

Table 1.

	x(ij)		
	j = 1	j = 2	
i = 1	x(11)	x(12)	x(1·)
i = 2	x(21)	x(22)	x(2·)
	x(·1)	x(·2)	x(··) = n

To answer this question one estimates the cell entries under the hypothesis of independence as a product of the marginals, that is, denoting the estimate by $x^*(ij)$ one uses $x^*(ij) = x(i\cdot)x(\cdot j)/n$. Some appropriate measure of the deviation between $x(ij)$ and $x^*(ij)$ is then used to determine whether the differences are "larger" than one would reasonably expect under the hypothesis of independence.

The estimated two-way table under the hypothesis or model of independence is given in Table 2.

Table 2.

ESTIMATE UNDER INDEPENDENCE

$x^*(ij)$

	$j = 1$	$j = 2$	
$i = 1$	$x(1\cdot)x(\cdot 1)/n$	$x(1\cdot)x(\cdot 2)/n$	$x(1\cdot)$
$i = 2$	$x(2\cdot)x(\cdot 1)/n$	$x(2\cdot)x(\cdot 2)/n$	$x(2\cdot)$
	$x(\cdot 1)$	$x(\cdot 2)$	n

Note that the estimated table has the same marginals as the observed table $x(ij)$.

A common statistical measure of the association or interaction between the variables of a two-way 2x2 contingency table is the cross-product ratio, or its logarithm. The cross-product ratio is defined by

$$\frac{x(11)x(22)}{x(12)x(21)} ,$$

though we shall be more concerned with its logarithm

$$\ell n \frac{x(11)x(22)}{x(12)x(21)} .$$

We shall use natural logarithms, that is, logarithms to the base e , rather than common logarithms to the base 10, because of the nature of the underlying mathematical statistical theory. Note that with the estimate for independence, or no association, the logarithm of the cross-product ratio is zero.

$$\ell n \frac{x^*(11)x^*(22)}{x^*(12)x^*(21)} = \ell n \frac{\frac{x(1\cdot)x(\cdot 1)}{n} \frac{x(2\cdot)x(\cdot 2)}{n}}{\frac{x(1\cdot)x(\cdot 2)}{n} \frac{x(2\cdot)x(\cdot 1)}{n}} = \ell n\ 1 = 0 .$$

The logarithm of the cross-product ratio is positive if the odds satisfy the inequalities

$$\frac{x(11)}{x(21)} > \frac{x(12)}{x(22)} \quad \text{or} \quad \frac{x(11)}{x(12)} > \frac{x(21)}{x(22)} ,$$

since then we get for the log-odds

$$\ell n \frac{x(11)x(22)}{x(12)x(21)} = \ell n \frac{x(11)}{x(21)} - \ell n \frac{x(12)}{x(22)} > 0$$

$$= \ell n \frac{x(11)}{x(12)} - \ell n \frac{x(21)}{x(22)} > 0 .$$

The logarithm of the cross-product ratio is negative if the odds satisfy the inequalities

$$\frac{x(11)}{x(21)} < \frac{x(12)}{x(22)} \quad \text{or} \quad \frac{x(11)}{x(12)} < \frac{x(21)}{x(22)} ,$$

since then we get for the log-odds

$$\ell n \frac{x(11)x(22)}{x(12)x(21)} = \ell n \frac{x(11)}{x(21)} - \ell n \frac{x(12)}{x(22)} < 0$$

$$= \ell n \frac{x(11)}{x(12)} - \ell n \frac{x(21)}{x(22)} < 0 .$$

The logarithm of the cross-product ratio thus varies from $-\infty$ to $+\infty$. Later we shall consider procedures for assessing the significance of the deviation of the logarithm of the cross-product ratio from zero, the value corresponding to no association or no interaction.

Similar procedures apply to the case of a two-way rxc contingency table, that is, one with r rows and c columns.

TABLE 3a

TWO-WAY rxc CONTINGENCY TABLE

i \ j	1	2	...	c	
1	x(11)	x(12)	...	x(1c)	x(1·)
2	x(21)	x(22)	...	x(2c)	x(2·)
⋮	...	...	...	...	...
r	x(r1)	x(r2)	...	x(rc)	x(r·)
	x(·1)	x(·2)	...	x(·c)	n

Under a hypothesis or model of independence of row and column categories $x^*(ij) = x(i\cdot)x(\cdot j)/n$. Even if the row categories, say, are not randomly observed but selected with respect to some characteristic, say time or space, the mathematical procedures are still the same for determining whether the column categories are homogeneous over the row categories, time or space for instance. In the latter case we may consider the two-way table as a set of one-way tables. Terms which cover both the case of independence and homogeneity are "association" or "interaction," that is, we question whether there is association or interaction among the variables.

The estimated two-way rxc contingency table under the hypothesis or model of independence is given in Table 3b.

TABLE 3b

ESTIMATE UNDER INDEPENDENCE

$x^*(ij)$

i \ j	1	2	...	c	
1	x(1•)x(•1)/n	x(1•)x(•2)/n	...	x(1•)x(•c)/n	x(1•)
2	x(2•)x(•1)/n	x(2•)x(•2)/n	...	x(2•)x(•c)/n	x(2•)
⋮	...	...	...	...	...
r	x(r•)x(•1)/n	x(r•)x(•2)/n	...	x(r•)x(•c)/n	x(r•)
	x(•1)	x(•2)	...	x(•c)	n

Note that the estimated table has the same marginals as the observed Table 3a.

A three-way contingency table arises when each observation has three classifications with different possible numbers of categories for each classification. The simplest three-way contingency table is 2x2x2, that is, with two categories for each classification.

In the general notation we have Table 4.

TABLE 4

	i = 1		i = 2		
	j = 1	j = 2	j = 1	j = 2	
k = 1	x(111)	x(121)	x(211)	x(221)	x(••1)
k = 2	x(112)	x(122)	x(212)	x(222)	x(••2)
	x(11•)	x(12•)	x(21•)	x(22•)	n

The two-way marginals are

$$x(11\cdot) = x(111) + x(112),$$
$$x(12\cdot) = x(121) + x(122),$$
$$x(21\cdot) = x(211) + x(212),$$
$$x(22\cdot) = x(221) + x(222),$$
$$x(1\cdot 1) = x(111) + x(121),$$
$$x(1\cdot 2) = x(112) + x(122),$$
$$x(2\cdot 1) = x(211) + x(221),$$
$$x(2\cdot 2) = x(212) + x(222),$$

$$x(\cdot 11) = x(111) + x(211) ,$$
$$x(\cdot 12) = x(112) + x(212) ,$$
$$x(\cdot 21) = x(121) + x(221) ,$$
$$x(\cdot 22) = x(122) + x(222) .$$

The one-way marginals are

$$x(1\cdot\cdot) = x(111) + x(112) + x(121) + x(122) = x(11\cdot) + x(12\cdot) ,$$
$$x(2\cdot\cdot) = x(211) + x(212) + x(221) + x(222) = x(21\cdot) + x(22\cdot) ,$$
$$x(\cdot 1\cdot) = x(111) + x(112) + x(211) + x(212) = x(11\cdot) + x(21\cdot) ,$$
$$x(\cdot 2\cdot) = x(121) + x(122) + x(221) + x(222) = x(12\cdot) + x(22\cdot) ,$$
$$x(\cdot\cdot 1) = x(111) + x(121) + x(211) + x(221) = x(1\cdot 1) + x(2\cdot 1) ,$$
$$x(\cdot\cdot 2) = x(112) + x(122) + x(212) + x(222) = x(1\cdot 2) + x(2\cdot 2) .$$

The entries $x(ijk)$ in Table 4 may also be considered as three-way marginals.

With more variables there are more possible questions of interest. One may be interested in whether any pair of the variables are independent or show no interaction or association. One may be interested in conditional independence, that is, whether a pair of variables are independent given the third variable. One may be interested in whether the three variables are mutually independent or whether one of the variables is independent of the pair of the other variables. These questions of independence, no interaction or association are all answered by considering estimates which are explicitly represented in terms of products of various marginals. We list some of these estimates.

Mutual independence of i, j, and k $\quad x_1^*(ijk) = x(i\cdot\cdot)x(\cdot j\cdot)x(\cdot\cdot k)/n^2,$

Independence of i and (jk) jointly $\quad x_a^*(ijk) = x(i\cdot\cdot)x(\cdot jk)/n ,$

Conditional independence of i and j given k $\quad x_b^*(ijk) = x(i\cdot k)x(\cdot jk)/x(\cdot\cdot k).$

As might be expected, these estimates also apply in the general three-way rxsxt contingency table.

We note that the estimate under mutual independence of i, j, and k has the same one-way marginals as the observed table $x(ijk)$,

$$x_1^*(111) = x(1\cdot\cdot)x(\cdot1\cdot)x(\cdot\cdot1)/n^2 ,$$

$$x_1^*(112) = x(1\cdot\cdot)x(\cdot1\cdot)x(\cdot\cdot2)/n^2 ,$$

$$x_1^*(121) = x(1\cdot\cdot)x(\cdot2\cdot)x(\cdot\cdot1)/n^2 ,$$

$$x_1^*(122) = x(1\cdot\cdot)x(\cdot2\cdot)x(\cdot\cdot2)/n^2 ,$$

$$x_1^*(211) = x(2\cdot\cdot)x(\cdot1\cdot)x(\cdot\cdot1)/n^2 ,$$

$$x_1^*(212) = x(2\cdot\cdot)x(\cdot1\cdot)x(\cdot\cdot2)/n^2 ,$$

$$x_1^*(221) = x(2\cdot\cdot)x(\cdot2\cdot)x(\cdot\cdot1)/n^2 ,$$

$$x_1^*(222) = x(2\cdot\cdot)x(\cdot2\cdot)x(\cdot\cdot2)/n^2 ,$$

$$\begin{aligned} x_1^*(1\cdot\cdot) &= x_1^*(111) + x_1^*(112) + x_1^*(121) + x_1^*(122) \\ &= x(1\cdot\cdot)x(\cdot1\cdot)/n + x(1\cdot\cdot)x(\cdot2\cdot)/n \\ &= x(1\cdot\cdot) , \end{aligned}$$

$$\begin{aligned} x_1^*(2\cdot\cdot) &= x_1^*(211) + x_1^*(212) + x_1^*(221) + x_1^*(222) \\ &= x(2\cdot\cdot)x(\cdot1\cdot)/n + x(2\cdot\cdot)x(\cdot2\cdot)/n \\ &= x(2\cdot\cdot) , \end{aligned}$$

$$\begin{aligned} x_1^*(\cdot1\cdot) &= x_1^*(111) + x_1^*(112) + x_1^*(211) + x_1^*(212) \\ &= x(1\cdot\cdot)x(\cdot1\cdot)/n + x(2\cdot\cdot)x(\cdot1\cdot)/n \\ &= x(\cdot1\cdot) , \end{aligned}$$

$$\begin{aligned} x_1^*(\cdot2\cdot) &= x_1^*(121) + x_1^*(122) + x_1^*(221) + x_1^*(222) \\ &= x(\cdot2\cdot) , \end{aligned}$$

$$\begin{aligned} x_1^*(\cdot\cdot1) &= x_1^*(111) + x_1^*(121) + x_1^*(211) + x_1^*(221) \\ &= x(\cdot\cdot1) , \end{aligned}$$

$$\begin{aligned} x_1^*(\cdot\cdot2) &= x_1^*(112) + x_1^*(122) + x_1^*(212) + x_1^*(222) \\ &= x(\cdot\cdot2) . \end{aligned}$$

However, the two-way marginals of the estimate under mutual independence of i , j , and k differ from the two-way marginals of the observed table $x(ijk)$. Thus, for example ,

$$x_1^*(11\cdot) = x_1^*(111) + x_1^*(112)$$
$$= x(1\cdot\cdot)x(\cdot1\cdot)x(\cdot\cdot1)/n^2 + x(1\cdot\cdot)x(\cdot1\cdot)x(\cdot\cdot2)/n^2$$
$$= x(1\cdot\cdot)x(\cdot1\cdot)/n \ ,$$

and the latter value is not necessarily equal to $x(11\cdot)$.

The estimate under the hypothesis or model of independence of i and (jk) jointly has the same one-way marginals and the same two-way jk-marginal as the observed table $x(ijk)$,

$$x_a^*(111) = x(1\cdot\cdot)x(\cdot11)/n \ ,$$
$$x_a^*(112) = x(1\cdot\cdot)x(\cdot12)/n \ ,$$
$$x_a^*(121) = x(1\cdot\cdot)x(\cdot21)/n \ ,$$
$$x_a^*(122) = x(1\cdot\cdot)x(\cdot22)/n \ ,$$
$$x_a^*(211) = x(2\cdot\cdot)x(\cdot11)/n \ ,$$
$$x_a^*(212) = x(2\cdot\cdot)x(\cdot12)/n \ ,$$
$$x_a^*(221) = x(2\cdot\cdot)x(\cdot21)/n \ ,$$
$$x_a^*(222) = x(2\cdot\cdot)x(\cdot22)/n \ ,$$
$$x_a^*(1\cdot\cdot) = x_a^*(111) + x_a^*(112) + x_a^*(121) + x_a^*(122)$$
$$= x(1\cdot\cdot)x(\cdot11)/n + x(1\cdot\cdot)x(\cdot12)/n + x(1\cdot\cdot)x(\cdot21)/n + x(1\cdot\cdot)x(\cdot22)/n$$
$$= x(1\cdot\cdot)[x(\cdot11) + x(\cdot12) + x(\cdot21) + x(\cdot22)]/n$$
$$= x(1\cdot\cdot) \ .$$

Similar results follow for the other one-way marginals.

$$x_a^*(\cdot11) = x_a^*(111) + x_a^*(211)$$
$$= x(1\cdot\cdot)x(\cdot11)/n + x(2\cdot\cdot)x(\cdot11)/n$$
$$= x(\cdot11) \ ,$$
$$x_a^*(\cdot12) = x_a^*(112) + x_a^*(212)$$
$$= x(1\cdot\cdot)x(\cdot12)/n + x(2\cdot\cdot)x(\cdot12)/n$$
$$= x(\cdot12) \ ,$$

$$x_a^*(\cdot 21) = x_a^*(121) + x_a^*(221)$$
$$= x(1\cdot\cdot)x(\cdot 21)/n + x(2\cdot\cdot)x(\cdot 21)/n$$
$$= x(\cdot 21)\ ,$$
$$x_a^*(\cdot 22) = x_a^*(122) + x_a^*(222)$$
$$= x(1\cdot\cdot)x(\cdot 22)/n + x(2\cdot\cdot)x(\cdot 22)/n$$
$$= x(\cdot 22)\ .$$

However, for the other two-way marginals, for example,

$$x_a^*(11\cdot) = x_a^*(111) + x_a^*(112)$$
$$= x(1\cdot\cdot)x(\cdot 11)/n + x(1\cdot\cdot)x(\cdot 12)/n$$
$$= x(1\cdot\cdot)[x(\cdot 11) + x(\cdot 12)]/n$$
$$= x(1\cdot\cdot)x(\cdot 1\cdot)/n\ ,$$

and the latter value is not necessarily equal to $x(11\cdot)$.

$$x_a^*(1\cdot 1) = x_a^*(111) + x_a^*(121)$$
$$= x(1\cdot\cdot)x(\cdot 11)/n + x(1\cdot\cdot)x(\cdot 21)/n$$
$$= x(1\cdot\cdot)[x(\cdot 11) + x(\cdot 21)]/n$$
$$= x(1\cdot\cdot)x(\cdot\cdot 1)/n\ ,$$

and the latter value is not necessarily equal to $x(1\cdot 1)$.

The estimate under the hypothesis or model of conditional independence of i and j given k has the same one-way marginals and the same two-way ik- and jk-marginals as the observed table $x(ijk)$,

$$x_b^*(111) = x(1\cdot 1)x(\cdot 11)/x(\cdot\cdot 1)\ ,$$
$$x_b^*(112) = x(1\cdot 2)x(\cdot 12)/x(\cdot\cdot 2)\ ,$$
$$x_b^*(121) = x(1\cdot 1)x(\cdot 21)/x(\cdot\cdot 1)\ ,$$
$$x_b^*(122) = x(1\cdot 2)x(\cdot 22)/x(\cdot\cdot 2)\ ,$$
$$x_b^*(211) = x(2\cdot 1)x(\cdot 11)/x(\cdot\cdot 1)\ ,$$

$$x_b^*(212) = x(2\cdot 2)x(\cdot 12)/x(\cdot\cdot 2)\ ,$$

$$x_b^*(221) = x(2\cdot 1)x(\cdot 21)/x(\cdot\cdot 1)\ ,$$

$$x_b^*(222) = x(2\cdot 2)x(\cdot 22)/x(\cdot\cdot 2)\ ,$$

$$\begin{aligned} x_b^*(1\cdot\cdot) &= x_b^*(111) + x_b^*(112) + x_b^*(121) + x_b^*(122) \\ &= x(1\cdot 1)x(\cdot 11)/x(\cdot\cdot 1) + x(1\cdot 2)x(\cdot 12)/x(\cdot\cdot 2) \\ &\quad + x(1\cdot 1)x(\cdot 21)/x(\cdot\cdot 1) + x(1\cdot 2)x(\cdot 22)/x(\cdot\cdot 2) \\ &= x(1\cdot 1) + x(1\cdot 2) = x(1\cdot\cdot)\ . \end{aligned}$$

Similar results follow for the other one-way marginals.

$$\begin{aligned} x_b^*(1\cdot 1) &= x_b^*(111) + x_b^*(121) \\ &= x(1\cdot 1)x(\cdot 11)/x(\cdot\cdot 1) + x(1\cdot 1)x(\cdot 21)/x(\cdot\cdot 1) \\ &= x(1\cdot 1)\ , \\ x_b^*(1\cdot 2) &= x_b^*(112) + x_b^*(122) \\ &= x(1\cdot 2)x(\cdot 12)/x(\cdot\cdot 2) + x(1\cdot 2)x(\cdot 22)/x(\cdot\cdot 2) \\ &= x(1\cdot 2)\ , \end{aligned}$$

and in a similar manner we have

$$\begin{aligned} x_b^*(2\cdot 1) &= x(2\cdot 1)\ ,\ x_b^*(2\cdot 2) = x(2\cdot 2)\ , \\ x_b^*(\cdot 11) &= x_b^*(111) + x_b^*(211) \\ &= x(1\cdot 1)x(\cdot 11)/x(\cdot\cdot 1) + x(2\cdot 1)x(\cdot 11)/x(\cdot\cdot 1) \\ &= x(\cdot 11)\ , \\ x_b^*(\cdot 12) &= x_b^*(112) + x_b^*(212) \\ &= x(1\cdot 2)x(\cdot 12)/x(\cdot\cdot 2) + x(2\cdot 2)x(\cdot 12)/x(\cdot\cdot 2) \\ &= x(\cdot 12)\ , \end{aligned}$$

and in a similar manner we have

$$x_b^*(\cdot 21) = x(\cdot 21)\ ,\ x_b^*(\cdot 22) = x(\cdot 22)\ .$$

However, for the other two-way marginals

$$x_b^*(11\cdot) = x_b^*(111) + x_b^*(112)$$
$$= x(1\cdot1)x(\cdot11)/x(\cdot\cdot1) + x(1\cdot2)x(\cdot12)/x(\cdot\cdot2) \ ,$$

and the latter value is not necessarily equal to $x(11\cdot)$.

We remark that one of the constraints in the determination of the estimates was that they have certain marginals the same as the observed table.

For the three-way 2x2x2 contingency table in addition to the classic types of independence, interaction or association, there arises an additional one, important historically and practically. This is known as no three-factor

or no second-order interaction. No three-factor or no second-order interaction implies that the logarithm of the association measured by the cross-product ratio for any two of the variables is the same for all the values of the third variable, that is, there is no second-order interaction if

$$(1) \qquad \begin{cases} \ln \dfrac{x(111)x(221)}{x(121)x(211)} = \ln \dfrac{x(112)x(222)}{x(122)x(212)} \ , & i, j \ , \\[2ex] \ln \dfrac{x(111)x(212)}{x(112)x(211)} = \ln \dfrac{x(121)x(222)}{x(122)x(221)} \ , & i, k \ , \\[2ex] \ln \dfrac{x(111)x(122)}{x(112)x(121)} = \ln \dfrac{x(211)x(222)}{x(212)x(221)} \ , & j, k \ . \end{cases}$$

One is concerned with the possible hypothesis or model of no second-order interaction when none of the other types of independence are found. However, in this case, the corresponding estimate cannot be expressed explicitly in terms of observed marginals although the estimate is constrained to have the same two-way marginals as the observed table. Straightforward iterative procedures exist to determine the estimate under the hypothesis or model of no second-order interaction. For the general three-way [illegible] contingency table there are of course many more relations among the log cross-product ratios like (1) which must be satisfied, but the iterative procedures to determine the estimate extend to the general case with no difficulty.

We may be concerned with a set of two-way tables for which it is of interest to determine whether they are homogeneous with respect to a third factor, say space or time. Such problems may also be treated as three-way contingency tables using the space or time factor as the third classification (Kullback, 1959).

For four-way and higher order contingency tables the problem of presentation of the data increases, as do the variety and number of questions about relationships of possible interest and varieties of interaction. The basic ideas, concepts, notation and terminology we have discussed for the two- and three-way contingency tables extend to the more general cases as we consider the methodology (Ku et al., 1971).

3. Log-linear Representation

1. Minimum Discrimination Information Estimation

To make the presentation more specific, and with no essential restriction on the generality, we discuss it in terms of the analysis of four-way contingency tables. Let us consider the collection of four-way contingency tables RxSxTxU of dimension rxsxtxu. For convenience let us denote the aggregate of all cell identifications, as well as their number, by Ω with individual cells identified by ω, so that the generic variable is $\omega = (i,j,k,\ell)$, $i=1,\ldots,r$, $j=1,\ldots,s$, $k=1,\ldots,t$, $\ell=1,\ldots,u$. In this case we also identify Ω as rstu. Suppose there are two probability distributions or contingency tables (we shall use these terms interchangeably) defined over the aggregate or space Ω, say $p(\omega)$, $\pi(\omega)$, $\sum_{\Omega} p(\omega) = 1$, $\sum_{\Omega} \pi(\omega) = 1$. The discrimination information is defined by

$$I(p:\pi) = \sum_{\Omega} p(\omega) \ln \frac{p(\omega)}{\pi(\omega)} .$$

For the various applications we shall consider the π-distribution, $\pi(\omega)$, according to the problem of interest, may either be specified, may be an estimated distribution, or may be an observed distribution. The p-distribution, $p(\omega)$, ranges over or is a member of a family $\mathcal{P}$ of distributions of interest satisfying certain restraints.

Of the various properties of $I(p:\pi)$ we mention in particular the fact that $I(p:\pi) > 0$ and $= 0$ if, and only if, $p(\omega) = \pi(\omega)$ (Kullback, 1959).

Many problems in the analysis of contingency tables may be characterized as estimating a distribution or contingency table subject to certain restraints and then comparing the estimated table with an observed table to determine whether the observed table satisfies a null hypothesis or model implied by the restraints. In accordance with the principle of minimum discrimination information estimation, we determine that member of the collection or family P of distributions, which minimizes the discrimination information $I(p:\pi)$. We denote the minimum discrimination information estimate by $p^*(\omega)$ so that

$$I(p^*:\pi) = \Sigma\, p^*(\omega)\ \ell n \frac{p^*(\omega)}{\pi(\omega)} = \min I(p:\pi),\ p,\ p^* \in P.$$

Unless otherwise stated, the summation is over Ω which will be omitted.

It may be shown that if $p(\omega)$ is any member of the family P of distributions, then

(1) $$I(p:\pi) = I(p^*:\pi) + I(p:p^*).$$

The pythagorean type property (1) plays an important role in the analysis of information tables.

In a wide class of problems which can be characterized as "smoothing", or fitting a model to an observed contingency table the restraints specify that the estimated distribution or contingency table have some set of marginals, or more generally, linear functions of observed cell entries, equal to those values for the observed contingency table. In such cases $\pi(\omega)$ is taken to be either the uniform distribution $\pi(ijk\ell) = 1/rstu$, or a distribution already estimated subject to restraints contained in and implied by the restraints under examination. The latter case includes the classical hypotheses of independence, conditional independence, homogeneity, conditional homogeneity and interaction, all of which can be considered as instances of generalized independence.

To test whether an observed contingency table is consistent with the null hypothesis, or model, as represented by the minimum discrimination information estimate, we compute a measure of the deviation between the observed distribution and the appropriate estimate by the minimum discrimination information statistic. For notational and computational convenience, let us denote the estimated contingency table in terms of occurrences by $x^*(\omega) = np^*(\omega)$ where n is the total number of occurrences. For the "smoothing" or fitting class of problems, that is, with the restraints implied by a set of observed marginals (those of a generalized independence hypothesis), or more generally, linear functions of observed all entries, the minimum discrimination information (m.d.i.) statistic is

$$2I(x:x^*) = 2\Sigma\, x(\omega)\, \ln \frac{x(\omega)}{x^*(\omega)}, \tag{2}$$

which is asymptotically distributed as a chi-squared variate with appropriate degrees of freedom under the null hypothesis.

The statistic in (2) is also minus twice the logarithm of the classic likelihood ratio statistic but this is not necessarily true for other kinds of applications of the general theory (Berkson, 1972).

2. Computational Procedures

An "experiment" has been designed and observations made resulting in a multidimensional contingency table with the desired classifications and categories. All the information the analyst hopes to obtain from the "experiment" is contained in the contingency table. In the process of analysis, the aim is to fit the observed table with a minimal or parsimonious number of parameters depending on some of the observed marginals, and/or some general linear combinations of observed cell entries, that is, essentially, to find out how much of this total information is contained in a summary consisting of sets of marginals, and/or some linear combinations of observed cell entries.

Indeed, the relationship between the concept of independence or association and interaction in contingency tables and the role the marginals play is evidenced in the historical developments in the extensive literature on the analysis of contingency tables.

Let us denote by $\underset{\sim}{x}$ the Ωx1 matrix of entries $x(\omega)$ of the observed contingency table arranged in lexicographic order, and denote by $\underset{\sim}{T}$ an Ωx(m+1) design matrix of rank $m+1 \leq \Omega$. We denote the columns of $\underset{\sim}{T}$ by $T_i(\omega)$, $1 \leq \omega \leq \Omega$, $0 \leq i \leq m$. The condition that the estimate $x^*(\omega)$ have some set of marginals, and/or some general linear combination of cell entries, equal to the corresponding values of the observed contingency table is written in matrix notation as

$$(3) \qquad \underset{\sim}{T}'\underset{\sim}{x}^* = \underset{\sim}{T}'\underset{\sim}{x} .$$

Those columns of $\underset{\sim}{T}$ which imply a marginal restraint are the indicator functions of the marginals, that is, the corresponding $T_i(\omega)$ will be one or zero for any cell ω, according as the cell ω does or does not, enter into the marginal in question. We usually take $T_0(\omega) = 1$, for all ω, so that $\Sigma x^*(\omega) = \Sigma\, x(\omega) = n$. In accordance with the minimum discrimination information theorem (Kullback, 1959), the m.d.i. estimate is the exponential family

$$(4) \qquad x^*(\omega) = \exp(\tau_0 T_0(\omega)+\tau_1 T_1(\omega)+\ldots+\tau_m T_m(\omega))n\pi(\omega).$$

If we denote the Ωx1 matrix whose entries are $\ln(x^*(\omega)/n\pi(\omega))$, in lexicographic order on ω by $\underset{\sim}{\ln}\,(x^*/n\pi)$, then we have from (4) the log-linear regression (Gokhale, 1971, 1972; Ku et al. 1974)

$$(5) \qquad \underset{\sim}{\ln}(x^*/n\pi) = \underset{\sim}{T}\,\underset{\sim}{\tau}\,,$$

where $\underset{\sim}{\tau}$ is the (m+1)x1 matrix of the parameters $\tau_0,\tau_1,\tau_2,\ldots,\tau_m$. We set the normalizing parameter τ_0=L and $\tau_1,\ldots,\tau_m$ are main effects and interactions. The parameters in (4) are to be determined so that $x^*(\omega)$ satisfies the condition (3). There are convergent iterative computer algorithms of proportional fitting (among others), which yield the estimate $x^*(\omega)$ satisfying (3), and then the parameters are determined from (5). The iteration may be described as successively cycling through adjustments of the marginals of interest starting with the $\pi(\omega)$ distribution until a desired accuracy of agreement between the set of observed marginals of interest and the computed marginals has been attained. See Ku et al. (1971). Note that although $n\pi(\omega)$ is here a constant and could be absorbed into τ_0 or L, we prefer to express it explicitly because there are cases in which $n\pi(\omega)$ is not a constant and the expression in (4) or (5) still applies (Ireland and Kullback, 1968a, b; Gokhale, 1971; Darroch and Ratcliff, 1972).

3. Analysis of Information

The analysis of information is based on the fundamental relation (1) for the minimum discrimination information statistics. Specifically if $np_a^*(\omega) = x_a^*(\omega)$ is the minimum discrimination information estimate corresponding to a set H_a of given marginals, and $x_b^*(\omega)$ is the minimum discrimination information estimate corresponding to a set H_b of given marginals, where H_a is explicitly or implicitly contained in H_b, then the basic relations are

$$(6)\quad \begin{cases} 2I(x:n\pi) = 2I(x_a^*:n\pi) + 2I(x:x_a^*) \\ 2I(x:n\pi) = 2I(x_b^*:n\pi) + 2I(x:x_b^*) \\ 2I(x_b^*:n\pi) = 2I(x_a^*:n\pi) + 2I(x_b^*:x_a^*) \\ 2I(x:x_a^*) = 2I(x_b^*:x_a^*) + 2I(x:x_b^*) \end{cases}$$

with a corresponding additive relation for the associated degrees of freedom.

In terms of the representation in (4) or (5), as an exponential family, the two extreme cases are the uniform distribution for which all τ's except L are zero, and the observed contingency table or distribution, the complete model, for which all Ω-1 = rstu - 1 τ's in addition to L are needed.

Measures of the form $2I(x:x_a^*)$, that is, the comparison of an observed contingency table with an estimated contingency table, are called measures of interaction or goodness-of-fit. Measures of the form $2I(x_b^*:x_a^*)$, comparing two estimated contingency tables, are called measures of effect, that is the effect of the marginals in the set H_b but not in the set H_a, or the taus in x_b^* but not in x_a^*. We note that $2I(x:x_a^*)$ tests a null hypothesis that the values of the τ parameters in the representation of the observed contingency table $x(\omega)$ but not in the representation of the estimated table $x_a^*(\omega)$ are zero and the number of these taus is the number of degrees of freedom. Similarly $2I(x_b^*:x_a^*)$ tests a null hypothesis that the values of the set of τ parameters in the representation of the estimated table $x_b^*(\omega)$ but not in the representation of the estimated table $x_a^*(\omega)$ are zero, and the number of these taus is the number of degrees of freedom. See section 5, The 2x2x2x2 Table.

We summarize the additive relationships of the m.d.i. statistics and the associated degrees of freedom in the Analysis of Information Table 1.

TABLE 1

ANALYSIS OF INFORMATION TABLE

Component due to		Information	D.F.
H_a:	Interaction	$2I(x:x_a^*)$	N_a
H_b:	Effect	$2I(x_b^*:x_a^*)$	$N_a - N_b$
	Interaction	$2I(x:x_b^*)$	N_b

Since measures of the form $2I(x:x_a^*)$ may also be interpreted as measures of the "variation unexplained" by the estimate x_a^*, the additive relationship leads to the interpretation of the ratio

$$(7) \qquad \frac{2I(x:x_a^*) - 2I(x:x_b^*)}{2I(x:x_a^*)} = \frac{2I(x_b^*:x_a^*)}{2I(x:x_a^*)} ,$$

as the percentage of the unexplained variation due to x_a^* accounted for by the additional constraints defining x_b^*. The ratio (7) is thus similar to the squared correlation coefficients associated with normal distributions (Goodman, 1970).

We remark that the marginals, explicit and implicit, of the estimated table $x_a^*(\omega)$, which form the set of restraints H_a used to generate $x_a^*(\omega)$ are the same as the corresponding marginals of the observed $x(\omega)$ table and all lower order implied marginals. It may be shown that $2I(x:x_a^*)$ is approximately a quadratic in the differences between the remaining marginals of the $x(\omega)$ table and the corresponding ones as calculated from $x_a^*(\omega)$.

Similarly, $2I(x_b^*:x_a^*)$ is also approximately a quadratic in the differences between those additional marginal restraints in H_b but not in H_a and the corresponding marginal values as computed from the $x_a^*(\omega)$ table.

The τ's are determined from the log-linear regression equations (5) as sums and differences of values of $\ell n\ x^*(\omega)$ or as linear combinations thereof. A variety of statistics have been presented in the literature for the analysis of contingency tables, which are quadratics in differences of marginal values or quadratics in the τ's or the linear combinations of logarithms of the observed or estimated values. The principle of minimum discrimination information estimation and its procedures thus provides a unifying relationship since such statistics may be seen as quadratic approximations of the minimum discrimination information statistic. We remark that the corresponding approximate X^2's are not generally additive (Berkson, 1972).

We mention the approximations in terms of quadratic forms in the marginals, or the τ's, as a possible bridge to relate the familiar procedures of classical regression analysis and the procedures proposed here. This may assist in understanding and interpreting the analysis of information tables (Kullback, 1959). The covariance matrix of the $T(\omega)$ functions or the taus can be obtained for either the observed table or any of the estimated tables, as well as the inverse matrices, as part of the output of the general computer program.

4. The 2x2 Table

It may be useful to reexamine the 2x2 table from the point of view of the preceding discussion. The algebraic details are simple in this case and exhibit the unification of the information theoretic development.

Suppose we have the observed 2x2 table in Figure 1

x(11)	x(12)	x(1.)
x(21)	x(22)	x(2.)
x(.1)	x(.2)	n

Figure 1

If we obtain the m.d.i. estimate fitting the one-way marginals, the generalized independence hypothesis is the classical independence hypothesis and the minimum discrimination information estimate is the usual $x^*(ij) = x(i.)x(.j)/n$. By the iterative scaling fitting procedure, we begin with $x^{(0)}(ij) = n/4$ in each cell and adjust the $x^{(0)}(ij)$ values by the ratios of the observed row marginals to those of $x^{(0)}(ij)$, that is,

$$x^{(1)}(ij) = x^{(0)}(ij)\,\frac{x(i.)}{n/2} = x(i.)2\ .$$

Then we adjust $x^{(1)}(ij)$ by the ratio of observed column marginals to the marginals of $x^{(1)}(ij)$,

$$x^{(2)}(ij) = x^{(1)}(ij)\,\frac{x(.j)}{n/2} = \frac{x(i.)}{2}\cdot\frac{x(.j)}{n/2}$$

$$= x(i.)x(.j)/n = x^*(ij).$$

Since the row and column marginals of $x^*(ij)$ are now the same as the observed values, no further iterative adjustment is necessary. For fitting a 2x2 table to externally specified marginals see Ireland and Kullback, 1968b or Fisher's 2x2 table in the examples.

The representation of the log-linear regression for the complete model is given in Figure 2. The entries in the columns τ_1, τ_2, τ_3

i	j	L	τ_1	τ_2	τ_3
1	1	1	1	1	1
1	2	1	1		
2	1	1		1	
2	2	1			

Figure 2

are, respectively, the values of the functions $T_1(ij)$, $T_2(ij)$, $T_3(ij)$ associated with the marginals $x(1.)$, $x(.1)$, $x(11)$, and the column headed L corresponds to the normalising factor.

We note the interpretation of Figure 2 as the lcg-linear relations

$$
(8) \qquad \begin{cases}
\ln \dfrac{x(11)}{n\pi} = L + \tau_1 + \tau_2 + \tau_3 \\[2ex]
\ln \dfrac{x(12)}{n\pi} = L + \tau_1 \\[2ex]
\ln \dfrac{x(21)}{n\pi} = L + \tau_2 \\[2ex]
\ln \dfrac{x(22)}{n\pi} = L\ .
\end{cases}
$$

From (8) we find

$$
\begin{cases}
L = \ln\ (x(22)/n/4), \\[1ex]
\tau_1 = \ln\ (x(12)/x(22)), \\[1ex]
\tau_2 = \ln\ (x(21)/x(22)), \\[1ex]
\tau_3 = \ln\ (x(11)x(22)/x(12)x(21),
\end{cases}
$$

or

$$
(9)\qquad \begin{cases} \tau_1 = \ln x(12) - \ln x(22), \\ \tau_2 = \ln x(21) - \ln x(22), \\ \tau_3 = \ln x(11) + \ln x(22) - \ln x(12) - \ln x(21). \end{cases}
$$

The design matrix $\underline{T}$ is the matrix of Figure 2, that is,

$$
\underline{T} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} .
$$

Define the diagonal matrix $\underline{D}$ with main diagonal the elements $x(ij)$, in lexicographic order, that is,

$$
\underline{D} = \begin{pmatrix} x(11) & 0 & 0 & 0 \\ 0 & x(12) & 0 & 0 \\ 0 & 0 & x(21) & 0 \\ 0 & 0 & 0 & x(22) \end{pmatrix},
$$

then the estimate of the covariance matrix of $x(1.)$, $x(.1)$, $x(11)$, for the observed contingency table is $\underline{S}_{22.1}$, where

$$
\underline{S} = \begin{pmatrix} \underline{S}_{11} & \underline{S}_{12} \\ \underline{S}_{21} & \underline{S}_{22} \end{pmatrix} = \underline{T}'\underline{D}\underline{T} ,
$$

$$
\underline{S}_{22.1} = \underline{S}_{22} - \underline{S}_{21}\, \underline{S}_{11}^{-1}\, \underline{S}_{12} ,
$$

and $\underline{S}_{11}$ is 1×1, $\underline{S}_{22}$ is 3×3, $\underline{S}_{21}' = \underline{S}_{12}$ is 1×3. It is found that

$$S_{22.1} = \begin{pmatrix} \frac{x(1.)x(2.)}{n} & x(11) - \frac{x(1.)x(.1)}{n} & \frac{x(11)x(2.)}{n} \\ x(11) - \frac{x(1.)x(.1)}{n} & \frac{x(.1)x(.2)}{n} & \frac{x(11)x(.2)}{n} \\ \frac{x(11)x(2.)}{n} & \frac{x(11)x(.2)}{n} & x(11) - \frac{x^2(11)}{n} \end{pmatrix} .$$

and the inverse matrix is

$$S_{22.1}^{-1} = \begin{pmatrix} \frac{1}{x(12)} + \frac{1}{x(22)} & \frac{1}{x(22)} & -\frac{1}{x(12)} - \frac{1}{x(22)} \\ \frac{1}{x(22)} & \frac{1}{x(21)} + \frac{1}{x(22)} & -\frac{1}{x(21)} - \frac{1}{x(22)} \\ -\frac{1}{x(12)} - \frac{1}{x(22)} & -\frac{1}{x(21)} - \frac{1}{x(22)} & \frac{1}{x(11)} + \frac{1}{x(12)} + \frac{1}{x(21)} + \frac{1}{x(22)} \end{pmatrix} .$$

The matrix $S_{22.1}^{-1}$ is the covariance matrix of the τ's in (9). Similar results hold in general and for estimated tables (Kullback, 1959).

Note that the value of the logarithm of the cross-product ratio, a measure of association or interaction, appears in the course of the analysis as the value of τ_3 for the observed values x(ij). For $x^*(ij)$, the estimate under the hypothesis of independence, the representation as in Figure 2 does not involve the last column, since $x^*(ij)$ is obtained by fitting the one-way marginals, and $\tau_3=0$.

The log-linear relations for the estimate $x^*(ij)$ are

$$(10) \quad \begin{cases} \ln \frac{x^*(11)}{n\pi} = L + \tau_1 + \tau_2 \\ \ln \frac{x^*(12)}{n\pi} = L + \tau_1 \\ \ln \frac{x^*(21)}{n\pi} = L + \tau_2 \\ \ln \frac{x^*(22)}{n\pi} = L \ , \end{cases}$$

where the numerical values of L, τ_1, τ_2 in (10) must of course depend on x^* and differ from the values in (8).

The minimum discrimination information statistic to test the null hypothesis or model of independence is $2I(x:x^*)$ with one degree of freedom. In this case the quadratic approximation is

$$(11) \qquad 2I(x:x^*) \quad \left(x(11) - \frac{x(1.)x(.1)}{n}\right)^2 \left(\frac{1}{x^*(11)} + \frac{1}{x^*(12)} + \frac{1}{x^*(21)} + \frac{1}{x^*(22)}\right).$$

Remembering that $x^*(ij) = x(i.)x(.j)/n$, the right-hand side of (11) may also be shown to be

$$(12) \qquad X^2 = \Sigma\ (x(ij) - x(i.)x(.j)/n)^2 / \frac{x(i.)x(.j)}{n},$$

the classical X^2-test for independence with one degree of freedom. Another test which has been proposed for the null hypothesis of no association or no interaction in the 2x2 table is

$$(\ln x(11) + \ln x(22) - \ln x(12) - \ln x(21))^2 \left(\frac{1}{x(11)} + \frac{1}{x(12)} + \frac{1}{x(21)} + \frac{1}{x(22)}\right)^{-1},$$

which may be shown to be a quadratic approximation for $2I(x:x^*)$ in terms of τ_3 with the covariance matrix estimated using the observed values and not the estimated values. We remark that if the observed values are used to estimate the covariance matrix then instead of the classical X^2-test in (12) there is derived the Neyman modified chi-square

$$X_1^2 = \Sigma\ (x(ij) - x(i.)x(.j)/n)^2/x(ij).$$

5. The 2x2x2x2 Table

A useful graphic representation of the log-linear regression (5) is given in Figure 3 for a 2x2x2x2 contingency table. This is the analogue of the design matrix in normal regression theory. The blank spaces in Figure 3 represent zero values. The ($ijk\ell$)-columns are the cell identifications in the same lexicographic order as the cell entries for the estimates in the computer output. Column 1 corresponds to L which is the normalising factor. Each of the columns 2 to 16 represents the corresponding values of the $T(\omega)$ functions, columns 2 to 5 those for the one-way marginals, columns 6 to 11 those for the two-way marginals, columns 12 to 15 those for the three-way marginals, and column 16 that for the four-way marginal. The tau parameter associated with the $T(\omega)$ function is given at the head of the column. The superscripts are useful identifications. The complete representation with all the columns of Figure 3 generates the observed values. Thus the rows represent

$$\ell n \frac{x(ijk\ell)}{n\pi(ijk\ell)} = L + \tau_1^i T_1^i(ijk\ell) + \ldots + \tau_{11}^{ij} T_{11}^{ij}(ijk\ell)$$

$$+ \ldots + \tau_{111}^{ijk} T_{111}^{ijk}(ijk\ell) + \ldots + \tau_{1111}^{ijk\ell} T_{1111}^{ijk\ell}(ijk\ell) ,$$

where $\pi(ijk\ell)$ in the 2x2x2x2 case is 1/2x2x2x2 and the numerical values of L and the taus depend on the observed values $x(ijk\ell)$. The design matrix corresponding to an estimate uses only those columns associated with the marginals explicit and implied in the fitting process. This is a reflection of the fact that higher order marginals imply certain lower order marginals, for example, the two-way marginal $x(ij..)$ implies, by summation over i and j, the one-way marginals $x(.j..)$, $x(i...)$, and the

ω				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
i	j	k	ℓ	L	τ_1^i	τ_1^j	τ_1^k	τ_1^ℓ	τ_{11}^{ij}	τ_{11}^{ik}	$\tau_{11}^{i\ell}$	τ_{11}^{jk}	$\tau_{11}^{j\ell}$	$\tau_{11}^{k\ell}$	τ_{111}^{ijk}	$\tau_{111}^{ij\ell}$	$\tau_{111}^{ik\ell}$	$\tau_{111}^{jk\ell}$	$\tau_{1111}^{ijk\ell}$
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	2	1	1	1	1		1	1		1			1				
1	1	2	1	1	1	1		1	1		1		1			1			
1	1	2	2	1	1	1			1										
1	2	1	1	1	1		1	1		1	1			1			1		
1	2	1	2	1	1		1			1									
1	2	2	1	1	1			1			1								
1	2	2	2	1	1														
2	1	1	1	1		1	1	1				1	1	1				1	
2	1	1	2	1		1	1					1							
2	1	2	1	1		1		1					1						
2	1	2	2	1		1													
2	2	1	1	1			1	1						1					
2	2	1	2	1			1												
2	2	2	1	1				1											
2	2	2	2	1															

Figure 3. Graphic representation

total n=x(....). The representation for the uniform distribution corresponds to column 1 only. The estimate $x_1^*(ijk\ell)$ based on fitting the one-way marginals will use only columns 1-5. The values of L and the taus for this estimate will be different from those for $x(ijk\ell)$ and depend on the estimate $x_1^*(ijk\ell)$. The representation in Figure 3 implies for $x_1^*(ijk\ell)$

$$\begin{cases} \ell n \dfrac{x_1^*(1111)}{n\pi} = L + \tau_1^i + \tau_1^j + \tau_1^k + \tau_1^\ell \\ \ell n \dfrac{x_1^*(1112)}{n\pi} = L + \tau_1^i + \tau_1^j + \tau_1^k \\ \quad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ \ell n \dfrac{x_1^*(2222)}{n\pi} = L \ . \end{cases}$$

The estimate $x_2^*(ijk\ell)$ based on fitting the two-way marginals will use columns 1-11 since the two-way marginals also imply the one-way marginals. The values of L and the taus for this estimate will be different from those for the observed values or other estimates and depend on the values of the estimate $x_2^*(ijk\ell)$. For the estimate fitting the two-way marginals the representation in Figure 3 implies

$$\begin{cases} \ell n \dfrac{x_2^*(1111)}{n\pi} = L + \tau_1^i + \tau_1^j + \tau_1^k + \tau_1^\ell + \tau_{11}^{ij} + \tau_{11}^{ik} + \tau_{11}^{i\ell} + \tau_{11}^{jk} + \tau_{11}^{j\ell} + \tau_{11}^{k\ell} \\ \ell n \dfrac{x_2^*(1112)}{n\pi} = L + \tau_1^i + \tau_1^j + \tau_1^k + \tau_{11}^{ij} + \tau_{11}^{ik} + \tau_{11}^{jk} \\ \quad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ \ell n \dfrac{x_2^*(2222)}{n\pi} = L \ . \end{cases}$$

The estimate $x_3^*(ijk\ell)$ based on fitting the three-way marginals will use columns 1-15 since the three-way marginals also imply the two-way and one-way marginals.

Note that in the graphic representation in figure 3 we set all taus with subscript i=2 and/or j=2 and/or k=2 and/or ℓ=2 equal to zero, by convention, to insure linear independence.

The analysis of information table corresponding to the hierarchical fitting of $x_1^*(ijk\ell)$, $x_2^*(ijk\ell)$, $x_3^*(ijk\ell)$ is shown in table 2.

TABLE 2

ANALYSIS OF INFORMATION

Component due to	Information	D.F.
All one-way marginals	$2I(x:x_1^*)$	11
All two-way marginals	$2I(x_2^*:x_1^*)$	6
	$2I(x:x_2^*)$	5
All three-way marginals	$2I(x_3^*:x_2^*)$	4
	$2I(x:x_3^*)$	1

$2I(x:x_1^*)$ tests the null hypothesis that the eleven taus of columns 6-16 are equal to zero.

$2I(x_2^*:x_1^*)$ tests the null hypothesis that the six taus of columns 6-11 are equal to zero.

$2I(x:x_2^*)$ tests the null hypothesis that the five taus of columns 12-16 are zero.

$2I(x_3^*:x_2^*)$ tests the null hypothesis that the four taus of columns 12-15 are zero.

$2I(x:x_3^*)$ tests the null hypothesis that the tau of column 16 is zero.

In the examples we shall see other tests on the interaction parameters (Kullback, 1974). We now consider a number of examples to illustrate more specifically various aspects of the analysis.

6. Algorithms to calculate quadratic approximations.

We now present algorithms to calculate quadratic approximations to $2I(x:x_a^*)$, $2I(x_b^*:x_a^*)$, $2I(x^*:x)$.

1. $2I(x:x_a^*)$.

   a) Compute x_a^*.

   b) Using the T design matrix corresponding to x(including the L column), compute the matrix $\underline{S} = \underline{T}'\underline{D}_a^*\underline{T}$, where $\underline{D}_a^*$ is a diagonal matrix whose entries are the values of x_a^* in the same order as for the rows of the T-matrix.

   c) Let $\underline{S} = \begin{pmatrix} \underline{S}_{11} & \underline{S}_{12} \\ \underline{S}_{21} & \underline{S}_{22} \end{pmatrix}$, where $\underline{S}_{11}$ is a 1x1 matrix,

   then $\underline{S}_{22.1} = \underline{S}_{22} - \underline{S}_{21} \underline{S}_{11}^{-1} \underline{S}_{12}$.

   d) Compute $\underline{S}_{22.1}^{-1}$.

   e) Consider the marginals which <u>do</u> <u>not</u> enter into the specification of x_a^*, and let $\underline{d}'$ be a one row matrix whose entries are the differences between the set of marginals just considered, in the x and x_a^* tables.

f) Let $\underline{B}$ be that submatrix of $\underline{S}_{22.1}^{-1}$ whose rows and columns correspond to the τ columns of the design matrix associated with the set of marginals in step e).

g) Compute $\underline{d}'\underline{Bd}$.

This is the "marginals" approximation to $2I(x:x_a^*)$.

h) Compute the set of τ's associated with the marginals considered in e) for the x distribution, and call the one row matrix of these τ's $\underline{\tau}'$.

Compute $\underline{\tau}'\underline{B}^{-1}\underline{\tau}$, where $\underline{B}^{-1}$ is the inverse of the matrix $\underline{B}$ in f).

$\underline{\tau}'\underline{B}^{-1}\underline{\tau}$ is the "tau" approximation to $2I(x:x_a^*)$.

i) The "marginals" approximation is also equal to

$$\sum \frac{(x - x_a^*)^2}{x_a^*}$$

2. $2I(x_b^* : x_a^*)$

a) Compute x_b^*, x_a^*.

b) Using the T design matrix corresponding to x_b^* (including the L column), compute the matrix $\underline{S} = \underline{T}'\underline{D}_a^*\underline{T}$, where $\underline{D}_a^*$ is a diagonal matrix whose entries are the values of x_a^* in the same order as for the rows of the T-matrix.

c) Let $\underline{S} = \begin{pmatrix} \underline{S}_{11} & \underline{S}_{12} \\ \underline{S}_{21} & \underline{S}_{22} \end{pmatrix}$, where $\underline{S}_{11}$ is a 1x1 matrix, then $\underline{S}_{22.1} = \underline{S}_{22} - \underline{S}_{21} \; \underline{S}_{11}^{-1} \; \underline{S}_{12}$.

d) Compute $\underline{S}_{22.1}^{-1}$.

e) Consider the marginals which enter into the specification of x_b^* but <u>not</u> in x_a^*, and let $\underline{d}'$ be a one row matrix whose entries are the differences between the set of marginals just considered in the x_b^* and x_a^* tables.

f) Let $\underline{B}$ be that submatrix of $\underline{S}_{22.1}^{-1}$ whose rows and columns correspond to the τ columns of the design matrix associated with the set of marginals in step e).

g) Compute $\underline{d}'\underline{Bd}$

This is the "marginals" approximation to $2I(x_b^*:x_a^*)$.

h) Compute the set of τ's associated with the marginals considered in e) for the x_b^* distribution and call the one row matrix of these τ's $\underline{\tau}'$.

Compute $\underline{\tau}'\underline{B}^{-1}\underline{\tau}$ where $\underline{B}^{-1}$ is the inverse of the matrix $\underline{B}$ in f).

$\underline{\tau}'\underline{B}^{-1}\underline{\tau}$ is the "tau" approximation to $2I(x_b^*:x_a^*)$.

i) The "marginals" approximation is also equal to

$$\sum \frac{(x_b^* - x_a^*)^2}{x_a^*} .$$

3. $2I(x^*:x)$.

a) Using the T design matrix corresponding to x* (including the L column), compute the matrix $\underline{S} = \underline{T}'\underline{D}_x\underline{T}$, where $\underline{D}_x$ is a diagonal matrix whose entries are the values of x in the same order as for the rows of the T-matrix.

b) Let $\underline{S} = \begin{pmatrix} \underline{S}_{11} & \underline{S}_{12} \\ \underline{S}_{21} & \underline{S}_{22} \end{pmatrix}$, where $\underline{S}_{11}$ is a 1x1 matrix, then

$$\underline{S}_{22.1} = \underline{S}_{22} - \underline{S}_{21} \quad \underline{S}_{11}^{-1} \quad \underline{S}_{12}.$$

c) Compute $\underline{S}_{22.1}^{-1}$.

d) Let $\underline{d}'$ be a one row matrix whose entries are the differences between the $\sum_{\omega} T(\omega)x^*(\omega)$ and $\sum_{\omega} T(\omega)x(\omega)$. In the case when $x^*(\omega)$ is specified by conditions external to the observed values, the value of $\sum_{\omega} T(\omega)x^*(\omega)$ is specified without having to compute $x^*(\omega)$.

e) Compute $\underline{d}'\underline{S}_{22.1}^{-1}\underline{d}$.

This is the approximation to $2I(x^*:x)$. Note that this can be obtained without computing x^*.

f) The approximation

$$\sum \frac{(x^*-x)^2}{x}$$

requires the prior computation of x^*.

4. Applications

In this chapter we consider eight examples illustrating various aspects of the model fitting methodology by the analysis of real data.

Example 1. Classification of multivariate dichotomous populations.

This example illustrates the analysis of a five-way 2x2x2x2x2 contingency table. It introduces the use of log-odds or logit representation, and the multiplicative version of the odds as a product of factors. It also illustrates the interpretation of the parameters, and the effect of interaction on the numerical value of the association between classifications. It considers several models with respect to the marginals fitted, the design matrices, and the detailed hierarchical analysis of information.

An Example of Multiway Contingency Table Analysis Applied to the Classification of Multivariate Dichotomous Populations

Introduction

Multiway contingency tables, or cross-classifications of vectors of discrete random variables, provide a useful approach to the analysis of multivariate discrete data. In the particular application we consider, the individual variates are dichotomous or binary. Note however that the procedures and analysis are not restricted to dichotomous or binary data but are also applicable to polychotomous variates.

For background on the study and problem leading to the data we consider see Solomon (1960). In Ku et al. (1969) minimum discrimination information procedures were applied to problems of multivariate binary data in information systems, such as communication, pattern recognition, and learning systems. In Cox (1972) there is a review of methods and models for the analysis of multivariate binary data and Solomon's data is given as a typical example. Martin and Bradley (1972) developed a model based on a set of orthogonal polynomials and applied it to Solomon's data. We remark that our procedure based on the principle of minimum discrimination information estimation applied to the analysis of multiway contingency tables yields a result practically equivalent to that of Martin and Bradley (1972). Goodman (1973) discusses Solomon's data in relation to methods for selecting models for contingency tables.

Solomon's Data

A total of 2982 high-school seniors were given an attitude questionnaire to assess their attitude towards science. The students were also classified on the basis of an IQ test into high IQ, the upper half, and low IQ, the lower half. The sixteen possible response vectors to each of four agree-disagree responses were tabulated. The problem of interest was to determine whether the response vectors could be used as a basis for classifying the students into one of two classes and evaluate possible classification procedures.

Contingency Table Analysis

We shall treat the data given in Table 1 as a five way 2x2x2x2x2 contingency table, denoting the original observations by $x(hijk\ell)$, where

Characteristic	Index	1	2
IQ	h	low IQ	high IQ
Response 1	i	disagree	agree
Response 2	j	disagree	agree
Response 3	k	disagree	agree
Response 4	ℓ	disagree	agree

As a first overview of the data to determine the marginals and their related interaction parameters which may furnish significant values in the log-linear representation of the exponential family of the estimates obtained by iterative scaling fitting, we list in Table 2a, Analysis of Information, a sequential hierarchical study of interaction and effect type measures Kullback (1970), Ku et al. (1971).

The first estimate we start with is

$$x_a^*(hijk\ell) = x(h\cdots\cdot)x(\cdot ijk\ell)/n$$

since the minimum discrimination information statistic (interaction type measure)

$$2I(x:x_a^*) = 2\Sigma\Sigma\Sigma\Sigma\Sigma\ x(hijk\ell)\ell n\ \frac{x(hijk\ell)n}{x(h\cdots\cdot)x(\cdot ijk\ell)}$$

tests a null hypothesis that the IQ groupings are homogeneous over the sixteen response vectors Kullback (1959, Chap. 8). This null hypothesis is rejected and the subsequent study of effect and interaction type measures is an attempt to find a good fit to the data and account for the total variation as measured by $2I(x:x_a^*)$. Although the association between IQ and the response to the first statement as measured by $2I(x_b^*:x_a^*) = 2.376$, 1 D.F., is not significant, it was decided to examine in detail the estimate $x_e^*(hijk\ell)$ whose numerical values are given in Table 1. It may be shown that

$$2I(x_b^*:x_a^*) = 2\Sigma\Sigma\ x(hi\cdots)\ell n\ \frac{x(hi\cdots)n}{x(h\cdots\cdot)x(\cdot i\cdots)}\ ,$$

and tests a null hypothesis that IQ is homogeneous over the response to the first question. The estimate $x_e^*(hijk\ell)$ was selected because the interaction type measure, $2I(x:x_e^*) = 16.307$, 11 D.F., represents an acceptable fit, the estimate is symmetric with respect to the four statements, and is comparable to the first-order model estimate of Martin and Bradley (1972), whose values are also listed in Table 1.

From the design matrix or log-linear representation in Fig. 1, we obtain the parametric representation for the log-odds (low IQ/high IQ)

$$\ell n(x_e^*(1ijk\ell)/x_e^*(2ijk\ell))$$

over the sixteen response vectors as given in Table 3a. Thus, for example

$$\ell n \frac{x_e^*(11111)}{x_e^*(21111)} = \tau_1^h + \tau_{11}^{hi} + \tau_{11}^{hj} + \tau_{11}^{hk} + \tau_{11}^{h\ell} ,$$

that is, a linear regression of the log-odds in terms of a constant τ_1^h and the main effects of each component of the response vector, namely, τ_{11}^{hi}, τ_{11}^{hj}, τ_{11}^{hk}, $\tau_{11}^{h\ell}$. The numerical values of the log-odds and the parameters are easily obtained from the entries in the computer output and are also given in Table 3a. It is clear that the odds may be expressed in a multiplicative model. The odds and the odds factors are easier to appreciate. From the log-odds representation above we find

$$\frac{x_e^*(11111)}{x_e^*(21111)} = \exp(\tau_1^h)\ \exp(\tau_{11}^{hi})\ \exp(\tau_{11}^{hj})\ \exp(\tau_{11}^{hk})\ \exp(\tau_{11}^{h\ell})$$

and from the values in Table 3a have

$$1.237 = (.682)(.816)(1.132)(1.406)(1.396).$$

We note from Table 3a that

$$\ell n \frac{x_e^*(1ijk1)}{x_e^*(2ijk1)} - \ell n \frac{x_e^*(1ijk2)}{x_e^*(2ijk2)} = \tau_{11}^{h\ell} = 0.3338 ,$$

that is, a change from disagree to agree on the fourth statement is associated with an increase of 0.3338 in the log-odds (low IQ/high IQ). Note also that $\tau_{11}^{h\ell}$ represents the association between IQ and response to the fourth statement as measured by the log-cross-product - ratio (log relative odds)

$$\tau_{11}^{h\ell} = \ell n \frac{x_e^*(1ijk1)\,x_e^*(2ijk2)}{x_e^*(2ijk1)\,x_e^*(1ijk2)} ,$$

and is the same for all eight levels of the responses to statements one, two and three.

Similarly, it is found that

$$\ell n \frac{x_e^*(1ij1\ell)}{x_e^*(2ij1\ell)} - \ell n \frac{x_e^*(1ij2\ell)}{x_e^*(2ij2\ell)} = \tau_{11}^{hk} = 0.3411 ,$$

$$\ell n \frac{x_e^*(1i1k\ell)}{x_e^*(2i1k\ell)} - \ell n \frac{x_e^*(1i2k\ell)}{x_e^*(2i2k\ell)} = \tau_{11}^{hj} = 0.1240 ,$$

$$\ell n \frac{x_e^*(11jk\ell)}{x_e^*(21jk\ell)} - \ell n \frac{x_e^*(12jk\ell)}{x_e^*(22jk\ell)} = \tau_{11}^{hi} = -0.2030 .$$

Classification

Since $x(1\cdot\cdot\cdot\cdot) = x_e^*(1\cdot\cdot\cdot\cdot) = 1491$, and $x(2\cdot\cdot\cdot\cdot) = x_e^*(2\cdot\cdot\cdot\cdot) = 1491$, we assign a response vector $(ijk\ell)$ to the region

E_1: classify as population h=1 (low IQ), when

$$\ln \frac{x_e^*(1ijk\ell)}{x_e^*(2ijk\ell)} \geq 0$$

and to the complementary region

E_2: classify as population h=2 (high IQ), when

$$\ln \frac{x_e^*(1ijk\ell)}{x_e^*(2ijk\ell)} < 0 .$$

If we set

$$\mu_1(E_1) = \sum_{(ijk\ell)\in E_1} \frac{x_e^*(1ijk\ell)}{1491}, \quad \mu_2(E_1) = \sum_{(ijk\ell)\in E_1} \frac{x_e^*(2ijk\ell)}{1491},$$

then the probability of error of the classification procedure is (Kullback, 1959, pp. 4, 69, 80),

$$\text{Prob Error} = p\mu_2(E_1) + q\mu_1(E_2) = (\mu_2(E_1) + \mu_1(E_2))/2$$

since here $p = x(2\cdot\cdot\cdot\cdot)/2982 = \frac{1}{2}$, $q = x(1\cdot\cdot\cdot\cdot)/2982 = \frac{1}{2}$.

The relevant computations with $x_e^*(hijk\ell)$ are given in Table 4(b) and show that the Prob. Error = 0.444. The corresponding computations with the original data $x(hikj\ell)$ are given in Table 4(a) and yield Prob. Error = 0.441.

Other Estimates

In view of the measure of the effect of the marginal $x(hi\cdot\cdot\ell)$ (and the associated interaction parameters) in Table 2a, $2I(x_m^*:x_g^*) = 4.316$, 1 D.F., and the marginal $x(h\cdot j\cdot\ell)$, $2I(x_p^*:x_n^*) = 3.181$, 1 D.F., the m.d.i. estimate $x_v^*(hijk\ell)$ fitting the marginals $x(\cdot ijk\ell)$, $x(h\cdot j\cdot\cdot)$, $x(h\cdot\cdot k\cdot)$, $x(hi\cdot\cdot\ell)$ and the m.d.i. estimate $x_w^*(hijk\ell)$ fitting the marginals $x(\cdot ijk\ell)$, $x(h\cdot\cdot k\cdot)$, $x(hi\cdot\cdot\ell)$, $x(h\cdot j\cdot\ell)$ were computed. The estimates are given in Table 1 and the relevant analysis of information given in Table 2b.

The values of the log-odds, parametric representation, and the associated interaction parameters are given in Table 3b for $x_v^*(hijk\ell)$ and in Table 3c for $x_w^*(hijk\ell)$. Note from Table 3b that

$$\ln\frac{x_v^*(11jk1)}{x_v^*(21jk1)} - \ln\frac{x_v^*(11jk2)}{x_v^*(21jk2)} = \tau_{11}^{h\ell} + \tau_{111}^{hi\ell} = 0.6469\ ,$$

$$\ln\frac{x_v^*(12jk1)}{x_v^*(22jk1)} - \ln\frac{x_v^*(12jk2)}{x_v^*(22jk2)} = \tau_{11}^{h\ell} = 0.2680\ ,$$

$$\ln\frac{x_v^*(11jk1)}{x_v^*(21jk1)} - \ln\frac{x_v^*(12jk1)}{x_v^*(22jk1)} = \tau_{11}^{hi} + \tau_{111}^{hi\ell} = -0.0276\ ,$$

$$\ln\frac{x_v^*(11jk2)}{x_v^*(21jk2)} - \ln\frac{x_v^*(12jk2)}{x_v^*(22jk2)} = \tau_{11}^{hi} = -0.4065\ ,$$

reflecting the interaction of the responses to the first and fourth statements.

From Table 3c, it is found for example, that

$$\ell n \frac{x_w^*(111k1)}{x_w^*(211k1)} - \ell n \frac{x_w^*(111k2)}{x_w^*(211k2)} = \tau_{11}^{h\ell} + \tau_{111}^{hi\ell} + \tau_{111}^{hj\ell} = 0.5806 ,$$

$$\ell n \frac{x_w^*(121k1)}{x_w^*(221k1)} - \ell n \frac{x_w^*(121k2)}{x_w^*(221k2)} = \tau_{11}^{h\ell} + \tau_{111}^{hj\ell} = 0.2030 ,$$

$$\ell n \frac{x_w^*(112k1)}{x_w^*(212k1)} - \ell n \frac{x_w^*(112k2)}{x_w^*(212k2)} = \tau_{11}^{h\ell} + \tau_{111}^{hi\ell} = 0.9371 ,$$

$$\ell n \frac{x_w^*(122k1)}{x_w^*(222k1)} - \ell n \frac{x_w^*(122k2)}{x_w^*(222k2)} = \tau_{11}^{h\ell} = 0.5595 ,$$

reflecting the interactions of the responses to the first, second and fourth statements.

The computation of the probability of error using the estimates $x_v^*(hijk\ell)$ and $x_w^*(hijk\ell)$ is shown in Table 4c and 4d respectively, and yields probabilities of error 0.444 and 0.446.

Remark

Martin and Bradley (1972) examined Solomon's data in terms of an estimate they called a first-order or linear model. These estimated values are given in Table 1. It turns out that although the underlying approaches are different, the Martin and Bradley parameters, their a_i, and estimates are practically the same as those for $x_e^*(hijk\ )$. From Martin and Bradley (1972, pp. 216-217) we note that

$$\ell n \frac{x_e^*(12222)}{x_e^*(22222)} = \tau_1^h = \ell n \frac{1+a_0+a_1+a_2+a_3+a_4}{1-a_0-a_1-a_2-a_3-a_4} ,$$

$$\ell n \frac{x_e^*(12221)}{x_e^*(22221)} = \tau_1^h + \tau_{11}^{h\ell} = \ell n \frac{1+a_0+a_1+a_2+a_3-a_4}{1-a_0-a_1-a_2-a_3+a_4} ,$$

$$\ell n \frac{x_e^*(12212)}{x_e^*(22212)} = \tau_1^h + \tau_{11}^{hk} = \ell n \frac{1+a_0+a_1+a_2-a_3+a_4}{1-a_0-a_1-a_2+a_3-a_4} ,$$

$$\ell n \frac{x_e^*(12122)}{x_e^*(22122)} = \tau_1^h + \tau_{11}^{hj} = \ell n \frac{1+a_0+a_1-a_2+a_3+a_4}{1-a_0-a_1+a_2-a_3-a_4} ,$$

$$\ell n \frac{x_e^*(11222)}{x_e^*(21222)} = \tau_1^h + \tau_{11}^{hi} = \ell n \frac{1+a_0-a_1+a_2+a_3+a_4}{1-a_0+a_1-a_2-a_3-a_4} ,$$

or to a first approximation of the logarithm

$$\tau_1^h = 2a_0+2a_1+2a_2+2a_3+2a_4 ,$$

$$\tau_1^h + \tau_{11}^{h\ell} = 2a_0+2a_1+2a_2+2a_3-2a_4 ,$$

$$\tau_1^h + \tau_{11}^{hk} = 2a_0+2a_1+2a_2-2a_3+2a_4 ,$$

$$\tau_1^h + \tau_{11}^{hj} = 2a_0+2a_1-2a_2+2a_3+2a_4 ,$$

$$\tau_1^h + \tau_{11}^{hi} = 2a_0-2a_1+2a_2+2a_3+2a_4 .$$

It is found that

$$\tau_{11}^{h\ell} = -4a_4 ,$$

$$\tau_{11}^{hk} = -4a_3 ,$$

$$\tau_{11}^{hj} = -4a_2 ,$$

$$\tau_{11}^{hi} = -4a_1 .$$

The values of the parameters given by Martin and Bradley (1972, Table 3, p. 217) are

$$a_0 = -0.042,\quad a_1 = 0.049,\quad a_2 = -0.031,\quad a_3 = -0.084,\quad a_4 = -0.082$$

so that

$$\tau_{11}^{h\ell} = 0.3338 = 0.334,\ -4a_4 = 0.328\ ,$$

$$\tau_{11}^{hk} = 0.3411 = 0.341,\ -4a_3 = 0.336\ ,$$

$$\tau_{11}^{hj} = 0.1240 = 0.124,\ -4a_2 = 0.124\ ,$$

$$\tau_{11}^{hi} = -0.2030 = -0.203,\ -4a_1 = -0.196\ .$$

The computation for the probability of error using the estimates are shown in Table 4e and yields a probability of error 0.445. (Martin and Bradley give a value of the risk as 0.455).

Solomon's Data-Classification Procedures

ij kℓ	Observed Low IQ $x(11jk\ell)$	Martin & Bradley	Estimates $x^*_e(11jk\ell)$	$x^*_v(11jk\ell)$	$x^*_w(11jk\ell)$	Observed High IQ $x(21jk\ell)$	Martin & Bradley	Estimates $x^*_e(21jk\ell)$	$x^*_v(21jk\ell)$	$x^*_w(21jk\ell)$
22 22	62	74.56	74.589	76.097	70.156	122	109.45	109.414	107.904	113.244
22 21	70	67.30	67.296	66.198	71.600	68	70.71	70.703	71.802	66.400
22 12	31	31.32	31.329	31.943	29.827	33	32.68	32.671	32.057	34.173
22 11	41	37.74	37.780	37.337	39.884	25	28.26	28.219	28.662	26.115
21 22	283	266.76	266.570	271.120	275.979	329	345.24	345.429	340.879	336.020
21 21	253	259.17	259.322	254.876	250.769	247	240.83	240.679	245.125	249.232
21 12	200	193.45	193.625	196.841	200.037	172	178.55	178.376	175.160	171.963
21 11	305	314.50	314.491	310.589	306.748	217	207.50	207.508	211.411	215.252
12 22	14	12.10	12.156	10.866	9.914	20	21.90	21.844	23.135	24.085
12 21	11	9.20	9.182	9.929	10.760	10	11.80	11.818	11.071	10.240
12 12	11	9.68	9.659	8.776	8.102	11	12.32	12.341	13.224	13.898
12 11	14	12.02	12.010	12.855	12.756	9	10.98	10.990	10.144	9.244
11 22	31	33.63	33.623	30.125	30.820	56	53.37	53.375	56.874	56.179
11 21	46	47.37	47.263	50.789	50.001	55	53.63	53.737	50.211	50.999
11 12	37	47.54	47.450	43.233	44.163	64	53.46	53.550	57.767	56.837
11 11	82	74.67	74.656	79.426	78.482	53	60.33	60.346	55.574	56.517
	1491					1491				

Table 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
h i j k l	L	h 1	i 1	j 1	k 1	l 1	hi 11	hj 11	hk 11	hl 11	ij 11	ik 11	il 11	jk 11	jl 11	kl 11	hij 111	hik 111	hil 111	hjk 111	hjl 111	hkl 111	ijk 111	ijl 111	ikl 111	jkl 111	hijk	hijl	hikl	hjkl	ijkl	hijkl
1 1 1 1 1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1 1 1 1 2	1	1	1	1	1		1	1	1		1	1		1			1	1		1			1				1					
1 1 1 2 1	1	1	1	1		1	1	1		1	1		1		1		1		1		1			1				1				
1 1 1 2 2	1	1	1	1			1	1			1						1															
1 1 2 1 1	1	1	1		1	1	1		1	1		1	1			1		1	1			1			1				1			
1 1 2 1 2	1	1	1		1		1		1			1						1														
1 1 2 2 1	1	1	1			1	1			1			1						1													
1 1 2 2 2	1	1	1				1																									
1 2 1 1 1	1	1		1	1	1		1	1	1				1	1	1				1	1	1				1				1		
1 2 1 1 2	1	1		1	1			1	1					1						1												
1 2 1 2 1	1	1		1		1		1		1					1						1											
1 2 1 2 2	1	1		1				1																								
1 2 2 1 1	1	1			1	1			1	1						1						1										
1 2 2 1 2	1	1			1				1																							
1 2 2 2 1	1	1				1				1																						
1 2 2 2 2	1	1																														
2 1 1 1 1	1		1	1	1	1					1	1	1	1	1	1							1	1	1	1					1	
2 1 1 1 2	1		1	1	1						1	1		1									1									
2 1 1 2 1	1		1	1		1					1		1		1									1								
2 1 1 2 2	1		1	1							1																					
2 1 2 1 1	1		1		1	1						1	1			1									1							
2 1 2 1 2	1		1		1							1																				
2 1 2 2 1	1		1			1							1																			
2 1 2 2 2	1		1																													
2 2 1 1 1	1			1	1	1								1	1	1										1						
2 2 1 1 2	1			1	1									1																		
2 2 1 2 1	1			1		1									1																	
2 2 1 2 2	1			1																												
2 2 2 1 1	1				1	1										1																
2 2 2 1 2	1				1																											
2 2 2 2 1	1					1																										
2 2 2 2 2	1																															
	x	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	x^*_e	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓							✓	✓	✓	✓					✓	
	x^*_v	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓				✓	✓	✓	✓					✓	
	x^*_w	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓		✓		✓	✓	✓	✓					✓	

Figure 1

Table 2a

Analysis of Information

Marginals Fitted	Information	D.F.
a) x(.ijkℓ),x(h....)	$2I(x:x^*_a) = 68.369$	15
b) x(.ijkℓ),x(hi...)	$2I(x^*_b:x^*_a) = 2.376$	1
	$2I(x:x^*_b) = 65.993$	14
c) x(.ijkℓ),x(hi...),x(h.j..)	$2I(x^*_c:x^*_b) = 4.265$	1
	$2I(x:x^*_c) = 61.728$	13
d) x(.ijkℓ),x(hi...),x(h.j..),x(h..k.)	$2I(x^*_d:x^*_c) = 25.230$	1
	$2I(x:x^*_d) = 36.498$	12
e) x(.ijkℓ),x(hi...),x(h.j..),x(h..k.),x(h...ℓ)	$2I(x^*_e:x^*_d) = 20.191$	1
	$2I(x:x^*_e) = 16.307$	11
f) x(.ijkℓ),x(h..k.),x(h...ℓ),x(hij..)	$2I(x^*_f:x^*_e) = 3.016$	1
	$2I(x:x^*_f) = 13.291$	10
g) x(.ijkℓ),x(h...ℓ),x(hij..),x(hi.k.)	$2I(x^*_g:x^*_f) = 0.042$	1
	$2I(x:x^*_g) = 13.249$	9
m) x(.ijkℓ),x(hij..),x(hi.k.),x(hi..ℓ)	$2I(x^*_m:x^*_g) = 4.316$	1
	$2I(x:x^*_m) = 8.933$	8
n) x(.ijkℓ),x(hij..),x(hi.k.),x(hi..ℓ),x(h.jk.)	$2I(x^*_n:x^*_m) = 0.983$	1
	$2I(x:x^*_n) = 7.950$	7
p) x(.ijkℓ),x(hij..),x(hi.k.),x(hi..ℓ),x(h.jk.),x(h.j.ℓ)	$2I(x^*_p:x^*_n) = 3.181$	1
	$2I(x:x^*_p) = 4.769$	6
q) x(.ijklℓ),x(hij..),x(hi.k.),x(hi..ℓ),x(h.jk.),x(h.j.ℓ), x(h..kℓ)	$2I(x^*_q:x^*_p) = 0.219$	1
	$2I(x:x^*_q) = 4.550$	5
r) x(.ijkℓ),x(hi..ℓ),x(h.j.ℓ),x(h..kℓ),x(hijk.)	$2I(x^*_r:x^*_q) = 0.346$	1
	$2I(x:x^*_r) = 4.204$	4

Analysis of Information (continued)

Marginals Fitted	Information	D.F.
	$2I(x:x^*_r) = 4.204$	4
s) $x(\cdot ijk\ell), x(h\cdot\cdot k\ell), x(hijk\cdot), x(hij\cdot\ell)$	$2I(x^*_s:x^*_r) = 2.303$	1
	$2I(x:x^*_s) = 1.901$	3
t) $x(\cdot ijk\ell), x(hijk\cdot), x(hij\cdot\ell), x(hi\cdot k\ell)$	$2I(x^*_t:x^*_s) = 1.375$	1
	$2I(x:x^*_t) = 0.526$	2
u) $x(\cdot ijk\ell), x(hijk\cdot), x(hij\cdot\ell), x(hi\cdot k\ell), x(h\cdot jk\ell)$	$2I(x^*_u:x^*_t) = 0.361$	1
	$2I(x:x^*_u) = 0.165$	1

Table 2b

Analysis of Information

Marginals Fitted	Information	D.F.
e) $x(\cdot ijk\ell), x(hi\cdot\cdot\cdot), x(h\cdot j\cdot\cdot), x(h\cdot\cdot k\cdot), x(h\cdot\cdot\cdot\ell)$	$2I(x:x^*_e) = 16.307$	11
v $x(\cdot ijk\ell), x(h\cdot j\cdot\cdot), x(h\cdot\cdot k\cdot), x(hi\cdot\cdot\ell)$	$2I(x^*_v:x^*_e) = 3.735$	1
	$2I(x:x^*_v) = 12.572$	10
w) $x(\cdot ijk\ell), x(h\cdot\cdot k\cdot), x(hi\cdot\cdot\ell), x(h\cdot j\cdot\ell)$	$2I(x^*_w:x^*_v) = 3.443$	1
	$2I(x:x^*_w) = 9.129$	9

Log-odds $\ell n \dfrac{x^*_e(1ij\ell)}{x^*_e(2ij\ell)}$

ijkℓ	Parametric representation					log-odds
1111	τ^h_1	$+\tau^{hi}_{11}$	$+\tau^{hj}_{11}$	$+\tau^{hk}_{11}$	$+\tau^{h\ell}_{11}$	0.2128
1112	τ^h_1	$+\tau^{hi}_{11}$	$+\tau^{hj}_{11}$	$+\tau^{hk}_{11}$		-0.1210
1121	τ^h_1	$+\tau^{hi}_{11}$	$+\tau^{hj}_{11}$		$+\tau^{h\ell}_{11}$	-0.1284
1122	τ^h_1	$+\tau^{hi}_{11}$	$+\tau^{hj}_{11}$			-0.4621
1211	τ^h_1	$+\tau^{hi}_{11}$		$+\tau^{hk}_{11}$	$+\tau^{h\ell}_{11}$	0.0888
1212	τ^h_1	$+\tau^{hi}_{11}$		$+\tau^{hk}_{11}$		-0.2450
1221	τ^h_1	$+\tau^{hi}_{11}$			$+\tau^{h\ell}_{11}$	-0.2524
1222	τ^h_1	$+\tau^{hi}_{11}$				-0.5861
2111	τ^h_1		$+\tau^{hj}_{11}$	$+\tau^{hk}_{11}$	$+\tau^{h\ell}_{11}$	0.4158
2112	τ^h_1		$+\tau^{hj}_{11}$	$+\tau^{hk}_{11}$		0.0820
2121	τ^h_1		$+\tau^{hj}_{11}$		$+\tau^{h\ell}_{11}$	0.0746
2122	τ^h_1		$+\tau^{hj}_{11}$			-0.2592
2211	τ^h_1			$+\tau^{hk}_{11}$	$+\tau^{h\ell}_{11}$	0.2918
2212	τ^h_1			$+\tau^{hk}_{11}$		-0.0420
2221	τ^h_1				$+\tau^{h\ell}_{11}$	-0.0494
2222	τ^h_1					-0.3831

$\tau^h_1 = -0.3831$, $\tau^{hi}_{11} = -0.2030$, $\tau^{hj}_{11} = 0.1240$

$\tau^{hk}_{11} = 0.3411$, $\tau^{h\ell}_{11} = 0.3338$

Table 3a

$$\text{Log-odds} \quad \ell n \frac{x_v^*(1ijk\ell)}{x_v^*(2ijk\ell)}$$

$ijk\ell$	Parametric representation						log-odds
1111	τ_1^h	$+\tau_{11}^{hi}$	$+\tau_{11}^{hj}$	$+\tau_{11}^{hk}$	$+\tau_{11}^{h\ell}$	$+\tau_{111}^{hi\ell}$	0.3571
1112	τ_1^h	$+\tau_{11}^{hi}$	$+\tau_{11}^{hj}$	$+\tau_{11}^{hk}$			-0.2898
1121	τ_1^h	$+\tau_{11}^{hi}$	$+\tau_{11}^{hj}$		$+\tau_{11}^{h\ell}$	$+\tau_{111}^{hi\ell}$	0.0115
1122	τ_1^h	$+\tau_{11}^{hi}$	$+\tau_{11}^{hj}$				-0.6355
1211	τ_1^h	$+\tau_{11}^{hi}$		$+\tau_{11}^{hk}$	$+\tau_{11}^{h\ell}$	$+\tau_{111}^{hi\ell}$	0.2366
1212	τ_1^h	$+\tau_{11}^{hi}$		$+\tau_{11}^{hk}$			-0.4101
1221	τ_1^h	$+\tau_{11}^{hi}$			$+\tau_{11}^{h\ell}$	$+\tau_{111}^{hi\ell}$	-0.1088
1222	τ_1^h	$+\tau_{11}^{hi}$					-0.7557
2111	τ_1^h		$+\tau_{11}^{hj}$	$+\tau_{11}^{hk}$	$+\tau_{11}^{h\ell}$		0.3847
2112	τ_1^h		$+\tau_{11}^{hj}$	$+\tau_{11}^{hk}$			0.1167
2121	τ_1^h		$+\tau_{11}^{hj}$		$+\tau_{11}^{h\ell}$		0.0390
2122	τ_1^h		$+\tau_{11}^{hj}$				-0.2290
2211	τ_1^h			$+\tau_{11}^{hk}$	$+\tau_{11}^{h\ell}$		0.2644
2212	τ_1^h			$+\tau_{11}^{hk}$			-0.0036
2221	τ_1^h				$+\tau_{11}^{h\ell}$		-0.0813
2222	τ_1^h						-0.3492

$$\tau_1^h = -0.3492, \quad \tau_{11}^{hi} = -0.4065, \quad \tau_{11}^{hj} = 0.1203$$

$$\tau_{11}^{hk} = 0.3457, \quad \tau_{11}^{h\ell} = 0.2680, \quad \tau_{111}^{hi\ell} = 0.3789$$

Table 3b

$$\text{Log-odds} \quad \ell n \frac{x^*_w(1ijk\ell)}{x^*_w(2ijk\ell)}$$

ijkℓ	Parametric representation							log-odds
1111	τ^h_1	$+\tau^{hi}_{11}$	$+\tau^{hj}_{11}$	$+\tau^{hk}_{11}$	$+\tau^{h\ell}_{11}$	$+\tau^{hi\ell}_{111}$	$+\tau^{hj\ell}_{111}$	0.3283
1112	τ^h_1	$+\tau^{hi}_{11}$	$+\tau^{hj}_{11}$	$+\tau^{hk}_{11}$				-0.2523
1121	τ^h_1	$+\tau^{hi}_{11}$	$+\tau^{hj}_{11}$		$+\tau^{h\ell}_{11}$	$+\tau^{hi\ell}_{111}$	$+\tau^{hj\ell}_{111}$	-0.0197
1122	τ^h_1	$+\tau^{hi}_{11}$	$+\tau^{hj}_{11}$					-0.6004
1211	τ^h_1	$+\tau^{hi}_{11}$		$+\tau^{hk}_{11}$	$+\tau^{h\ell}_{11}$	$+\tau^{hi\ell}_{111}$		0.3976
1212	τ^h_1	$+\tau^{hi}_{11}$		$+\tau^{hk}_{11}$				0.5396
1221	τ^h_1	$+\tau^{hi}_{11}$			$+\tau^{h\ell}_{11}$	$+\tau^{hi\ell}_{111}$		0.0495
1222	τ^h_1	$+\tau^{hi}_{11}$						-0.8876
2111	τ^h_1		$+\tau^{hj}_{11}$	$+\tau^{hk}_{11}$	$+\tau^{h\ell}_{11}$		$+\tau^{hj\ell}_{111}$	0.3542
2112	τ^h_1		$+\tau^{hj}_{11}$	$+\tau^{hk}_{11}$				0.1512
2121	τ^h_1		$+\tau^{hj}_{11}$		$+\tau^{h\ell}_{11}$		$+\tau^{hj\ell}_{111}$	0.0061
2122	τ^h_1		$+\tau^{hj}_{11}$					-0.1968
2211	τ^h_1			$+\tau^{hk}_{11}$	$+\tau^{h\ell}_{11}$			0.4235
2212	τ^h_1			$+\tau^{hk}_{11}$				-0.1360
2221	τ^h_1				$+\tau^{h\ell}_{11}$			0.0754
2222	τ^h_1							-0.4841

$\tau^h_1 = -0.4841$, $\tau^{hi}_{11} = -0.4035$, $\tau^{hj}_{11} = 0.2873$

$\tau^{hk}_{11} = 0.3481$, $\tau^{h\ell}_{11} = 0.5595$, $\tau^{hi\ell}_{111} = 0.3776$

$\tau^{hj\ell}_{111} = -0.3565$

Table 3c

E_1: {ijkℓ: ℓn odds ≥ 0}

E_1: Observations

ijkℓ	x(1ijkℓ)	x(2ijkℓ)
1111	82	53
1211	14	9
1221	11	10
2111	305	217
2112	200	172
2121	253	247
2211	41	25
2221	70	68
	976	801

$$\mu_2(E_1) = \frac{801}{1491}, \quad \mu_1(E_2) = \frac{1491-976}{1491}$$

$$\text{Prob. Error} = \frac{1}{2}\,\frac{801+515}{1491}$$

$$= \frac{1316}{2\text{x}1491} = 0.441$$

(a)

E_1: x_e^*

ijkℓ	x_e^*(1ijkℓ)	x_e^*(2ijkℓ)
1111	74.656	60.346
1211	12.010	10.990
2111	314.491	207.508
2112	193.625	178.376
2121	259.322	240.679
2211	37.780	28.219
	891.884	726.118

$$\mu_2(E_1) = \frac{726.118}{1491}, \quad \mu_1(E_2) = \frac{1491-891.884}{1491}$$

$$\text{Prob. Error} = \frac{1}{2}\,\frac{726.118+599.116}{1491}$$

$$= \frac{1325.234}{2982}$$

$$= 0.444$$

(b)

Table 4

$E_1 : x_v^*$

ijkℓ	$x_v^*(1ijk\ell)$	$x_v^*(2ijk\ell)$
1111	79.426	55.574
1121	50.789	50.211
1211	12.855	10.144
2111	310.589	211.411
2112	196.841	175.160
2121	254.876	245.125
2211	37.337	28.662
	942.713	776.287

$$\mu_2(E_1) = \frac{776.287}{1491}$$

$$\mu_1(E_2) = \frac{1491-942.713}{1491}$$

$$\text{Prob. Error} = \frac{1}{2}\ \frac{776.287+548.287}{1491}$$

$$= \frac{1324.574}{2982}$$

$$= 0.444$$

(c)

$E_1 : x_w^*$

ijkℓ	$x_w^*(1ijk\ell)$	$x_w^*(2ijk\ell)$
1111	78.482	56.517
1211	13.756	9.244
1212	8.102	13.898
1221	10.760	10.240
2111	306.748	215.252
2112	200.037	171.963
2121	250.769	249.232
2211	39.884	26.115
2221	71.600	66.401
	980.138	818.862

$$\mu_2(E_1) = \frac{818.862}{1491}$$

$$\mu_1(E_2) = \frac{1491-980.138}{1491}$$

$$\text{Prob. Error} = \frac{1}{2}\ \frac{818.862+510.862}{1491}$$

$$= \frac{1329.724}{2982}$$

$$= 0.446$$

(d)

Table 4

Martin and Bradley

E_1	$\hat{x}(1ijk\ell)$	$\hat{x}(2ijk\ell)$
1111	74.67	60.33
1211	12.02	10.98
2111	314.50	207.50
2112	193.45	178.55
2121	259.17	240.83
2211	37.74	28.26
	891.55	726.45

$$\mu_2(E_1) = \frac{726.45}{1491}, \quad \mu_1(E_2) = \frac{1491-891.55}{1491}$$

$$\text{Prob. Error} = \frac{1}{2}\ \frac{726.45+599.45}{1491}$$

$$= \frac{1325.90}{2982}$$

$$= 0.445$$

Table 4(e)

Example 2. Leukemia death observation at ABCC. This example illustrates the analysis of a three-way 5x6x2 contingency table. It illustrates the estimation procedure for the hypothesis of no second-order interaction. It also illustrates the use of a cell, other than the last one, as the reference cell. Details of the computation of the covariance matrix of a set of estimated parameters of interest is given. Confidence intervals for the parameters are computed using the multiple comparison lemma.

The Analysis of Leukemia

Death Observation at ABCC

Sugiura and Otake (1974) have considered the analysis of k 2xc contingency tables and have applied their procedures to the data in Table I. We propose to apply the minimum discrimination information estimation and associated concepts to the analysis of the data in Table I. We denote the occurrences in the three-way contingency Table I by x(ijk) with the notation

Variable	Index	1	2	3	4	5	6
Age	i	0-9	10-19	20-34	35-49	50+	
Dose	j	Not in city	0-9	10-49	50-99	100-199	200+
Mortality	k	Dead	Alive				

We get the minimum discrimination information estimates fitting the sets of marginals

a) x(ij.), x(..k)

b) x(ij.), x(i.k)

c) x(ij.), x(i.k), x(.jk)

d) x(ij.), x(.jk)

We start with the set of marginals x(ij.), x(..k) because $x_a^*(ijk) = x(ij.)\ x(..k)/n.$ is the m.d.i. or maximum-likelihood estimate under the null hypothesis that mortality is homogenous over the age by dose combinations.

We summarize the results in the Analysis of Information Table.

Analysis of Information Table

Component due to	Information		D.F.
a) x(ij.), x(..k)	$2I(x:x_a^*)$	= 205.983	29
b) x(ij.), x(i.k)	$2I(x_b^*:x_a^*)$	= 2.326	4
	$2I(x:x_b^*)$	= 203.657	25
c) x(ij.), x(i.k), x(.jk)	$2I(x_c^*:x_b^*)$	= 175.810	5
	$2I(x:x_c^*)$	= 27.847	20
a) x(ij.), x(..k)	$2I(x:x_a^*)$	= 205.983	29
d) x(ij.), x(.jk)	$2I(x_d^*:x_a^*)$	= 173.502	5
	$2I(x:x_d^*)$	= 32.481	24
c) x(ij.), x(.jk), x(i.k)	$2I(x_c^*:x_d^*)$	= 4.634	4
	$2I(x:x_c^*)$	= 27.847	20

We may draw the following inferences from the Analysis of Information Table.

1. Mortality is not homogeneous over the age by dose combinations ($2I(x:x_a^*) = 205.983$, 29 D.F.)

2. The effects of age by mortality are not significant ($2I(x_b^*:x_a^*) = 2.326$, 4 D.F., $2I(x_c^*:x_d^*) = 4.634$, 4 D.F.)

3. The effects of dose by mortality are highly significant ($2I(x_c^*:x_b^*) = 175.810$, 5 D.F., $2I(x_d^*:x_a^*) = 173.502$, 5 D.F.)

Since the value of $2I(x:x_d^*) = 32.481$, 24 D.F. is not significant at the 10% level, we obtained the complete output for x_d^* and the estimates are shown in Table IIb. However since four OUTLIER values were indicated for x_d^*, and for comparison with the results of Sugiura and Otake, it was decided to perform a more complete analysis with the estimate fitting all the two-way marginals, that is, the estimate corresponding to an hypothesis of no second-order interaction. This estimate is given in Table IIa and we have called it $x_2^*(ijk)$, that is, $x_2^*(ijk) \equiv x_c^*(ijk)$.

Again for easier comparison with the results of Sugiura and Otake we selected the cell (512) as the reference cell so that the log-linear representation of $x_2^*(ijk)$ is given by

$$\ln \frac{x_2^*(ijk)}{n(1/60)} = L + \tau_1^i T_1^i(ijk) + \ldots + \tau_4^i T_4^i(ijk) + \tau_2^j T_2^j(ijk) + \ldots + \tau_6^j T_6^j(ijk) +$$

$$+ \tau_1^k T_1^k(ijk) + \tau_{12}^{ij} T_{12}^{ij}(ijk) + \ldots + \tau_{46}^{ij} T_{46}^{ij}(ijk) + \tau_{11}^{ik} T_{11}^{ik}(ijk) + \ldots$$

$$+ \tau_{41}^{ik} T_{41}^{ik}(ijk) + \tau_{21}^{jk} T_{21}^{jk}(ijk) + \ldots + \tau_{61}^{jk} T_{61}^{jk}(ijk)$$

where $L = 1$, the taus are main effect and interaction parameters and the $T(ijk)$ are the explanatory variables, the indicator functions of the corresponding marginals, e.g. $\sum_{ijk} T_{12}^{ij}(ijk)x_2^*(ijk) = x_2^*(12.) = x(12.)$ etc.

From the log-linear representation of $x_2^*(ijk)$ we have the log-linear representation of the mortality log-odds or logit as

$$\ell n \frac{x_2^*(ij1)}{x_2^*(ij2)} = \tau_1^k + \tau_{i1}^{ik} + \tau_{j1}^{jk}$$

where $\tau_{51}^{ik}=0=\tau_{11}^{jk}$.

Since the computer output includes log's of the x_2^* we can evaluate the tau parameters, for example, as follows

$$\ell n \frac{x_2^*(511)}{x_2^*(512)} = \tau_1^k$$

$$\ell n \frac{x_2^*(111)}{x_2^*(112)} = \tau_1^k + \tau_{11}^{ik}$$

$$\vdots$$

$$\ell n \frac{x_2^*(411)}{x_2^*(412)} = \tau_1^k + \tau_{41}^{ik}$$

$$\ell n \frac{x_2^*(521)}{x_2^*(522)} = \tau_1^k + \tau_{21}^{jk}$$

$$\vdots$$

$$\ell n \frac{x_2^*(561)}{x_2^*(562)} = \tau_1^k + \tau_{61}^{jk}$$

The following are the values obtained

$$\tau_1^{k} = -7.4714 \qquad \tau_{21}^{jk} = 0.5017$$

$$\tau_{11}^{ik} = -0.0849 \qquad \tau_{31}^{jk} = 0.9685$$

$$\tau_{21}^{ik} = -0.4515 \qquad \tau_{41}^{jk} = 1.2848$$

$$\tau_{31}^{ik} = -0.2655 \qquad \tau_{51}^{jk} = 2.2293$$

$$\tau_{41}^{ik} = 0.0371 \qquad \tau_{61}^{jk} = 3.4785$$

Sugiura and Otake used the representation for the log-odds

$$\log \{p_{ij}/(1-p_{ij})\} = \mu + \alpha_i + \beta_j$$

where $\sum_{i=1}^{5} \alpha_i = 0$, $\beta_1 = 0$ and give the estimates

$$\hat{\alpha}_1 = 0.068 \qquad \hat{\beta}_2 = 0.502$$

$$\hat{\alpha}_2 = -0.299 \qquad \hat{\beta}_3 = 0.969$$

$$\hat{\alpha}_3 = -0.113 \qquad \hat{\beta}_4 = 1.285$$

$$\hat{\alpha}_4 = 0.190 \qquad \hat{\beta}_5 = 2.229$$

$$\hat{\alpha}_5 = 0.153 \qquad \hat{\beta}_6 = 3.478$$

We note that $\tau_{21}^{jk} = \hat{\beta}_2, \ldots, \tau_{61}^{jk} = \hat{\beta}_6$ and

$$\mu + \alpha_5 = \tau_1^{k}$$

$$\mu + \alpha_1 = \tau_1^{k} + \tau_{11}^{ik}$$

$$\mu + \alpha_2 = \tau_1^{k} + \tau_{21}^{ik}$$

$$\mu + \alpha_3 = \tau_1^{k} + \tau_{31}^{ik}$$

$$\mu + \alpha_4 = \tau_1^{k} + \tau_{41}^{ik}$$

that is

$$\alpha_1 = \tau_{11}^{ik} - (\tau_{11}^{ik} + \tau_{21}^{ik} + \tau_{31}^{ik} + \tau_{41}^{ik})/5$$

$$\alpha_2 = \tau_{21}^{ik} - (\tau_{11}^{ik} + \tau_{21}^{ik} + \tau_{31}^{ik} + \tau_{41}^{ik})/5$$

$$\alpha_3 = \tau_{31}^{ik} - (\tau_{11}^{ik} + \tau_{21}^{ik} + \tau_{31}^{ik} + \tau_{41}^{ik})/5$$

$$\alpha_4 = \tau_{41}^{ik} - (\tau_{11}^{ik} + \tau_{21}^{ik} + \tau_{31}^{ik} + \tau_{41}^{ik})/5$$

$$\alpha_5 = -(\tau_{11}^{ik} + \tau_{21}^{ik} + \tau_{31}^{ik} + \tau_{41}^{ik})/5$$

yielding $\alpha_1 = 0.0680$, $\alpha_2 = -0.2986$, $\alpha_3 = -0.1126$,

$\alpha_4 = 0.1900$, $\alpha_5 = 0.1529$.

We determine the covariance matrix of the tau's in the logit representation as follows.

Let T denote the 60x40 matrix whose columns are

$$L,\ \overbrace{T_1^i(ijk),\ldots,T_4^i(ijk)}^{4},\ \overbrace{T_2^j(ijk),\ldots,T_6^j(ijk)}^{5},\ \overbrace{T_{12}^{ij}(ijk),\ldots,T_{46}^{ij}(ijk)}^{20}$$

$$T_1^k(ijk),\ \underbrace{T_{11}^{ik}(ijk),\ldots,T_{41}^{ik}(ijk)}_{4},\ \underbrace{T_{21}^{jk}(ijk),\ldots,T_{61}^{jk}(ijk)}_{5}$$

and let D denote a 60x60 diagonal matrix whose diagonal values are $x_2^*(ijk)$ (in the same ijk sequence as the T(ijk) functions).

Compute the 40x40 matrix $S = T'DT$

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

where $\underline{S}_{11}$ is 30x30 and $\underline{S}_{22}$ is 10x10

The covariance matrix of $\tau_1^k, \tau_{11}^{ik}, \ldots, \tau_{41}^{ik}, \tau_{21}^{jk}, \ldots, \tau_{61}^{jk}$

is then given by

$$S_{22.1}^{-1} = (S_{22} - S_{21}S_{11}^{-1}S_{12})^{-1}.$$

The covariance matrix thus obtained is given in Table III.

To compute confidence intervals for the τ^{jk}'s, following the procedure suggested by Sugiura and Otake using the multiple comparison lemma, Ferguson (1967, p. 282), we computed $\sqrt{11.070 \times V_{\tau_{j1}^{jk}}}$ using the variances in Table III and obtained the following confidence intervals

τ_{21}^{jk}	-0.5463	1.5497
τ_{31}^{jk}	-0.2295	2.1665
τ_{41}^{jk}	-0.2762	2.8458
τ_{51}^{jk}	0.9233	3.5353
τ_{61}^{jk}	2.4185	4.5385

The confidence intervals for the τ^{ik}'s were obtained by computing $\sqrt{9.488 V_{\tau_{i1}^{ik}}}$ using the variances in Table III leading to

τ_{11}^{ik}	-0.9689	0.7991
τ_{21}^{ik}	-0.4515	0.4455
τ_{31}^{ik}	-1.1525	0.6215
τ_{41}^{ik}	-0.7759	0.8501

To relate with the bounds given by Sugiura and Otake for the α's, since we have seen that

$$\alpha_5 = -(\tau_{11}^{ik} + \tau_{21}^{ik} + \tau_{31}^{ik} + \tau_{41}^{ik})/5$$

we have that

$$\mathrm{Var}(\alpha_5) = \frac{1}{25}\left\{\mathrm{Var}(\tau_{11}^{ik}) + \ldots + \mathrm{Var}(\tau_{41}^{ik}) + 2\sum_{m<n} \mathrm{cov}(\tau_{ml}^{ik}, \tau_{nl}^{ik})\right\}$$

and from the entries in Table III we finally find

$\mathrm{Var}(\alpha_5) = 0.0339$, leading to the interval

α_5 (-0.4141, 0.7199).

We did not trouble to compute the others as it is evident that the results are the same.

In the output corresponding to fitting all the two-way marginals, the entry corresponding to the cell x(111) had a large OUTLIER value (5.239). Accordingly we fitted an estimate fitting all the two-way marginals but omitting the values x(111), x(112). This estimate is denoted by $x_e^*(ijk)$ and its values are given in Table IIc.

The associated Analysis of Information is

Analysis of Information

Component due to	Information	D.F.
x(ij.), x(i.k), x(.jk)	$2I(x:x_2^*) = 27.847$	20
as above but omitting x(111),x(112)	$2I(x_e^*:x_2^*) = 6.223$	1
	$2I(x:x_e^*) = 21.614$	19

Removing x(111), x(112) from the estimation gives an improved fit. We did not carry out any extensive analysis with $x_e^*(ijk)$ but did note the approximate equality of

$$\tau_{61}^{jk} - \tau_{51}^{jk},\ \tau_{51}^{jk} - \tau_{41}^{jk},\ \tau_{41}^{jk} - \tau_{31}^{jk},\ \tau_{31}^{jk} - \tau_{21}^{jk}$$

when computed for x_e^* and x_2^*, the respective values being

x_e^*	x_2^*
1.249	1.249
0.949	0.944
0.320	0.316
0.466	0.467

TABLE I
Original Data x(ijk)

Age	i	Not in city j=1 dead k=1	Not in city j=1 alive k=2	0-9 j=2 dead k=1	0-9 j=2 alive k=2	10-49 j=3 dead k=1	10-49 j=3 alive k=2	50-99 j=4 dead k=1	50-99 j=4 alive k=2	100-199 j=5 dead k=1	100-199 j=5 alive k=2	200+ j=6 dead k=1	200+ j=6 alive k=2
0-9	1	0	5015	7	10752	3	2989	1	694	4	418	11	387
10-19	2	5	5973	4	11811	6	2620	1	771	3	792	6	820
20-34	3	2	5669	8	10828	3	2798	1	797	3	596	9	624
35-49	4	3	6158	19	12645	4	3566	2	972	1	694	10	608
50+	5	3	3695	7	9053	3	2415	2	655	2	393	6	289
		13	26510	45	55089	19	14388	7	3889	13	2893	42	2728

TABLE IIa
Estimates

x_2^*(ijk) Fitting marginals x(ij.), x(i.k), x(.jk)

i	j=1 k=1	j=1 k=2	j=2 k=1	j=2 k=2	j=3 k=1	j=3 k=2	j=4 k=1	j=4 k=2	j=5 k=1	j=5 k=2	j=6 k=1	j=6 k=2
1	2.621	5012.379	9.282	10749.719	4.115	2987.887	1.311	693.689	2.040	419.960	6.632	391.369
2	2.165	5975.828	7.066	11807.930	2.504	2623.496	1.010	770.990	2.668	792.332	9.588	816.411
3	2.474	5668.520	7.804	10828.191	3.216	2797.783	1.257	796.743	2.420	596.580	8.829	624.170
4	3.637	6157.359	12.341	12651.648	5.546	3564.455	2.075	971.925	3.794	691.206	11.608	606.391
5	2.103	3695.898	8.507	9051.488	3.619	2414.382	1.349	655.652	2.078	392.922	5.343	289.657

TABLE IIb

$x_d^*(ijk)$ Fitting marginals x(ij.), x(.jk), $x_d^*(ijk) = x(ij.)x(.jk)/x(.j.)$

	j=1		j=2		j=3		j=4		j=5		j=6	
i	k=1	k=2	k=1	k=2	k=1	k=2	k=1	k=2	k=1	k=2	k=1	k=2
1	2.458	5012.543	8.781	10750.215	3.946	2988.055	1.249	693.751	1.888	420.112	6.035	391.965
2	2.930	5975.070	9.643	11805.352	3.463	2622.537	1.387	770.613	3.556	791.443	12.524	813.476
3	2.780	5668.219	8.844	10827.152	3.694	2797.307	1.434	796.566	2.679	596.320	9.598	623.402
4	3.020	6157.980	10.336	12653.660	4.708	3565.293	1.750	972.250	3.109	691.891	9.370	608.629
5	1.813	3696.189	7.395	9052.602	3.189	2414.812	1.180	655.819	1.767	393.233	4.473	290.527

TABLE IIc

$x_e^*(ijk)$ - one outlier removed from $x_2^*(ijk)$, that is x(111), x(112)

	j=1		j=2		j=3		j=4		j=5		j=6	
i	k=1	k=2	k=1	k=2	k=1	k=2	k=1	k=2	k=1	k=2	k=1	k=2
1	0	5015	10.303	10748.695	4.561	2987.441	1.459	693.541	2.279	419.720	7.398	390.602
2	2.715	5975.285	6.870	11808.125	2.431	2623.570	0.984	771.015	2.612	792.388	9.388	816.612
3	3.095	5667.902	7.573	10828.422	3.117	2797.884	1.223	796.777	2.364	596.635	8.628	624.371
4	4.554	6156.441	11.984	12652.012	5.378	3564.625	2.020	971.980	3.710	691.290	11.354	606.645
5	2.636	3695.365	8.270	9051.727	3.514	2414.487	1.314	655.685	2.034	392.966	5.231	289.769

TABLE III

Covariance Matrix τ_1^k, τ_{11}^{ik}, τ_{21}^{ik}, τ_{31}^{ik}, τ_{41}^{ik}, τ_{21}^{jk}, τ_{31}^{jk}, τ_{41}^{jk}, τ_{51}^{jk}, τ_{61}^{jk}

	τ_1^k	τ_{11}^{ik}	τ_{21}^{ik}	τ_{31}^{ik}	τ_{41}^{ik}	τ_{21}^{jk}	τ_{31}^{jk}	τ_{41}^{jk}	τ_{51}^{jk}	τ_{61}^{jk}
τ_1^k	0.1140	-0.0445	-0.0438	-0.0443	-0.0441	-0.0782	-0.0782	-0.0783	-0.0769	-0.0755
τ_{11}^{ik}		0.0824	0.0438	0.0438	0.0438	0.0010	0.0007	0.0019	0.0016	0.0002
τ_{21}^{ik}			0.0849	0.0444	0.0441	0.0016	0.0027	0.0023	-0.0017	-0.0041
τ_{31}^{ik}				0.0829	0.0440	0.0019	0.0021	0.0018	0.0000	-0.0024
τ_{41}^{ik}					0.0697	0.0013	0.0010	0.0009	-0.0004	-0.0014
τ_{21}^{jk}						0.0993	0.0770	0.0770	0.0769	0.0769
τ_{31}^{jk}							0.1298	0.0770	0.0769	0.0768
τ_{41}^{jk}								0.2202	0.0770	0.0769
τ_{51}^{jk}									0.1544	0.0771
τ_{61}^{jk}										0.1015

Example 3. **Automobile accident data.** This example illustrates the analysis of a four-way 3x4x3x2 contingency table. It points out that the model fitted determines the form of the log-odds or logit representation, but the converse is not true. The covariance matrix of the estimated parameters is given.

Example

Automobile Accident Data - Driver Ejection

Data used on this example are taken from a study of the relationship between car size and accident injuries as given in Kihlberg et al. (1964). The observed data are given in Table 1 and the observed occurrences are denoted by $x(ijk\ell)$ where

Characteristic	Index	1	2	3	4
Car weight	i	Small	Compact	Standard	
Accident type	j	Collision with vehicle	Collision with object	Rollover without collision	Other rollover
Severity	k	Not severe	Mod. severe	Severe	
Driver Ejection	ℓ	Not ejected	Ejected		

A condensed 2x2x2x2 version of this data was studied by Bhapkar and Koch (1968) and Ku et al. (1968).

Since the question of interest is the possible relation of driver ejection on car weight, accident type and severity, we start the fitting sequence with the marginals $x(ijk.)$, $x(...\ell)$. This first estimate, $x_a^*(ijk\ell) = x(ijk.)x(...\ell)/n$, corresponds to a null hypothesis that driver ejection is homogeneous over the 36 combinations of the other characteristics. As may be seen from the analysis of information table this hypothesis is clearly rejected by the data. It is found that fitting the model incorporating in addition to $x(ijk.)$ the marginals $x(i..\ell)$, $x(.j.\ell)$, $x(..k\ell)$, that is, the interactions of car weight, accident type, and severity respectively with driver ejection, a satisfactory fit to the observed data is obtained. The models fitting in addition three-way marginals $x(ij.\ell)$, etc., showed no significant effects for the associated interaction parameters. The results are summarized in the analysis of information table.

Analysis of Information

Component due to			D.F.
a) $x(ijk.)$, $x(...\ell)$	$2I(x:x_a^*)$	= 613.102	35
b) $x(ijk.)$, $x(i..\ell)$, $x(.j.\ell)$, $x(..k\ell)$	$2I(x_b^*:x_a^*)$	= 587.584	7
	$2I(x:x_b^*)$	= 25.518	28
c) $x(ijk.)$, $x(ij.\ell)$, $x(i.k\ell)$, $x(.jk\ell)$	$2I(x_c^*:x_b^*)$	= 14.491	16
	$2I(x:x_c^*)$	= 11.028	12

The fitted values $x_b^*(ijk\ell)$ are given in Table 2. The log-linear regression representation of $x_b^*(ijk\ell)$ contains the parameters L (a normalizing constant), τ_1^i, τ_2^i, τ_1^j, τ_2^j, τ_3^j, τ_1^k, τ_2^k, τ_1^ℓ, τ_{11}^{ij}, τ_{12}^{ij}, τ_{13}^{ij}, τ_{21}^{ij}, τ_{22}^{ij}, τ_{23}^{ij}, τ_{11}^{ik}, τ_{12}^{ik}, τ_{21}^{ik}, τ_{22}^{ik}, $\tau_{11}^{i\ell}$, $\tau_{21}^{i\ell}$, τ_{11}^{jk}, τ_{12}^{jk}, τ_{21}^{jk}, τ_{22}^{jk}, τ_{31}^{jk}, τ_{32}^{jk}, $\tau_{11}^{j\ell}$, $\tau_{21}^{j\ell}$, $\tau_{31}^{j\ell}$, $\tau_{11}^{k\ell}$, $\tau_{21}^{k\ell}$, τ_{111}^{ijk}, τ_{112}^{ijk}, τ_{121}^{ijk}, τ_{122}^{ijk}, τ_{131}^{ijk}, τ_{132}^{ijk}, τ_{211}^{ijk}, τ_{212}^{ijk}, τ_{221}^{ijk}, τ_{222}^{ijk}, τ_{231}^{ijk}, τ_{232}^{ijk}. The 28 additional parameters which would appear in the complete model for $x(ijk\ell)$ are hypothesized as zero and represent the 28 degrees of freedom of $2I(x:x_b^*)$. The log-odds or logit representation for the estimate x_b^* is

$$\ell n \frac{x_b^*(ijk1)}{x_b^*(ijk2)} = \tau_1^\ell + \tau_{i1}^{i\ell} + \tau_{j1}^{j\ell} + \tau_{k1}^{k\ell} .$$

Parameters not involving ℓ are common to numerator and denominator of the odds and drop out. The values of the parameters may be obtained as

$$\tau_1^{\ell} = \ln \frac{x_b^*(3431)}{x_b^*(3432)}$$

$$\tau_{11}^{i\ell} = \ln \frac{x_b^*(1431)}{x_b^*(1432)} - \tau_1^{\ell}$$

$$\tau_{21}^{i\ell} = \ln \frac{x_b^*(2431)}{x_b^*(2432)} - \tau_1^{\ell}$$

etc.

The values of the parameters are (in this case provided as computer output)

$\tau_1^{\ell} = -0.0083$	$\tau_{11}^{j\ell} = 1.3665$	$\tau_{11}^{k\ell} = 1.6085$
$\tau_{11}^{i\ell} = -0.2936$	$\tau_{21}^{j\ell} = 1.1139$	$\tau_{21}^{k\ell} = 0.8823$.
$\tau_{21}^{i\ell} = -0.0788$	$\tau_{31}^{j\ell} = -0.2405$	

We recall that any parameter with a subscript i=3 and/or j=4 and/or k=3 and/or ℓ=2 is by convention zero.

It is important to note that the estimate $x_2^*(ijk\ell)$ obtained by fitting the two-way marginals $x(ij..)$, $x(i.k.)$, $x(i..\ell)$, $x(.jk.)$, $x(.j.\ell)$,$x(..k\ell)$ would also have the log-odds or logit representation

$$\ln \frac{x_2^*(ijk1)}{x_2^*(ijk2)} = \tau_1^{\ell} + \tau_{i1}^{i\ell} + \tau_{j1}^{j\ell} + \tau_{k1}^{k\ell} .$$

The values of the parameters would depend however on the values of the estimate $x_2^*(ijk\ell)$.

The model fitted determines the form of the log-odds or logit representation but the converse is not true.

For easier interpretation of the numerical values we use the representation of the estimated odds as the multiplicative model

$$\frac{x_b^*(ijk1)}{x_b^*(ijk2)} = \exp(\tau_1^{\ell})\exp(\tau_{i1}^{i\ell})\ \exp(\tau_{j1}^{j\ell})\ \exp(\tau_{k1}^{k\ell})$$

The factors which determine the odds of not ejected for any combination of the characteristics are:

Factors

Base	Car weight	Accident type		Severity	
0.99	Small 0.75	Collision with vehicle	3.92	Not severe	5.00
	Compact 0.92	Collision with object	3.05	Mod. severe	2.42
	Standard 1.00	Rollover without collision	0.79	Severe	1.00
		Other rollover	1.00		

By selecting the combination of characteristics with the largest factors, it is seen that the best odds for not ejected, 19.40, occur for

Standard, Collision with vehicle, Not severe.

By selecting the combination of characteristics with the smallest factors, it is seen that the worst odds for Not ejected, 0.59, occur for

Small, Rollover without collision, Severe.

The observed odds for Not ejected from the original data are 4124/707=5.83. The estimated odds for any combination of characteristics is easily obtained from the values of x_b^*.

The covariance matrix of the parameters for the estimate x_b^* is given in Table 3.

Table 1

Accident Data - Drivers Alone - Observed

Accident type	Accident severity	Not Ejected			Ejected		
		Small	Compact	Standard	Small	Compact	Standard
Collision with vehicle	Not severe	95	166	1279	8	7	65
	Mod. severe	31	34	506	2	5	51
	Severe	11	17	186	4	5	54
Collision with object	Not severe	34	55	599	5	6	46
	Mod. Severe	8	34	241	2	4	26
	Severe	5	10	39	0	1	30
Rollover without Collision	Not severe	23	13	65	6	5	11
	Mod. severe	22	17	118	18	9	63
	Severe	5	2	23	5	6	33
Other Rollover	Not severe	9	10	83	6	2	11
	Mod. severe	23	26	177	13	16	78
	Severe	8	9	86	7	6	86
		274	398	3452	76	72	559

Table 2

Accident data - Drivers Alone - Estimate x_b^*

Accident type	Accident severity	Not ejected Small	Not ejected Compact	Not ejected Standard	Ejected Small	Ejected Compact	Ejected Standard
Collision with vehicle	Not severe	96.349	163.874	1278.209	6.651	9.126	65.790
	Mod. severe	28.879	34.973	503.433	4.121	4.027	53.567
	Severe	11.154	17.212	190.913	3.846	4.788	49.087
Collision with object	Not severe	35.817	56.919	604.917	3.183	4.081	40.082
	Mod. severe	8.448	33.095	234.832	1.552	4.905	32.167
	Severe	3.463	8.099	89.406	1.537	2.901	29.594
Rollover without Collision	Not severe	21.572	18.000	60.475	7.428	5.000	15.525
	Mod. severe	23.367	16.516	121.512	16.633	9.484	64.488
	Severe	3.676	3.351	24.535	6.324	4.649	31.465
Other Rollover	Not severe	11.804	9.849	78.213	3.196	2.151	15.787
	Mod. severe	23.082	28.936	179.924	12.918	13.064	75.076
	Severe	6.377	7.174	85.645	8.623	7.826	86.355
		273.988	397.998	3452.014	76.012	72.002	558.983

Table 3

Covariance matrix - parameters of estimate x_b^*

τ_1^{ℓ}	$\tau_{11}^{i\ell}$	$\tau_{21}^{i\ell}$	$\tau_{11}^{j\ell}$	$\tau_{21}^{j\ell}$	$\tau_{31}^{j\ell}$	$\tau_{11}^{k\ell}$	$\tau_{21}^{k\ell}$
.0017	.0003	.0003	.0005	.0003	.0003	.0005	.0003
	.0039	-.0003	.0000	-.0001	.0005	.0001	.0001
		.0027	.0001	.0000	.0001	.0001	.0000
			.0008	-.0005	-.0004	.0003	.0000
				.0012	-.0003	.0002	.0000
					.0036	-.0001	.0003
						.0008	-.0006
							.0011

Example 4. Minnesota high school graduates of June 1938. This example illustrates the analysis of a four-way 2x3x7x4 contingency table. In particular the "dependent" classification is not dichotomous as in the previous examples but has four categories. The final model leads to log-odds representations involving main effects and interactions.

Example

Classification of Minnesota High School

Graduates of June 1938

The data of this 2x3x7x4 contingency table represents a four-way cross classification of the April 1939 status of 13,968 Minnesota High School graduates of June 1938. The data was presented by Hoyt et al. (1959). They formulated and tested various hypotheses of independence using chi-squared statistics. The same data was also used by Kullback et al. (1962b) to illustrate the use of the minimum discrimination information statistics in the analysis of various hypotheses of independence and homogeneity. Patil (1974) condensed the original data into a 4x3x7 table by summing over the sex classification and tested for no second-order interaction in the three-way table by an asymptotic chi-squared statistic.

We shall examine models fitting certain sets of marginals and analyze the data on the basis of the log-linear representation of a model that well fits the data. The original data is listed in Table 1 where we denote the occurrences in the cells by x(hijk), with

Characteristic	Index	1	2	3	4	5	6	7
Sex	h	Male	Female					
H.S. Rank	i	Lowest third	Middle third	Upper third				
Father's Occupational Level	j	1	2	3	4	5	6	7
Post H.S. Status	k	Enrolled in College	Noncollegiate school	Employed full time	Other			

The problem is to determine the relationship of post high-school status on the other variables. Note that here the 'dependent" variable is polychotomous. We summarize in the analysis of information Table 3, the results of fitting three models to the data, or the sets of marginals,

H_a: $x(hij\cdot)$, $x(\cdot\cdot\cdot k)$,

H_b: $x(hij\cdot)$, $x(h\cdot\cdot k)$, $x(\cdot i\cdot k)$, $x(\cdot\cdot jk)$,

H_c: $x(hij\cdot)$, $x(\cdot i\cdot k)$, $x(h\cdot jk)$.

The estimate x_a^*, corresponding to H_a, is to determine whether the occurrences of post high-school status are homogeneously distributed over the 42 combinations of sex, high-school rank, and father's occupational level. We note that $x_a^*(hijk) = x(hij\cdot)\, x(\cdot\cdot\cdot k)/n$. Since the data do not support the null hypothesis of homogeneity we consider the estimate x_b^* corresponding to H_b. This estimate will provide a log-odds or logit representation in terms of a linear combination of the main effects of sex, high-school rank and father's occupational level on post high-school status. Since the fit of the estimate x_b^* to the data was not considered satisfactory the effects of various interactions associated with three-way marginals were examined. The interaction with the largest effect, for the additional degrees of freedom, turned out to be that of sex x father's occupational level x post high-school status, that is, associated with the marginal $x(h\cdot jk)$. It was decided to analyze the data in terms of the estimate x_c^* corresponding to H_c. The values of $x_c^*(hijk)$ are listed in Table 2.

From the log-linear representation of the estimate x_c^*, we arrive at the following representation for the log-odds

$$\ell n \frac{x_c^*(hij1)}{x_c^*(hij4)} = \tau_1^k + \tau_{h1}^{hk} + \tau_{i1}^{ik} + \tau_{j1}^{jk} + \tau_{hj1}^{hjk} ,$$

$$\ell n \frac{x_c^*(hij2)}{x_c^*(hij4)} = \tau_2^k + \tau_{h2}^{hk} + \tau_{i2}^{ik} + \tau_{j2}^{jk} + \tau_{hj2}^{hjk} ,$$

$$\ell n \frac{x_c^*(hij3)}{x_c^*(hij4)} = \tau_3^k + \tau_{h3}^{hk} + \tau_{i3}^{ik} + \tau_{j3}^{jk} + \tau_{hj3}^{hjk} .$$

The values of the parameters in the log-odds representations are:

$\tau_1^k = -1.0345$	$\tau_2^k = -2.2548$	$\tau_3^k = -1.7189$
$\tau_{11}^{hk} = 0.9935$	$\tau_{12}^{hk} = -0.3523$	$\tau_{13}^{hk} = -0.1111$
$\tau_{11}^{ik} = -1.5908$	$\tau_{12}^{ik} = -1.0060$	$\tau_{13}^{ik} = -1.0682$
$\tau_{21}^{ik} = -0.8912$	$\tau_{22}^{ik} = -0.4542$	$\tau_{23}^{ik} = -0.4934$
$\tau_{11}^{jk} = 2.2731$	$\tau_{12}^{jk} = 0.9905$	$\tau_{13}^{jk} = 0.8593$
$\tau_{21}^{jk} = 1.2332$	$\tau_{22}^{jk} = 0.9822$	$\tau_{23}^{jk} = 0.6872$
$\tau_{31}^{jk} = 0.4009$	$\tau_{32}^{jk} = 0.3932$	$\tau_{33}^{jk} = 0.6333$
$\tau_{41}^{jk} = 1.1259$	$\tau_{42}^{jk} = 0.8881$	$\tau_{43}^{jk} = 0.6099$
$\tau_{51}^{jk} = 0.6194$	$\tau_{52}^{jk} = 0.3995$	$\tau_{53}^{jk} = 0.5254$
$\tau_{61}^{jk} = -0.0321$	$\tau_{62}^{jk} = -0.1397$	$\tau_{63}^{jk} = 0.1989$
$\tau_{111}^{hjk} = -0.7277$	$\tau_{112}^{hjk} = -1.3054$	$\tau_{113}^{hjk} = -0.4037$
$\tau_{121}^{hjk} = -0.6340$	$\tau_{122}^{hjk} = -0.8018$	$\tau_{123}^{hjk} = -0.3643$
$\tau_{131}^{hjk} = -1.0923$	$\tau_{132}^{hjk} = -0.8080$	$\tau_{133}^{hjk} = -0.9709$
$\tau_{141}^{hjk} = -0.8463$	$\tau_{142}^{hjk} = -0.7581$	$\tau_{143}^{hjk} = -0.5573$

$\tau^{hijk}_{151} = -0.6402$ $\tau^{hijk}_{152} = -0.8605$ $\tau^{hijk}_{153} = -0.5503$

$\tau^{hijk}_{161} = -0.7587$ $\tau^{hijk}_{162} = -0.2334$ $\tau^{hijk}_{163} = -0.4397$

All parameters with subscripts h=2 and/or i=3 and/or j=7 and/or k=4 are zero by convention.

From the representation for the log-odds it is seen that the association between high-school rank and post high-school status is independent of the combination of sex and father's occupational level, that is,

$$\ln \frac{x_c^*(hij1)}{x_c^*(h1j4)} - \ln \frac{x_c^*(h2j1)}{x_c^*(h2j4)} = \ln \frac{x_c^*(h1j1)x_c^*(h2j4)}{x_c^*(h1j4)x_c^*(h2j1)}$$

$$= \tau^{ik}_{11} - \tau^{ik}_{21} = -0.6996 \ ,$$

$$\ln \frac{x_c^*(h2j1)x_c^*(h3j4)}{x_c^*(h2j4)x_c^*(h3j1)} = \tau^{ik}_{21} = -0.3912 \ ,$$

$$\ln \frac{x_c^*(h1j2)x_c^*(h2j4)}{x_c^*(h1j4)x_c^*(h2j2)} = \tau^{ik}_{12} - \tau^{ik}_{22} = -0.5518 \ ,$$

$$\ln \frac{x_c^*(h2j2)x_c^*(h3j4)}{x_c^*(h2j4)x_c^*(h3j2)} = \tau^{ik}_{22} = -0.4542 \ ,$$

$$\ln \frac{x_c^*(h1j3)x_c^*(h2j3)}{x_c^*(h1j4)x_c^*(h2j4)} = \tau^{ik}_{13} - \tau^{ik}_{23} = -0.5748 \ ,$$

$$\ln \frac{x_c^*(h2j3)x_c^*(h3j3)}{x_c^*(h2j4)x_c^*(h3j4)} = \tau^{ik}_{23} = -0.4934 \ .$$

The association between sex and post high-school status is of course dependent on father's occupational level, that is,

$$\ell n \frac{x_c^*(1ij1)}{x_c^*(1ij4)} - \ell n \frac{x_c^*(2ij1)}{x_c^*(2ij4)} = \tau_{11}^{hk} + \tau_{1j1}^{hjk} ,$$

$$\ell n \frac{x_c^*(1ij2)}{x_c^*(1ij4)} - \ell n \frac{x_c^*(2ij2)}{x_c^*(2ij4)} = \tau_{12}^{hk} + \tau_{1j2}^{hjk} ,$$

$$\ell n \frac{x_c^*(1ij3)}{x_c^*(1ij4)} - \ell n \frac{x_c^*(2ij3)}{x_c^*(2ij4)} = \tau_{13}^{hk} + \tau_{1j3}^{hjk} .$$

We summarize the numerical values below.

j	$\tau_{11}^{hk} + \tau_{1j1}^{hjk}$	$\tau_{12}^{hk} + \tau_{1j2}^{hjk}$	$\tau_{13}^{hk} + \tau_{1j3}^{hjk}$
1	0.2658	-1.6577	-0.5148
2	0.3595	-1.1541	-0.4754
3	-0.0988	-1.1603	-1.0820
4	0.1472	-1.1104	-0.6684
5	0.3533	-1.2128	-0.6619
6	0.2348	-0.5857	-0.5508
7	0.9935	-0.3523	-0.1111

We remark that father's occupational level 3 shows a peculiarity as compared to other values in the first column above. Kullback et al. (1962b, p. 593) noted that there was an unusually larger number of girls than boys for the third category of father's occupation. Apparently there was a tendency for the girls not to enroll in college as compared to the boys. In particular, for example, the association between sex and collegiate or noncollegiate school is

$$\ln \frac{x^*_c(1ij1)}{x^*_c(1ij2)} - \ln \frac{x^*_c(2ij1)}{x^*_c(2ij2)} = \tau^{hk}_{11} + \tau^{hjk}_{1j1} - \tau^{hk}_{12} - \tau^{hjk}_{1j2} .$$

From the preceding results we have

j	$\tau^{hk}_{11} + \tau^{hjk}_{1j1} - \tau^{hk}_{12} - \tau^{hjk}_{1j2}$
1	1.9235
2	1.5136
3	1.0615
4	1.2576
5	1.5661
6	0.8205
7	1.3458

The association between father's occupational level and post high-school status is dependent on the sex, that is,

$$\ln \frac{x^*_c(hi11)}{x^*_c(hi14)} - \ln \frac{x^*_c(hi71)}{x^*_c(hi74)} = \tau^{hk}_{11} + \tau^{hjk}_{h11} ,$$

$$\ln \frac{x^*_c(hi21)}{x^*_c(hi24)} - \ln \frac{x^*_c(hi71)}{x^*_c(hi74)} = \tau^{jk}_{21} + \tau^{hjk}_{h21} ,$$

etc.

$$\ln \frac{x^*_c(hi12)}{x^*_c(hi14)} - \ln \frac{x^*_c(hi72)}{x^*_c(hi74)} = \tau^{jk}_{12} + \tau^{hjk}_{h12} ,$$

$$\ln \frac{x^*_c(hi22)}{x^*_c(hi24)} - \ln \frac{x^*_c(hi72)}{x^*_c(hi74)} = \tau^{jk}_{22} + \tau^{hjk}_{h22} ,$$

etc.

$$\ell n \frac{x_c^*(hi13)}{x_c^*(hi14)} - \ell n \frac{x_c^*(hi73)}{x_c^*(hi74)} = \tau_{13}^{jk} + \tau_{h13}^{hjk} ,$$

$$\ell n \frac{x_c^*(hi23)}{x_c^*(hi24)} - \ell n \frac{x_c^*(hi73)}{x_c^*(hi74)} = \tau_{23}^{jk} + \tau_{h23}^{hjk} ,$$

etc.

A tabulation of these associations is

	h=1			h=2		
j	k=1	k=2	k=3	k=1	k=2	k=3
1	1.5094	-0.3149	0.4556	2.2731	0.9905	0.8593
2	0.5992	0.1804	0.3229	1.2332	0.9822	0.6872
3	-0.6914	-0.4148	-0.3376	0.4009	0.3932	0.6333
4	0.2796	0.1300	0.0526	1.1259	0.8881	0.6099
5	-0.0208	-0.4610	-0.0254	0.6194	0.3995	0.5254
6	-0.7908	-0.3731	-0.2408	-0.0321	-0.1397	0.1989

In particular, the association between father's occupational levels 1 and 2 and post high-school status of collegiate and noncollegiate school, for boys, is

$$\ell n \frac{x_c^*(1i11)}{x_c^*(1i12)} - \ell n \frac{x_c^*(1i21)}{x_c^*(1i22)} = \tau_{11}^{jk} + \tau_{111}^{hjk} - \tau_{12}^{jk} - \tau_{112}^{hjk} - \tau_{21}^{jk} - \tau_{121}^{hjk} + \tau_{22}^{jk} + \tau_{122}^{hjk} .$$

We shall not pursue this matter any further here. The reader should be able to examine any particular associations of interest.

Table 1

Frequency for each High-School Rank x Post High-School Status x Sex x Father's Occupational Level Combination

x(hijk)

			High-School Rank											
			Lowest Third				Middle Third				Upper Third			
Post High-School Status*			1	2	3	4	1	2	3	4	1	2	3	4
Sex (1)	Father's Occupational Levels	1	87	3	17	105	216	4	14	118	256	2	10	53
		2	72	6	18	209	159	14	28	22[illegible]	176	3	22	95
		3	52	17	14	541	119	13	44	578	119	10	33	257
		4	88	9	14	328	158	15	36	304	144	12	20	115
		5	32	1	12	124	43	5	7	119	42	2	7	56
		6	14	2	5	148	24	6	15	131	24	2	4	61
		7	20	3	4	109	41	5	13	88	32	2	4	41
Sex (2)	Father's Occupational Levels	1	53	7	13	76	163	30	28	118	309	17	38	89
		2	36	16	11	111	116	41	53	214	225	49	68	210
		3	52	28	49	521	162	64	129	708	243	79	184	448
		4	48	18	29	191	130	47	62	305	237	57	63	219
		5	12	5	10	101	35	11	37	152	72	20	21	95
		6	9	1	15	130	19	13	22	174	42	10	19	105
		7	3	1	6	98	25	9	15	158	36	14	19	93

*Categories of post high-school status: (1) enrolled in college; (2) enrolled in non-collegiate school; (3) employed full-time; (4) other.

Table 2

Estimated Frequency for each High-School Rank x Post High-School Status x Sex x Father's Occupational Level Combination

$x^{*}_{c}(hijk)$

		High-School Rank											
		Lowest Third				Middle Third				Upper Third			
	Post High-School Status*	1	2	3	4	1	2	3	4	1	2	3	4
Sex (1)	Father's Occupational Levels 1	96.076	2.062	9.106	104.751	214.142	3.964	17.913	115.981	248.782	2.975	13.981	55.269
	2	74.160	6.726	15.853	208.256	160.787	12.579	30.337	224.296	172.053	8.695	21.809	[illegible]3.448
	3	52.918	9.622	21.244	540.216	114.549	17.964	40.588	580.899	122.534	12.414	29.169	254.885
	4	84.275	10.004	18.926	325.787	159.270	16.308	31.570	305.852	146.455	9.687	19.503	115.361
	5	25.645	2.277	7.194	133.882	44.415	3.401	10.997	115.186	46.9[illegible]	2.322	7.808	49.932
	6	13.027	2.727	6.363	146.881	24.402	4.406	10.520	136.671	24.571	2.866	7.117	56.447
	7	20.818	2.871	5.867	106.443	37.619	4.474	9.357	95.549	34.562	2.655	5.776	36.[illegible]08
Sex (2)	Father's Occupational Levels 1	53.675	7.884	11.104	76.337	168.174	21.306	30.708	118.813	303.151	24.810	37.188	87.850
	2	29.353	12.096	14.462	118.093	111.908	39.776	48.665	223.655	235.739	54.129	68.873	193.253
	3	54.868	28.839	58.878	507.420	151.976	68.898	143.943	698.185	250.157	73.263	159.179	471.394
	4	44.660	18.647	22.674	200.023	134.884	48.576	60.442	300.100	235.456	54.778	70.884	214.878
	5	13.289	5.649	10.239	98.774	40.924	15.005	27.967	151.105	64.787	15.346	29.744	98.121
	6	9.223	4.386	9.883	131.508	24.155	9.909	22.846	171.091	36.622	9.705	23.271	106.401
	7	6.054	3.206	5.149	83.592	22.830	10.430	17.139	156.602	35.117	10.364	17.712	98.306

*Categories of post high-school status: (1) enrolled in college (2) enrolled in non-collegiate school; (3) employed full-time; (4) other

Table 2

Estimated Frequency for each High-School Rank x Post High-School Status x Sex x Father's Occupational Level Combination

$x_c^*(hijk)$

		High-School Rank											
		Lowest Third				Middle Third				Upper Third			
Post High-School Status*		1	2	3	4	1	2	3	4	1	2	3	4
Sex (1) — Father's Occupational Levels	1	96.076	2.062	9.106	104.751	214.142	3.964	17.913	115.981	248.782	2.975	13.981	55.269
	2	74.160	6.726	15.853	208.256	160.787	12.579	30.337	224.296	172.053	8.695	21.809	93.448
	3	52.918	9.622	21.244	540.216	114.549	17.964	40.578	580.899	122.534	12.414	29.169	254.885
	4	84.275	10.004	18.926	325.787	159.270	16.308	31.570	305.852	146.455	9.637	19.503	115.361
	5	25.645	2.277	7.194	133.882	44.415	3.401	10.997	115.186	46.939	2.322	7.808	49.932
	6	13.027	2.727	6.363	146.881	24.402	4.406	10.520	136.671	24.571	2.866	7.117	56.447
	7	20.818	2.871	5.867	106.443	37.619	4.474	9.357	95.549	34.562	2.655	5.776	36.008
Sex (2) — Father's Occupational Levels	1	53.675	7.884	11.104	76.337	168.174	21.306	30.708	118.813	303.151	24.810	37.188	87.850
	2	29.353	12.096	14.462	118.093	111.908	39.776	48.665	223.655	235.739	54.129	68.873	193.253
	3	54.868	28.839	58.878	507.420	151.976	68.898	143.943	698.185	250.157	73.263	159.179	471.394
	4	44.660	18.647	22.674	200.023	134.884	48.576	60.442	300.100	235.456	54.778	70.884	214.878
	5	13.289	5.649	10.289	98.774	40.924	15.005	27.967	151.105	64.787	15.346	29.744	98.121
	6	9.223	4.386	9.883	131.508	24.155	9.909	22.846	171.091	36.622	9.705	23.271	106.401
	7	6.054	3.206	5.149	83.592	22.830	10.430	17.139	156.602	35.117	10.364	17.712	93.506

*Categories of post high-school status: (1) enrolled in college (2) enrolled in non-collegiate school; (3) employed full-time; (4) other

Table 3

Analysis of Information

Component due to	Information	D.F.
a) x(hij•), x(•••k)	$2I(x.x_a^*) = 2824.434$	123
b) x(hij•), x(h••k), x(•i•k), x(••jk)	$2I(x_b^*:x_a^*) = 2672.724$	27
	$2I(x:x_b^*) = 151.710$	96
c) x(hij•), x(•i•k), x(h•jk)	$2I(x_c^*:x_b^*) = 52.850$	18
	$2I(x:x_c^*) = 98.860$	78

Example 5. Coronary heart disease risk. This example illustrates the analysis of a three-way 2x4x4 contingency table. It illustrates the test of equality of certain parameters in the model of no second-order interaction, both by computing the estimate implied by the hypothesized relation among some of the parameters, and also by computing the appropriate quadratic approximation.

Example

Coronary Heart Disease Risk

We are indebted to Professor S. Greenhouse and J. Cornfield (1962) for calling our attention to this set of data.

In this example we analyze data from a 3-way, R x S x T, table resulting from a coronary heart disease study. We denote the observed values by f(ijk), where

Characteristic		Index	1	2	3	4
Coronary heart disease	R	i	yes	no		
Serum cholesterol, mg/100 cc	S	j	< 200	200-219	220-259	260 +
Blood pressure, mm Hg	T	k	< 127	127-146	147-166	167 +

We ask the reader's indulgence for not using the notation used elsewhere in this report, that is, x(ijk), $x^*_a(ijk)$, etc.

The complete 2 x 4 x 4 table is given in Fig. 1. A preliminary analysis is given in the analysis of information table shown in Fig. 2, where the various sets of marginal constraints and the corresponding information values and degrees of freedom are listed. Interaction hypotheses corresponding to sets of marginal constraints in the table are

$$H_a:\quad p(ijk) = p(i\cdot\cdot)p(\cdot jk)$$

H_b : $p(ijk) = \dfrac{p(ij\cdot)p(\cdot jk)}{p(\cdot j\cdot)}$

H_2 : no second-order interaction.

The effects due to addition of each of the three 2-way marginal tables are shown immediately above these interactions. We note that both the information values and the degrees of freedom are additive.

This analysis indicated that a fit to this set of data could be made adequately using as explanatory variables the marginal cell frequencies of three marginal tables of dimensions 2 x 4, 2 x 4, and 4 x 4. The hypothesis tested was that of no second-order interaction in the sense of Bartlett [1935], as discussed by Ku et al. (1971). We start with H_a because our first concern is whether the incidence of coronary heart disease is homogeneous over the factors serum cholesterol and blood pressure. Thus considering $2I(f:f_a)$ in Fig. 2 as the total "unexplained variation" we may set up the summary analysis of information table in Fig. 3.

The interpretation of the no second-order interaction hypothesis is:

a. The association between blood pressure and heart disease is the same for different levels of cholesterol,

b. The association between cholesterol level and heart disease is the same for different levels of blood pressure,

c. The association between cholesterol level and blood pressure is the same for subjects with and without heart disease. For the estimate f_2^* under the model of no second-order interaction the log-odds (logit) of the estimated incidence of coronary heart disease is a linear additive function of an average effect, an effect due to cholesterol and an effect due to blood pressure, i.e.,

$$\ln \frac{f_2^*(1jk)}{f_2^*(2jk)} = \tau_1^i + \tau_{1j}^{ij} + \tau_{1k}^{ik} .$$

Values of f_2^* are shown in Fig. 4 and the design matrix in Fig. 5. We note that there are 22 parameters, in addition to τ_0, to be estimated from the f_2^* values. A complete model would include nine additional parameters, which, under the no second-order interaction hypothesis, are equal to zero, i.e.,

$$\tau_{111}^{ijk} = \tau_{112}^{ijk} = \tau_{113}^{ijk} = 0 ,$$

$$\tau_{121}^{ijk} = \tau_{122}^{ijk} = \tau_{123}^{ijk} = 0 ,$$

$$\tau_{131}^{ijk} = \tau_{132}^{ijk} = \tau_{133}^{ijk} = 0 .$$

We note that the number of parameters in the complete model is 23 + 9 = 32, that is, the number of cells.

The computation of the τ parameter estimates is straightforward, e.g.,

$$\tau_1^i = \ln \frac{f_2^*(144)}{f_2^*(244)} = - 0.9374 ,$$

etc. The values of the τ's are listed in Fig. 6. For simplicity we use τ with no further diacritical marking.

When the "dependent" variable or response variable is dichotomous, odds and log-odds have long been used as indices indicative of risk. The estimated log-odds,

$$\ell n \frac{f_2^*(1jk)}{f_2^*(2jk)} = \tau_1^i + \tau_{1j}^{ij} + \tau_{1k}^{ik} ,$$

and the estimated odds,

$$\frac{f_2^*(1jk)}{f_2^*(2jk)}$$

are given in Fig. 7.

From the design matrix or the representation of the log-odds we can compute the difference in log-odds of risk of heart disease for change in blood pressure and constant cholesterol concentration in terms of the τ parameters, e.g.,

$$\ell n \frac{f_2^*(1j2)}{f_2^*(2j2)} - \ell n \frac{f_2^*(1j1)}{f_2^*(2j1)} = \ell n \frac{f_2^*(112)}{f_2^*(212)} - \ell n \frac{f_2^*(111)}{f_2^*(211)}$$

$$= \tau_{12}^{ik} - \tau_{11}^{ik} = -0.0415 .$$

Similarly,

$$\ell n \frac{f_2^*(1j3)}{f_2^*(2j3)} - \ell n \frac{f_2^*(1j2)}{f_2^*(2j2)} = 0.5738 ,$$

$$\ell n \frac{f_2^*(1j4)}{f_2^*(2j4)} - \ell n \frac{f_2^*(1j3)}{f_2^*(2j3)} = 0.6681 .$$

The differences in log-odds for change in cholesterol level and constant blood pressure are:

$$\ln \frac{f_2^*(12k)}{f_2^*(22k)} - \ln \frac{f_2^*(11k)}{f_2^*(21k)} = -0.2079 \, ,$$

$$\ln \frac{f_2^*(13k)}{f_2^*(23k)} - \ln \frac{f_2^*(12k)}{f_2^*(22k)} = 0.7702 \, ,$$

$$\ln \frac{f_2^*(14k)}{f_2^*(24k)} - \ln \frac{f_2^*(13k)}{f_3^*(23k)} = 0.7818 \, .$$

The differences in log-odds for change in cholesterol level and change in blood pressure are

$$\ln \frac{f_2^*(122)}{f_2^*(222)} - \ln \frac{f_2^*(111)}{f_2^*(211)} = -0.2494 \, ,$$

$$\ln \frac{f_2^*(133)}{f_2^*(233)} - \ln \frac{f_2^*(122)}{f_2^*(222)} = 1.3440 \, ,$$

$$\ln \frac{f_2^*(144)}{f_2^*(244)} - \ln \frac{f_2^*(133)}{f_2^*(233)} = 1.4499 \, .$$

In view of the negative values of the changes in log-odds represented by $\tau_{12}^{ik} - \tau_{11}^{ik}$, $\tau_{12}^{ij} - \tau_{11}^{ij}$, we may wish to check the hypothesis that

$$\tau_{11}^{ij} = \tau_{12}^{ij} \, ; \quad \tau_{11}^{ik} = \tau_{12}^{ik} \, ,$$

which would imply that the risk does not begin to manifest itself significantly until the cholestrol level and blood pressure exceed some minimum level, that is, a threshold effect. Let

$$Z_1 = \tau_{12}^{ij} - \tau_{11}^{ij} = -0.2079$$

$$Z_2 = \tau_{12}^{ik} - \tau_{11}^{ik} = -0.0415 .$$

The variance-covariance matrix of the taus for f_2^* is obtained as follows (a weighted version of Kullback (1959, p. 217):

Compute $\underset{\sim}{S} = \underset{\sim}{T}'\underset{\sim}{D}\underset{\sim}{T}$ where $\underset{\sim}{T}$ is the 32 x 23 design matrix for the log-linear representation of f_2^* in Fig. 5 , and $\underset{\sim}{D}$ is a diagonal matrix whose entries are the values of f_2^* in the order of the rows of the design matrix.

Partition the matrix $\underset{\sim}{S}$ as

$$\begin{pmatrix} \underset{\sim}{S}_{11} & \underset{\sim}{S}_{12} \\ \underset{\sim}{S}_{21} & \underset{\sim}{S}_{22} \end{pmatrix}$$

where $\underset{\sim}{S}_{11}$ is 1 x 1 .

Then the variance-covariance matrix of the taus is

$$\left(\underset{\sim}{S}_{22} - \underset{\sim}{S}_{21}\underset{\sim}{S}_{11}^{-1}\underset{\sim}{S}_{12}\right)^{-1} \text{ or } \underset{\sim}{S}_{22\cdot 1}^{-1} .$$

The covariance matrix of Z_1, Z_2 is found to be:

$$a_{11} = \sigma^{8,8} + \sigma^{9,9} - 2\sigma^{8,9} = 0.2175$$

$$a_{12} = a_{21} = \sigma^{8,11} - \sigma^{9,11} - \sigma^{8,12} + \sigma^{9,12} = -0.0013$$

$$a_{22} = \sigma^{11,11} + \sigma^{12,12} - 2\sigma^{11,12} = 0.0922 \ .$$

We found $$\underset{\sim}{A}^{-1} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} = \begin{pmatrix} 4.5981 & 0.0648 \\ 0.0648 & 10.8469 \end{pmatrix},$$

$$X^2 = (Z_1, Z_2)\ \underset{\sim}{A}^{-1} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = 0.2185$$

does not exceed the upper 5% critical value of a chi-squared variate with 2 degrees of freedom.

For this particular hypothesis, we may alternatively revise the design matrix by combining the columns τ_{11}^{ij} with τ_{12}^{ij}, and τ_{11}^{ik} with τ_{12}^{ik}, and use the iterative procedure suggested by Gokhale [1972], Kullback [1973] for "unusual marginal totals" to obtain the estimated cell frequencies. The resulting estimates f_d^* are given in Fig. 8. In Fig. 9 are listed the log-odds

$$\ln \frac{f_d^*(1jk)}{f_d^*(2jk)}$$

and the odds $f_d^*(1jk)/f_d^*(2jk)$. The associated analysis of information table is shown in Fig. 10. Note that $2I(f_2^*:f_d^*)$ is a test of the hypothesis that $\tau_{11}^{ij} = \tau_{12}^{ij}$, $\tau_{11}^{ik} = \tau_{12}^{ik}$ and is approximated by the test previously given as a quadratic chi-squared variate.

		j: Serum cholesterol, mg/100 cc	k: blood pressure, mm Hg 1 < 127	2 127-146	3 147-166	4 167 +	Total
	1	< 200	2	3	3	4	12
CHD	2	200-219	3	2	0	3	8
i = 1	3	220-259	8	11	6	6	31
	4	260 +	7	12	11	11	41
		j total	20	28	20	24	92
	1	< 200	117	121	47	22	307
NCHD	2	200-219	85	98	43	20	246
i = 2	3	220-259	119	209	68	43	439
	4	260 +	67	99	46	33	245
		j total	388	527	204	118	1237
		Total	408	555	224	142	1329

Figure 1. Coronary Heart Disease Risk

Component due to	Information	D.F.
1) $f(i\cdot\cdot)$, $f(\cdot j\cdot)$, $f(\cdot\cdot k)$	$2I(f:f_1^*) = 83.149$	24
a) $f(i\cdot\cdot)$, $f(\cdot jk)$		
ST effect	$2I(f_a^*:f_1^*) = 24.423$	9
Independence R x ST	$2I(f:f_a^*) = 58.726$	15
b) $f(\cdot jk)$, $f(ij\cdot)$		
RS effect/ST	$2I(f_b^*:f_a^*) = 31.921$	3
Conditional independence		
R x T/S	$2I(f:f_b^*) = 26.805$	12
2) $f(\cdot jk)$, $f(ij\cdot)$, $f(i\cdot k)$		
RT effect/ST, RS	$2I(f_2^*:f_b^*) = 18.730$	3
Second-order interaction	$2I(f:f_2^*) = 8.075$	9

Figure 2. Analysis of Information - Coronary Heart Disease Risk Data

Component due to		Information	D.F.
f(i··), f(·jk),	Total	$2I(f:f_a^*) = 58.726$	15
f(·jk), f(ij·),	Cholesterol effect	$2I(f_b^*:f_a^*) = 31.921$	3
f(·jk), f(ij·), f(i·k),	Blood Pressure effect given Cholesterol	$2I(f_2^*:f_b^*) = 18.730$	3
Second-order interaction	(Residual)	$2I(f:f_2^*) = 8.075$	9

Figure 3. Analysis of Information

	j: Serum cholesterol, mg/100 cc	k: blood pressure, mm Hg 1 < 127	2 127-146	3 147-166	4 167 +	Total
CHD i = 1	1 < 200	3.550	3.553	2.488	2.409	12.000
	2 200-219	2.144	2.340	1.754	1.762	8.000
	3 220-259	6.501	10.827	6.227	7.446	31.001
	4 260 +	7.805	11.287	9.531	12.382	40.998
	Total	20.000	28.000	20.000	23.999	91.999
NHCD i = 2	1 < 200	115.450	120.447	47.512	23.591	307.000
	2 200-219	85.856	97.660	41.246	21.238	246.000
	3 220-259	120.499	209.173	67.773	41.554	438.999
	4 260 +	66.196	99.720	47.469	31.617	245.002
	Total	388.001	527.000	204.000	118.000	1237.001
	TOTAL	408.001	555.000	224.000	141.999	1329.000

Figure 4. Estimated Cell Frequencies under No Second-Order Interaction Hypothesis, f_2^*, Coronary Heart Disease Risk.

i	j	k	τ_0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
				i 1	j 1	j 2	j 3	k 1	k 2	k 3	ij 11	ij 12	ij 13	ik 11	ik 12	ik 13	jk 11	jk 12	jk 13	jk 21	jk 22	jk 23	jk 31	jk 32	jk 33
1	1	1	1	1	1			1			1			1			1								
1	1	2	1	1	1				1		1				1			1							
1	1	3	1	1	1					1	1					1			1						
1	1	4	1	1	1						1														
1	2	1	1	1		1		1				1		1						1					
1	2	2	1	1		1			1			1			1						1				
1	2	3	1	1		1				1		1				1						1			
1	2	4	1	1		1						1													
1	3	1	1	1			1	1					1	1									1		
1	3	2	1	1			1		1				1		1									1	
1	3	3	1	1			1			1			1			1									1
1	3	4	1	1			1						1												
1	4	1	1	1				1						1											
1	4	2	1	1					1						1										
1	4	3	1	1						1						1									
1	4	4	1	1																					
2	1	1	1		1			1									1								
2	1	2	1		1				1									1							
2	1	3	1		1					1									1						
2	1	4	1		1																				
2	2	1	1			1		1												1					
2	2	2	1			1			1												1				
2	2	3	1			1				1												1			
2	2	4	1			1																			
2	3	1	1				1	1															1		
2	3	2	1				1		1															1	
2	3	3	1				1			1															1
2	3	4	1				1																		
2	4	1	1					1																	
2	4	2	1						1																
2	4	3	1							1															
2	4	4	1																						

log-linear representation

	τ_0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
f^*_1	✓	✓	✓	✓	✓	✓	✓	✓															
f^*_a	✓	✓	✓	✓	✓	✓	✓	✓							✓	✓	✓	✓	✓	✓	✓	✓	✓
f^*_b	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓	✓	✓	✓	✓	✓	✓	✓	✓
f^*_2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Figure 5. Design Matrix - Coronary Heart Disease Risk

$\tau_1^i = -0.9374$	$\tau_{11}^{ij} = -1.3441$	$\tau_{11}^{jk} = 0.8491$
$\tau_1^j = -0.2929$	$\tau_{12}^{ij} = -1.5520$	$\tau_{12}^{jk} = 0.4817$
$\tau_2^j = -0.3979$	$\tau_{13}^{ij} = -0.7818$	$\tau_{13}^{jk} = 0.2938$
$\tau_3^j = 0.2733$	$\tau_{11}^{ik} = -1.2004$	$\tau_{21}^{jk} = 0.6580$
$\tau_1^k = 0.7389$	$\tau_{12}^{ik} = -1.2419$	$\tau_{22}^{jk} = 0.3770$
$\tau_2^k = 1.1481$	$\tau_{13}^{ik} = -0.6681$	$\tau_{23}^{jk} = 0.2574$
$\tau_3^k = 0.4064$		$\tau_{31}^{jk} = 0.3527$
		$\tau_{32}^{jk} = 0.4675$
		$\tau_{33}^{jk} = 0.0828$

NOTE: Any tau parameter corresponding to a subscript $i = 2$, and/or $j = 4$, and/or $k = 4$ is zero.

Figure 6. Values of Estimates of Tau Parameters

	k = 1	k = 2	k = 3	k = 4
j = 1	- 3.482	- 3.523	- 2.950	- 2.281
	.0307	.0295	.0523	.1022
j = 2	- 3.690	- 3.731	- 3.158	- 2.489
	.0250	.0240	.0245	.0830
j = 3	- 2.920	- 2.961	- 2.387	- 1.719
	.0539	.0518	.0919	.1792
j = 4	- 2.138	- 2.179	- 1.605	- 0.937
	0.1179	0.1132	0.2009	0.3918

Figure 7. Log-odds and Odds

Entries are log-odds $\ln \dfrac{f_2^*(1jk)}{f_2^*(2jk)}$

and odds $\dfrac{f_2^*(1jk)}{f_2^*(2jk)}$

	j: Serum cholesterol, mg/100 cc		k: blood pressure, mm Hg 1 < 127	2 127-146	3 147-166	4 167 +	Total
CHD $i = 1$	1	< 200	3.189	3.323	2.289	2.225	11.026
	2	200-219	2.358	2.680	1.969	1.968	8.975
	3	220-259	6.350	11.000	6.217	7.437	31.001
	4	260 +	7.640	11.460	9.525	12.374	40.999
			19.537	28.463	20.000	24.001	92.001
NCHD $i = 2$	1	< 200	115.811	120.677	47.711	23.775	307.974
	2	200-219	85.692	97.320	41.031	21.032	245.025
	3	220-259	120.650	209.000	67.783	41.566	438.999
	4	260 +	66.360	99.539	47.475	31.626	245.000
		Total	388.463	526.536	204.000	117.999	1236.998
		TOTAL	408.000	554.999	224.000	142.000	1328.999

Figure 8. Estimate under $\tau_{11}^{ij} = \tau_{12}^{ij}$, $\tau_{11}^{ik} = \tau_{12}^{ik}$, f_d^*,
Coronary Heart Disease Risk

Serum cholesterol		Blood pressure k = 1	2	3	4
	j = 1	- 3.592	- 3.592	- 3.037	- 2.369
		0.0275	0.0275	0.0480	0.0936
	j = 2	- 3.592	- 3.592	- 3.037	- 2.369
		0.0275	0.0275	0.0480	0.0936
	j = 3	- 2.944	- 2.944	- 2.389	- 1.721
		0.0526	0.0526	0.0917	0.1788
	j = 4	- 2.162	- 2.162	- 1.606	- 0.938
		0.1151	0.1151	0.2006	0.3912

Figure 9. The log-odds $\ln f^*_d(1jk)/f^*_d(2jk)$, and the odds $f^*_d(1jk)/f^*_d(2jk)$.

$$\ln f^*_d(1jk)/f^*_d(2jk) = \tau^{i}_{1} + \tau^{ij}_{1j} + \tau^{ik}_{1k}$$

$$\tau^{i}_{1} = - 0.9384$$

$$\tau^{ij}_{11} = \tau^{ij}_{12} = - 1.4306 \qquad \tau^{ik}_{11} = \tau^{ik}_{12} = - 1.2232$$

$$\tau^{ij}_{13} = - 0.7828 \qquad \tau^{ik}_{13} = - 0.6678$$

Component due to	Information	D.F.
a) $f(i\cdot\cdot)\ f(\cdot jk)$	$2I(f:f_a^*) = 58.726$	15
d) $f(\cdot jk)$, $f(ij\cdot)$, $j = 3,4$ $f(i\cdot k)$, $k = 3,4$	$2I(f_d^*:f_a^*) = 50.429$	4
$f(i1\cdot) + f(i2\cdot)$; $f(i\cdot 1) + f(i\cdot 2)$	$2I(f:f_d^*) = 8.297$	11
2) $f(\cdot jk)$, $f(ij\cdot)$, $f(i\cdot k)$	$2I(f_2^*:f_d^*) = 0.222$	2
	$2I(f:f_2^*) = 8.075$	9

Figure 10. Analysis of Information

Example 6. Hospital data. This example illustrates the analysis of a pair of related three-way 2x2x2 contingency tables. In particular it illustrates the procedure to obtain an estimate satisfying certain observed marginal restraints and having certain of the tau parameters predetermined, that is, the "inheritance" of certain parameters. It also mentions that the T-functions of the two-way marginals are the products of the T-functions of the related one-way marginals.

Example

Hospital Data

The data used are from the field of hospital administration and relate to the matter of innovation in hospitals. We begin with the assumption that the use of electronic data processing (EDP) in hospitals in the late 1960's was innovative. This assumption is substantiated by a variety of surveys of the use of EDP in hospitals, Hammon et al. (1972). On this basis the data in a survey of hospitals using EDP conducted by Herner and Co. were combined with data from the Guide Issue of Hospitals for the same period so that a file of records reflecting characteristics of hospitals and levels at which EDP was used by these hospitals was created. The hospitals in this survey were selected by stratified sampling. The stratification (fixed variable) was on the basis of hospital size. All hospitals in the large-size category (200 or more beds) were included in the survey and a ten percent sample was taken of those in the small size category. The data from these files were tabulated and arranged in multiway contingency tables. The analysis of the tables for the large and small hospitals will be described here and interrelated. See Kullback and Reeves (1974).

On the basis of these analyses we conclude that there is a distinct relation of innovation on location and length of stay with a common factor for large and small hospitals. The association (measured by the logarithm of the cross-product ratio) between use of EDP and length of stay is the same for the large and small hospitals. The log-odds (logit) of use of EDP in descending order of magnitude within the large hospitals and within the small hospitals are parallel in terms of the combinations of the factors location and length of stay. The usage of EDP is generally greater in the large hospitals than in the small hospitals except that the best log-odds for the small hospitals is greater than the poorest log-odds for the large hospitals.

In a study to identify characteristics which distinguish hospitals which use EDP from those which do not, that is, to identify characteristics which are significantly associated with use of EDP, data on 1176 hospitals, 923 large and 253 small, were collected with respect to use, location, and length of stay. The data appear in the two three-way 2×2×2 contingency tables 1 and 2. In order to determine the relation among the free variables use, location and length of stay, indexed by size of hospital, and interactions that may exist among these characteristics it seems intuitively clear that an analysis based only on two-way tables would not suffice.

We shall denote the occurrences in the observed tables 1 and 2 respectively by x(ijk), y(ijk) with

i=1, user; i=2, non-user

j=1, urban; j=2, rural

k=1, short; k=2, long.

The proposed procedure provides estimates for the original data analogous to a regression procedure using sets of observed marginals as explanatory variables and we shall try to find an estimate which does not differ significantly from the observed data. The set of acceptable estimates will indicate the nature of the significant interactions for which we can compute numerical measures.

As a first step in the analysis we shall find "smoothed" estimates of the original data. We shall do this for the large hospitals also even though the data for all large hospitals was collected. We examine the minimum discrimination information estimates obtained by a convergent iterative algorithm starting with a uniform table and successively adjusting for sets of observed marginals. It turns out that the sets of two-way marginals are best and the resultant estimates provide a satisfactory fit. The estimated tables have the same two-way and also the same one-way marginals as the original tables . These estimates which we denote by $x_2^*(ijk)$, $y_2^*(ijk)$ respectively for the large and small hospitals are given in tables 3 and 4 and imply no second-order (three-factor) interaction. Note that the estimate for the observed $y(122)=0$ is $y_2^*(122)=0.137$.

The estimates are given analytically by the log-linear representation of an exponential family

$$\ell n\frac{x_2^*(ijk)}{n\pi(ijk)} = L+\tau_1 T_1(ijk)+\tau_2 T_2(ijk)+\tau_3 T_3(ijk) + \tau_4 T_4(ijk)+\tau_5 T_5(ijk)+\tau_6 T_6(ijk) \quad (1)$$

where $n=\sum\sum\sum x(ijk)$, $\pi(ijk)=1/2\times 2\times 2$, L is a normalizing constant, the taus are main-effect and interaction parameters, and the $T(ijk)$ are a set of linearly independent random variables, in this case the indicator functions of the respective marginals. A similar representation holds for $y_2^*(ijk)$. The log-linear representations are shown graphically in Fig. 1 . The values in the various columns of Fig. 1, zeros or ones, are the values of the respective functions $T(ijk)$. Note that

$$T_4(ijk)=T_1(ijk)T_2(ijk),T_5(ijk)=T_1(ijk)T_3(ijk),$$

$$T_6(ijk)=T_2(ijk)T_3(ijk).$$

To test the goodness-of-fit of the estimates we compute the statistics [3,4]

$$2I(x:x_2^*)=2\sum\sum\sum x(ijk)\ell n(x(ijk)/x_2^*(ijk))=0.481, \text{ 1 D.F.}$$

$$2I(y:y_2^*)=2\sum\sum\sum y(ijk)\ell n(y(ijk)/y_2^*(ijk))=0.294, \text{ 1 D.F.}$$

Since the statistics are asymptotically distributed as χ^2 we conclude that the "smoothed" values x_2^*,y_2^* are good estimates and we shall use them in our subsequent analysis.

From the log-linear representation (1) or the graphical presentation in Fig. 1, we find that the log-odds or logits of the use of EDP for large hospitals is given by the parametric representation

$$\begin{aligned}
\ell n\frac{x_2^*(111)}{x_2^*(211)} &= \tau_1 + \tau_4 + \tau_5 \\
\ell n\frac{x_2^*(112)}{x_2^*(212)} &= \tau_1 + \tau_4 \\
\ell n\frac{x_2^*(121)}{x_2^*(221)} &= \tau_1 \qquad + \tau_5 \\
\ell n\frac{x_2^*(122)}{x_2^*(222)} &= \tau_1
\end{aligned} \tag{2}$$

where the values of the parameters for the estimate $x_2^*(ijk)$ are found to be

$$\tau_1 = -1.4842,\ \tau_4 = 0.5113,\ \tau_5 = 1.5103.$$

From (2) we also see that for the large hospitals

$$\tau_4 = \ell n\frac{x_2^*(111)x_2^*(221)}{x_2^*(211)x_2^*(121)} = \ell n\frac{x_2^*(112)x_2^*(222)}{x_2^*(212)x_2^*(122)} = 0.5113,$$

that is, the association between usage and location for either short or long stay. Similarly

$$\tau_5 = \ell n\frac{x_2^*(111)x_2^*(212)}{x_2^*(211)x_2^*(112)} = \ell n\frac{x_2^*(121)x_2^*(222)}{x_2^*(221)x_2^*(122)} = 1.5103,$$

that is, the association between usage and stay for for either urban or rural location.

For the small hospitals the log-odds or logits are

$$\ell n\frac{y_2^*(111)}{y_2^*(211)} = \tau_1 + \tau_4 + \tau_5$$

$$\ell n\frac{y_2^*(112)}{y_2^*(212)} = \tau_1 + \tau_4$$

$$\ell n\frac{y_2^*(121)}{y_2^*(221)} = \tau_1 \qquad + \tau_5$$

$$\ell n\frac{y_2^*(122)}{y_2^*(222)} = \tau_1$$

where the values of the parameters for the estimate $y_2^*(ijk)$ are found to be

$$\tau_1 = -3.3357, \ \tau_4 = 1.3088, \ \tau_5 = 0.9836.$$

For the small hospitals we also have

$$\tau_4 = \ell n \frac{y_2^*(111)y_2^*(221)}{y_2^*(211)y_2^*(121)} = \ell n \frac{y_2^*(112)y_2^*(222)}{y_2^*(212)y_2^*(122)} = 1.3088,$$

that is, the association between usage and location for either short or long stay. Similarly

$$\tau_5 = \ell n \frac{y_2^*(111)y_2^*(212)}{y_2^*(211)y_2^*(112)} = \ell n \frac{y_2^*(121)y_2^*(222)}{y_2^*(221)y_2^*(122)} = 0.9836,$$

that is, the association between usage and stay for either urban or rural locations.

Since the data for the large hospitals reflect observations over all such hospitals, it will be of interest to determine whether there exists a suitable estimate for the small hospitals, other than $y_2^*(ijk)$, which will have some of its interactions (associations) the same as the corresponding values for the large hospitals. This can be accomplished by using the iterative algorithm fitting various subsets of marginals of $y_2^*(ijk)$ (or the original $y(ijk)$) but starting with a distribution which has the same tau parameters as $x_2^*(ijk)$. The tau parameters of $x_2^*(ijk)$ not affected by the iterative fitting procedure will be "inherited" by the resultant estimate. We shall use the table $v(ijk)=(253/923)x_2^*(ijk)$ which has the same tau parameters as the $x_2^*(ijk)$ table with total adjusted to be the same as the observed total of small hospitals.

We summarize the procedure: starting the iterative fitting algorithm with $v(ijk)$ (recall that $y(ijk)$ and $y_2^*(ijk)$ have the same two-way and one-way marginals)

	Marginals fitted	Estimate	Tau parameters "inherited" from v(ijk)
a)	$y(i.k), y(.jk)$	$u_a^*(ijk)$	τ_4
b)	$y(ij.), y(.jk)$	$u_b^*(ijk)$	τ_5
c)	$y(ij.), y(i.k)$	$u_c^*(ijk)$	τ_6
d)	$y(.jk), y(i..)$	$u_d^*(ijk)$	τ_4, τ_5
e)	$y(i.k), y(.j.)$	$u_e^*(ijk)$	τ_4, τ_6
f)	$y(ij.), y(..k)$	$u_f^*(ijk)$	τ_5, τ_6
g)	$y(i..), y(.j.), y(..k)$	$u_g^*(ijk)$	τ_4, τ_5, τ_6

In order to test whether the u^* estimates differ significantly from the y_2^* estimates, that is, whether the interaction parameters in y_2^* differ significantly from the interaction parameters in u^* "inherited" from x_2^*

or v, we compute the statistic

$$2I(y_2^*:u_m^*)=2\sum\sum\sum y_2^*(ijk)\,\ell n(y_2^*(ijk)/u_m^*(ijk))$$

which is asymptotically distributed as χ^2 with 1 D.F. for m=a,b,c, 2 D.F. for m=d,e,f, 3 D.F. for m=g.

The only case which yielded a non-significant value was $u_b^*(ijk)$ for which

$$2I(y_2^*:u_b^*) = 0.408, \quad 1 \text{ D.F.}$$

The values of $u_b^*(ijk)$ are given in Table 5.

The log-linear representation for $u_b^*(ijk)$ in terms of v(ijk) is

$$\ell n\frac{u_b^*(ijk)}{v(ijk)} = L+\tau_1T_1(ijk)+\tau_2T_2(ijk)+\tau_3T_3(ijk) +\tau_4T_4(ijk)+\tau_6T_6(ijk) \tag{3}$$

Note that τ_5 does not appear explicitly in (3). By using the log-linear representation for v(ijk) itself we also get the reparametrization or log-linear representation for $u_b^*(ijk)$ in terms of the uniform distribution

$$\ell n\frac{u_b^*(ijk)}{n\pi(ijk)} = L+\tau_1T_1(ijk)+\tau_2T_2(ijk)+\tau_3T_3(ijk) +\tau_4T_4(ijk)+\tau_5T_5(ijk)+\tau_6T_6(ijk) \tag{4}$$

We remark that the numerical values of the taus in (3) and (4) are not the same.

The log-odds or logits of the use of EDP for small hospitals may now be given by the parametric representation

$$\begin{aligned}
\ell n\frac{u_b^*(111)}{u_b^*(211)} &= \tau_1 + \tau_4 + \tau_5 \\
\ell n\frac{u_b^*(112)}{u_b^*(212)} &= \tau_1 + \tau_4 \\
\ell n\frac{u_b^*(121)}{u_b^*(221)} &= \tau_1 \qquad\; + \tau_5 \\
\ell n\frac{u_b^*(122)}{u_b^*(222)} &= \tau_1
\end{aligned} \tag{5}$$

where the values of the parameters in (5) are

$$\tau_1 = -3.8569,\ \tau_4 = 1.3354,\ \tau_5 = 1.5103.$$

For the small hospitals we now have the associations

$$\tau_4 = \ell n\frac{u_b^*(111)\,u_b^*(221)}{u_b^*(211)\,u_b^*(121)} = \ell n\frac{u_b^*(112)\,u_b^*(222)}{u_b^*(212)\,u_b^*(122)} = 1.3354$$

and

$$\tau_5 = \ell n\frac{u_b^*(111)\,u_b^*(212)}{u_b^*(211)\,u_b^*(112)} = \ell n\frac{u_b^*(121)\,u_b^*(222)}{u_b^*(221)\,u_b^*(122)} = 1.5103.$$

Note that τ_4, the association between usage and location for the small hospitals is still different from that for the large hospitals, but that the association between usage and stay, τ_5, is now the same for both large and small hospitals.

Arranging the log-odds of usage in descending order of magnitude within the large hospitals and within the small hospitals we find

Large hospitals	Factors	Small hospitals
$\ell n \frac{x_2^*(111)}{x_2^*(211)} = 0.5374$	Urban, Short	$\ell n \frac{u_b^*(111)}{u_b^*(211)} = -1.0111$
$\ell n \frac{x_2^*(121)}{x_2^*(221)} = 0.0262$	Rural, Short	$\ell n \frac{u_b^*(121)}{u_b^*(221)} = -2.3466$
$\ell n \frac{x_2^*(112)}{x_2^*(212)} = -0.9729$	Urban, Long	$\ell n \frac{u_b^*(112)}{u_b^*(212)} = -2.5214$
$\ell n \frac{x_2^*(122)}{x^*(222)} = -1.4841$	Rural, Long	$\ell n \frac{u_b^*(122)}{u_b^*(222)} = -3.8569$

Table 1

Large Hospitals x(ijk)

	Urban		Rural		
	Short	Long	Short	Long	
User	376	40	52	15	483
Non-user	217	112	54	57	440
	593	152	106	72	923

Table 2

Small Hospitals y(ijk)

	Urban		Rural		
	Short	Long	Short	Long	
User	28	2	11	0	41
Non-user	80	14	114	4	212
	108	16	125	4	253

Table 3

Large Hospitals $x_2^*(ijk)$

	Urban		Rural		
	Short	Long	Short	Long	
User	374.305	41.694	53.695	13.306	483.000
Non-user	218.693	110.308	52.307	58.692	440.000
	592.998	152.002	106.002	71.998	923.000

Table 4

Small Hospitals $y_2^*(ijk)$

	Urban		Rural		
	Short	Long	Short	Long	
User	28.137	1.863	10.863	0.137	41.000
Non-user	79.863	14.137	114.137	3.863	212.000
	108.000	16.000	125.000	4.000	253.000

Table 5

Small Hospitals $u_b^*(ijk)$

	Urban		Rural		
	Short	Long	Short	Long	
User	28.810	1.190	10.917	0.083	41.000
Non-user	79.190	14.810	114.083	3.917	212.000
	108.000	16.000	125.000	4.000	253.000

Figure 1

Log-linear Representation

i j k	L	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6
1 1 1	1	1	1	1	1	1	1
1 1 2	1	1	1		1		
1 2 1	1	1		1		1	
1 2 2	1	1					
2 1 1	1		1	1			1
2 1 2	1		1				
2 2 1	1			1			
2 2 2	1						

Example 7. Partitioning using OUTLIERS

Outliers are observations in one or more cells of a contingency table which apparently deviate significantly from a fitted model. These outliers may lead one to reject a model which fits the other observations.

In other cases even though a model seems to fit, the outliers contribute much more than reasonable to the measure of deviation between the data and the fitted values of the model. In other words, the outliers make up a large percentage of the "unexplained variation" $2I(x:x^*)$.

A clue to possible outliers is provided by the output of the computer program. In the computer output for each estimate five entries are listed for each cell. The fourth of these is titled OUTLIER and its numerical value provides a lower bound for the decrease in the corresponding $2I(x:x^*)$, if that cell were not included in the fitting procedure. Since the reduction in the degrees of freedom is one for each omitted cell, values of OUTLIER greater than say 3.5 are of interest. The basis for the OUTLIER computation and interpretation follows. Let x_a^* denote the minimum discrimination information estimate subject to certain marginal restraints. Let x_b^* denote the minimum discrimination information estimate subject to the same marginal restraints as x_a^* except that the value $x(\omega_1)$, say, is not included, so that $x_b^*(\omega_1) = x(\omega_1)$. The basic additivity property of the minimum discrimination information statistics states that

$$2I(x:x_a^*) = 2I(x_b^*:x_a^*) + 2I(x:x_b^*)$$

or

$$2I(x:x_a^*) - 2I(x:x_b^*) = 2I(x_b^*:x_a^*) \ .$$

These results are summarized in the Analysis of Information Table.

TABLE

ANALYSIS OF INFORMATION TABLE

Component due to	Information	D.F.
H_a :	$2I(x:x_a^*)$	N_a
H_b : Same as H_a but omitting $x(\omega_1)$	$2I(x_b^*:x_a^*)$	1
	$2I(x:x_b^*)$	$N_b = N_a - 1$

But

$$2I(x_b^*:x_a^*) = 2\left(x_b^*(\omega_1)\ \ln \frac{x_b^*(\omega_1)}{x_a^*(\omega_1)} + \sum_{\Omega-\omega_1} x_b^*(\omega)\ \ln \frac{x_b^*(\omega)}{x_a^*(\omega)}\right)$$

$$(1) \qquad = 2\left(x(\omega_1)\ \ln \frac{x(\omega_1)}{x_a^*(\omega_1)} + \sum_{\Omega-\omega_1} x_b^*(\omega)\ \ln \frac{x_b^*(\omega)}{x_a^*(\omega)}\right) ,$$

and using the convexity property which implies that

$$(2) \qquad \sum_{\Omega-\omega_1} x_b^*(\omega)\ \ell n \frac{x_b^*(\omega)}{x_a^*(\omega)} \geq \left(\sum_{\Omega-\omega_1} x_b^*(\omega)\right) \ell n \frac{\left(\sum_{\Omega-\omega_1} x_b^*(\omega)\right)}{\left(\sum_{\Omega-\omega_1} x_a^*(\omega)\right)}$$

$$= (n - x_b^*(\omega_1))\ \ell n \frac{n - x_b^*(\omega_1)}{n - x_a^*(\omega_1)} \ ,$$

we get from (1) that

$$(3) \qquad 2I(x_b^*:x_a^*) \geq 2\left(x(\omega_1)\ \ell n \frac{x(\omega_1)}{x_a^*(\omega_1)} + \left(\sum_{\Omega-\omega_1} x_b^*(\omega)\right) \ell n \frac{\left(\sum_{\Omega-\omega_1} x_b^*(\omega)\right)}{\left(\sum_{\Omega-\omega_1} x_a^*(\omega)\right)}\right)$$

$$= 2\left(x(\omega_1)\ \ell n \frac{x(\omega_1)}{x_a^*(\omega_1)} + (n - x(\omega_1))\ \ell n \frac{n - x(\omega_1)}{n - x_a^*(\omega_1)}\right).$$

The last value can be computed and is listed as the OUTLIER entry for each cell of the computer output for the estimate x_a^* .

The ratio

$$\frac{2I(x:x_a^*) - 2I(x:x_b^*)}{2I(x:x_a^*)} = \frac{2I(x_b^*:x_a^*)}{2I(x:x_a^*)} \ ,$$

then indicates the percentage of the "unexplained variation" due to the outlier value.

This property is also utilized in the next example. See Ireland (1972) and Ireland and Kullback (1974) for further discussion and application.

Partitioning Using Outliers

We shall use the OUTLIER feature of the CONTAB program to partition a 2x7 table into homogeneous segments.

Table Ia presents data on leukemia cases observed. Denoting the entries in the observed table by x(ij), i=1,2, j=1,2,...,7 we first test whether the incidence of leukemia is homogeneous over the doses by fitting the marginals x(i.), x(.j). The corresponding output is shown in Table II. We observe that large OUTLIER values are associated with values of j=1,2,6,7 and that $2I(x:x^*) = 44.65$, 6D.F.

Since the doses are arranged on a scale we repeat the process omitting the cells corresponding to x(ij), i=1,2, j=6,7. The corresponding output is shown in Table III. We observe that a large OUTLIER value is associated with j=3 and that $2I(x:x^*) = 18.92$, 4 D.F.

We continue the process using the original cells corresponding to j=3,4,5. The computer output is given in Table IV. Now there are no large OUTLIER values and $2I(x:x^*) = 0.09$, 2 D.F. For the original cells with j=6,7 the computer output is given in Table V and again there are no large OUTLIERS and $2I(x:x^*) = 0.37$, 1 D.F. For the original cells with j=1,2 the computer output is given in Table VI and again, there are no large OUTLIERS and $2I(x:x^*) = 0.91$, 1 D.F.

We may summarize in the Analysis of Information Tables.

Component due to	Information		D.F.
cells $j=1,\ldots,7$	$2I(x:x^*)$	$= 44.649$	6
omit cells $j=6,7$	$2I(x_a^*:x^*)$	$= 25.734$	2
cells $j=1,\ldots 5$ [1]	$2I(x:x_a^*)$	$= 18.915$	4
omit cells, $j=1,2$	$2I(x_b^*:x_a^*)$	$= 18.826$	2
cells $j=3,4,5$ [2]	$2I(x:x_b^*)$	$= 0.089$	2
	$2I(x:x^*)$	$= 44.649$	6
omit cells, $j=1,2,3,4,5$	$2I(x_c^*:x^*)$	$= 44.283$	5
cells $j=6,7$ [3]	$2I(x:x_c^*)$	$= 0.366$	1
	$2I(x:x^*)$	$= 44.649$	6
omit cells $j=3,4,5,6,7$	$2I(x_d^*:x^*)$	$= 43.740$	5
cell $j=1,2$ [4]	$2I(x:x_d^*)$	$= 0.909$	1

[1] Note that $x_a^*(ij) = x(i.)x(.j)/n$, $i=1,2, j=1,2,\ldots,5$

$x_a^*(ij) = x(ij)$, $i=1,2$, $j=6,7$

[2] Note that $x_b^*(ij) = x(i.)x(.j)/n$, $i=1,2$, $j=3,4,5$

$x_b^*(ij) = x(ij)$, $i=1,2$, $j=1,2,6,7$

[3] Note that $x_c^*(ij) = x(i.)x(.j)/n$, $i=1,2$, $j=6,7$

$x_c^*(ij) = x(ij)$, $i=1,2$, $j=1,2,3,4,5$

4/ Note that $x_d^*(ij) = x(i.)x(.j)/n$, $i=1,2$, $j=1,2$,

$$x_d^*(ij) = x(ij) \quad , \ i=1,2, \ j=3,4,5,6,7$$

We now define an overall estimate by

$$x_e^*(ij) = x_d^*(ij), \ i=1,2, \ j=1,2$$
$$x_e^*(ij) = x_b^*(ij), \ i=1,2, \ j=3,4,5$$
$$x_e^*(ij) = x_c^*(ij), \ i=1,2, \ j=6,7$$

and we have for the associated min-discrimination information statistic

$$2I(x:x_e^*) = 1.364, \ 4 \text{ D.F.}$$

The values of $x_e^*(ij)$ are given in Table lb.

The data of Table Ia comes from Sugiura, N. and Otake, M. (1973). Approximate distribution of the maximum of c-l X^2 statistics (2x2) derived from 2xC contingency table. Communications in Statistics 1(1), 9-16. We arrived at the same partitioning by a different approach.

Table Ia

Number of Leukemia Cases Observed for the Period
1 Oct 1950 - 30 Sept 1966 Among Hiroshima Male
Survivors for the Extended Life Span Study Sample
at ABCC Aged 15-19 at the Time of Atomic Bomb

Dose (rad)	<5	5~	20~	50~	100~	200~	300+	Total
Leukemia	2	0	3	2	2	2	5	16
Not Leukemia	4601	1161	477	271	243	98	149	7000
Total	4603	1161	480	273	245	100	154	7016

Table Ib.
Values of estimate $x_e^*(ij)$

Dose (rad)	<5	5~	20~	50~	100~	200~	300+	Total
Leukemia	1.597	0.403	3.367	1.915	1.718	2.756	4.244	16
Not Leukemia	4601.398	1160.597	476.633	271.085	243.282	97.244	149.756	6999.995
	4602.995	1161.000	480.000	273.000	245.000	100.000	154.000	7015.995

GWU CONTAB II. . FACTORS: I * J

Table II
Cells 1-7

RESIDUALS: I * J.

			1	2	3	4	5	6	7
1	OBSERVED	1	2.000000	0.000001	3.000000	2.000000	2.000000	2.000000	5.000000
1	PREDICTED	2	10.497149	2.647662	1.094641	0.622577	0.558723	0.228050	0.351197
1	RESIDUAL	3	-8.497149	-2.647661	1.905359	1.377422	1.441277	1.771950	4.648803
1	OUTLIER	4	10.374859	5.284845	2.244734	1.917824	2.217709	5.146365	17.204038
1	LOG RATIO	5	-2.683566	-4.060993	-4.944243	-5.508558	-5.616772	-6.212829	-4.081278
2	OBSERVED	6	4601.000000	1161.000000	477.000000	271.000000	243.000000	98.000000	149.000000
2	PREDICTED	7	4592.500000	1158.357051	478.905273	272.377197	244.441269	99.771927	153.648758
2	RESIDUAL	8	8.500000	2.647949	-1.905273	-1.377197	-1.441269	-1.771927	-4.648758
2	OUTLIER	9	0.038748	0.013617	0.000460	0.005398	0.008474	0.023613	0.140827
2	LOG RATIO	10	3.397510	2.020083	1.136832	0.572518	0.464305	-0.431763	-0.000001

156

GWU CONTAB II SUMMARY. . FACTORS: I * J

SAMPLE SIZE 7016.000000

HYPOTHESIS	NONZERO EFFECTS			SMOOTH	ZERO	2NI(X:X*)	D.F.	PROB	I*	IC*
1	MARGINALS	RESIDUALS								
	I									
	J									
		OUTLIERS								
		1	1							
		1	2							
		1	6							
		1	7	0.0000	0.000001	44.649	6	0.0000	0.00	0.00

GWU CONTAB II. . FACTORS: I * J

Table III

Cells 1-5

RESIDUALS: I * J.

			1	2	3	4	5
1	OBSERVED	1	2.000000	0.000001	3.000000	2.000000	2.000000
1	PREDICTED	2	6.126442	1.545253	0.638864	0.363354	0.326087
1	RESIDUAL	3	-4.126442	-1.545252	2.361135	1.636645	1.673913
1	OUTLIER	4	3.771570	3.082119	4.561651	3.548321	3.911049
1	LOG RATIO	5	-3.687312	-5.064739	-2.947989	-6.512304	-6.620515
2	OBSERVED	6	4601.000000	1161.000000	477.000000	271.000000	243.000000
2	PREDICTED	7	4596.871094	1159.454834	479.361064	272.636475	244.673859
2	RESIDUAL	8	4.128906	1.545166	-2.361084	-1.636475	-1.673859
2	OUTLIER	9	0.004099	0.005678	0.011730	0.004946	0.008045
2	LOG RATIO	10	2.933204	1.555779	0.672527	0.108213	-0.000000

GWU CONTAB II SUMMARY. FACTORS: I * J

SAMPLE SIZE 6762.000000

HYPOTHESIS	NONZERO EFFECTS	SMOOTH	ZERO	2NI(X:X*)	D.F.	PROB	I*	IC*
2	MARGINALS RESIDUALS							
	I							
	J							
	OUTLIERS							
	1 3	0.0000	0.000001	18.915	4	0.0008	0.00	0.00

Table IV

Original j = 3,4,5 only

GWU CONTAB II. . FACTORS: I * J

RESIDUALS: I * J.

			1	2	3
1	OBSERVED	1	3.000000	2.000000	2.000000
1	PREDICTED	2	3.366735	1.914830	1.718437
1	RESIDUAL	3	-0.366735	0.085170	0.281563
1	OUTLIER	4	0.040434	0.004038	0.043937
1	LOG RATIO	5	-4.280277	-4.844592	-4.952806
2	OBSERVED	6	477.000000	271.000000	243.000000
2	PREDICTED	7	476.633301	271.085205	243.281601
2	RESIDUAL	8	0.366699	-0.085205	-0.281601
2	OUTLIER	9	-0.000173	-0.001247	0.000072
2	LOG RATIO	10	0.672526	0.108213	-0.000001

GWU CONTAB II SUMMARY. . FACTORS: I * J

SAMPLE SIZE 998.000000

HYPOTHESIS	NONZERO EFFECTS	SMOOTH	ZERO	2NI(X:X*)	D.F.	PROB	I*	IC*
3	MARGINALS RESIDUALS I J	0.0000	0.000001	0.089	2	0.9566	0.00	0.00

GWU CONTAB II. . FACTORS: I * J

Table V

Original j = 6,7 only

RESIDUALS: I * J.

			1	2
1	OBSERVED	1	2.000000	5.000000
1	PREDICTED	2	2.755905	4.244093
1	RESIDUAL	3	-0.755905	0.755907
1	OUTLIER	4	0.231305	0.129557
1	LOG RATIO	5	-3.995261	-3.563478
2	OBSERVED	6	98.000000	149.000000
2	PREDICTED	7	97.244080	149.755875
2	RESIDUAL	8	0.755920	-0.755875
2	OUTLIER	9	0.009494	0.009243
2	LOG RATIO	10	-0.431783	-0.000001

GWU CONTAB II SUMMARY. . FACTORS: I * J

SAMPLE SIZE 254.000000

HYPOTHESIS	NONZERO EFFECTS	SMOOTH	ZERO	2NI(X:X*)	D.F.	PROB	I*	IC*
4	MARGINALS RESIDUALS							
	I							
	J	0.0000	0.000001	0.366	1	0.5449	0.00	0.00

GWU CONTAB II. . FACTORS: I * J

Table VI

Original j = 1,2 only

RESIDUALS: I * J.

			1	2
1	OBSERVED	1	2.000000	0.000001
1	PREDICTED	2	1.597155	0.402845
1	RESIDUAL	3	0.402845	-0.402844
1	OUTLIER	4	0.094638	0.802505
1	LOG RATIO	5	-6.568467	-7.965894
2	OBSERVED	6	4601.000000	1161.000000
2	PREDICTED	7	4601.398438	1160.596924
2	RESIDUAL	8	-0.398438	0.403076
2	OUTLIER	9	-0.000765	0.000905
2	LOG RATIO	10	1.377425	-0.000001

GWU CONTAB II SUMMARY. . FACTORS: I * J

SAMPLE SIZE 5764.000000

HYPOTHESIS	NONZERO EFFECTS	SMOOTH	ZERO	2NI(X:X*)	D.F.	PROB	I*	IC*
5	MARGINALS RESIDUALS							
	I J	0.0000	0.000001	0.909	1	0.3405	0.00	0.00

Example 8. Respiratory data. This example deals with two three-way 9x2x2 contingency tables which are essentially marginal tables of a higher dimensional table, not available to us, listing data on respiratory symptoms among a group of British coal miners. It illustrates the use of OUTLIER to partition second-order interaction in a three-way contingency table. Also illustrated are multivariate logit analysis and the relations among the parameters implied by logit linearity. The generalized iterative scaling algorithm of Darroch and Ratcliff (1972) is used to obtain the m.d.i. estimates under the hypothesis of logit linearity.

EXAMPLE - RESPIRATORY DATA

This example deals with two three-way contingency tables arising from respiratory symptoms among the same group of British coal miners. The analyses progressively consider more complex hypotheses because of basic differences in certain properties of the two sets of data. Among other features the example illustrates a test of the hypothesis of no second-order interaction in a three-way contingency table, multivariate logit analysis, and the partitioning of second-order interaction in a three-way contingency table.

The techniques are based on the principle of minimum discrimination information estimation, the associated log-linear representation and analysis of information tables (see Ku et al. 1971, Kullback 1959, pp. 36-54, 155-186; 1970). The computational procedures for this example utilized the Deming-Stephan iterative marginal fitting algorithm and its extension to general linear constraints by Darroch and Ratcliff (1972). Since our m.d.i. estimates are constrained to satisfy certain linear relations based on observed values, they are maximum likelihood estimates and the associated m.d.i. test statistics are log-likelihood ratio statistics. The log-linear model has been discussed in many papers and further references may be found in Dempster (1971), Gokhale (1971), Ku et al. (1971), Plackett (1969).

In Grizzle (1971) a model developed by Grizzle, Starmer, and Koch (1969) is specialized to the case of fitting models to correlated logits. Grizzle (1971, p. 1060) says, "Unfortunately a test of the goodness-of-fit of the logit model to the joint response data has not been developed." For its methodological interest, we first consider the problem as presented by Grizzle (1971) from the minimum discrimination information estimation approach. Our results (maximum likelihood) are numerically in close agreement with those of Grizzle (BAN), but also include estimates of the cell entries under the logit model and a test of the goodness-of-fit to the joint response data.

In Table 1 is given a 9x2x2 contingency table of coal-miners classified as smokers without radiological pneumoconiosis, between the ages of 20 and 64 years inclusive at the time of their examination, showing the occurrence of breathlessness and wheeze over nine age groupings. We denote the observed frequency in any cell by x(ijk) with

Variable		Index	1	2	3	4	...	9
Age Group	A	i	20-24	25-29	30-34	35-39	...	60-64
Breathlessness	B	j	yes	no				
Wheeze	W	k	yes	no				

These data are discussed and analysed from a different point of view by Ashford and Sowden (1970), Mantel and Brown (1973).

A log-linear representation of the observed values x(ijk) in Table 1 is given in columns 1-36 of Fig. 1. The representation in Fig. 1 is a graphic presentation of the design matrix of the complete log-linear regression

$$\ln \frac{x(ijk)}{n\pi(ijk)} = L + \tau_1^A T_1^A(ijk) + \ldots + \tau_8^A T_8^A(ijk) + \tau_1^B T_1^B(ijk)$$

(1) $$+ \tau_1^W T_1^W(ijk) + \tau_{11}^{AB} T_{11}^{AB}(ijk) + \ldots + \tau_{81}^{AB} T_{81}^{AB}(ijk) + \tau_{11}^{AW} T_{11}^{AW}(ijk)$$

$$+\ldots+ \tau_{81}^{AW}T_{81}^{AW}(ijk) + \tau_{11}^{BW}T_{11}^{BW}(ijk) + \tau_{111}^{ABW}T_{111}^{ABW}(ijk)$$

$$+\ldots+ \tau_{811}^{ABW}T_{811}^{ABW}(ijk) \ ,$$

where $\Pi(ijk) = 1/9\text{x}2\text{x}2$, n is the total number of observations, L is a normalizing factor (the negative of the logarithm of a moment generating function) and the T(ijk) are linearly independent indicator functions (explanatory variables) taking on the values given by the columns of Fig. 1 and whose mean values are the various marginals.

Since Grizzle (1971) is concerned with the marginal logits of breathlessness and wheeze, this means implicitly that one is concerned with the minimum discrimination information estimate, or log-linear representation, obtained by fitting the marginals x(ij.) and x(i.k). If we denote this estimate by $x_d^*(ijk)$, then its log-linear representation or design matrix is given by columns 1-27 of Fig. 1. It may be verified that x_d^* has the explicit form $x_d^*(ijk) = x(ij.)x(i.k)/x(i..)$ and consequently we have the marginal logits

$$\ell n \frac{x_d^*(i1k)}{x_d^*(i2k)} = \ell n \frac{x(i1.)x(i.k)x(i..)}{x(i..)x(i2.)x(i.k)} = \ell n \frac{x(i1.)}{x(i2.)} \qquad \text{(breathlessness)}$$

$$\ell n \frac{x_d^*(ij1)}{x_d^*(ij2)} = \ell n \frac{x(ij.)x(i.1)x(i..)}{x(i..)x(ij.)x(i.2)} = \ell n \frac{x(i.1)}{x(i.2)} \qquad \text{(wheeze)} \ .$$

The values of $\ell n(x(i1.)/x(i2.))$ and $\ell n(x(i.1)/x(i.2))$ are given in Grizzle (1971, p. 1060) and the values of $x_d^*(ijk)$ are given in Table 2.

From Fig. 1 we have the parametric representation

$$\ell n \frac{x_d^*(i1k)}{x_d^*(i2k)} = \tau_1^B + \tau_{i1}^{AB} \ ; \quad \ell n \frac{x_d^*(ij1)}{x_d^*(ij2)} = \tau_1^W + \tau_{i1}^{AW} \ , \qquad i=1,2,\ldots,8$$

$$\ln \frac{x_d^*(91k)}{x_d^*(92k)} = \tau_1^B \; ; \qquad \ln \frac{x_d^*(9j1)}{x_d^*(9j2)} = \tau_1^W \; .$$

The values of the parameters in the parametric representation of the logits are

$\tau_1^B = -\ 0.3196$, $\tau_1^W = -\ 0.2263$, and

		τ_{i1}^{AB}	τ_{i1}^{AW}
	1	- 4.4762	- 2.6512
	2	- 3.6872	- 2.3380
	3	- 3.0106	- 1.8714
	4	- 2.4191	- 1.6241
i =	5	- 1.8993	- 1.1955
	6	- 1.4214	- 0.8840
	7	- 0.7828	- 0.5713
	8	- 0.4394	- 0.3466
	9	0	0

In particular, Grizzle's objective was to calculate two lines relating the marginal logits to age, that is, to estimate and test the hypothesis

$$\ln \frac{x_d^*(i1k)}{x_d^*(i2k)} = \alpha_1 + i\beta_1 \; ; \; \ln \frac{x_d^*(ij1)}{x_d^*(ij2)} = \alpha_2 + i\beta_2, \; i=1,\ldots,9.$$

But this hypothesis implies that the first-order differences in logits across age groups is constant, or in view of the parametric representation, that the first-order differences in the effect parameters are constant. These chains of equalities permit us to express the parameters τ_{i1}^{AB} , τ_{i1}^{AW} in terms of τ_{11}^{AB} and τ_{11}^{AW} as

$$\tau_{i1}^{AB} = \frac{9-i}{8}\,\tau_{11}^{AB}\,,\quad \tau_{i1}^{AW} = \frac{9-i}{8}\,\tau_{11}^{AW}\,,\qquad i=1,\dots,8.$$

These relations among the parameters mean that in the log-linear representation the terms

$$\dots\ \tau_{11}^{AB}T_{11}^{AB}(ijk) + \tau_{21}^{AB}T_{21}^{AB}(ijk) + \dots + \tau_{81}^{AB}T_{81}^{AB}(ijk)\ \dots$$

reduce to

$$\tau_{11}^{AB}\left(T_{11}^{AB}(ijk) + \frac{7}{8}T_{21}^{AB}(ijk) + \frac{6}{8}T_{31}^{AB}(ijk) + \dots + \frac{1}{8}T_{81}^{AB}(ijk)\right)$$

and the terms

$$\dots\ \tau_{11}^{AW}T_{11}^{AW}(ijk) + \tau_{21}^{AW}T_{21}^{AW}(ijk) + \dots + \tau_{81}^{AW}T_{81}^{AW}(ijk)\ \dots$$

reduce to

$$\tau_{11}^{AW}\left(T_{11}^{AW}(ijk) + \frac{7}{8}T_{21}^{AW}(ijk) + \frac{6}{8}T_{31}^{AW}(ijk) + \dots + \frac{1}{8}T_{81}^{AW}(ijk)\right).$$

If we denote the estimate satisfying logit linearity by x_m^* then its design matrix or log-linear representation is given by Columns 1-11, 37, 38 of Fig. 1, where we use τ^{AB} and τ^{AW} respectively instead of τ_{11}^{AB} and τ_{11}^{AW}.

The values of x_m^* were determined using the generalised iterative scaling procedure of Darroch and Ratcliff (1972) subject to the constraints

$$x_m^*(i..) = x(i..),\quad x_m^*(.j.) = x(.j.),\quad x_m^*(..k) = x(..k),$$

$$\sum_{i=1}^{8} \frac{9-i}{8} x_m^*(i1.) = \sum_{i=1}^{8} \frac{9-i}{8} x(i1.), \quad \sum_{i=1}^{8} \frac{9-i}{8} x_m^*(i.1) = \sum_{i=1}^{8} \frac{9-i}{8} x(i.1).$$

The values of $x_m^*(ijk)$ are given in Table 3. The values of the tau parameters appearing in the linear model of the logits are

$$\tau_1^B = 0.2098, \quad \tau^{AB} = -4.0996, \quad \tau_1^W = -0.1841, \quad \tau^{AW} = -2.6068.$$

The corresponding values of the logit representation in terms of the α's and β's as used by Grizzle (1971) are obtained from

$$\begin{cases} \alpha_1 + 9\beta_1 = \tau_1^B \\ \alpha_1 + \beta_1 = \tau_1^B + \tau^{AB} \end{cases}, \qquad \begin{cases} \alpha_2 + 9\beta_2 = \tau_1^W \\ \alpha_2 + \beta_2 = \tau_1^W + \tau^{AW} \end{cases},$$

or

$$\alpha_1 = -4.8219, \quad \beta_1 = 0.5125, \quad \alpha_2 = -3.1167, \quad \beta_2 = 0.3259.$$

We also note that

$$\mathrm{Var}(\alpha_1) = \mathrm{Var}(\tau_1^B) + (81/64)\ \mathrm{Var}(\tau^{AB}) + (18/8)\ \mathrm{Cov}(\tau_1^B, \tau^{AB})$$

$$\mathrm{Var}(\beta_1) = (1/64)\ \mathrm{Var}(\tau^{AB})$$

$$\mathrm{Var}(\alpha_2) = \mathrm{Var}(\tau_1^W) + (81/64)\ \mathrm{Var}(\tau^{AW}) + (18/8)\ \mathrm{Cov}(\tau_1^W, \tau^{AW})$$

$$\mathrm{Var}(\beta_2) = (1/64)\ \mathrm{Var}(\tau^{AW}).$$

The variance-covariance matrix of the taus for x_m^* is obtained as follows (a weighted version of the procedure used in Kullback 1959,

p. 217). Compute $S = T'DT$ where T is the design matrix for the log-linear representation of x_m^* (columns 1-11, 37, 38 of Fig. 1), and D is a diagonal matrix whose entries are the values of $x_m^*(ijk)$ in the order of the rows of the design matrix. Partition the matrix S as

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \text{ where } S_{11} \text{ is } 1x1.$$

Then the variance-covariance matrix of the taus is $(S_{22} - S_{21}\, S_{11}^{-1}\, S_{12})^{-1}$.

For comparison we list the values as given by Grizzle (1971) and as computed from x_m^*.

	Grizzle (1971)	x_m^*
α_1:	-4.8174 ± 0.0848	-4.8219 ± 0.0835
β_1:	0.5123 ± 0.0124	0.5125 ± 0.0129
α_2:	-3.1135 ± 0.0558	-3.1167 ± 0.0549
β_2:	0.3253 ± 0.0090	0.3258 ± 0.0089

The associated analysis of information table 4 provides a basis for tests of significance and goodness-of-fit.

Table 4

Analysis of Information

Component due to	Information	D.F.
Interaction (linear logit model)	$2I(x:x_m^*) = 3077.154$	23
Effect	$2I(x_d^*:x_m^*) = 25.300$	14
Interaction (marginal logits)	$2I(x:x_d^*) = 3051.854$	9

We infer from $2I(x:x_m^*)$ and $2I(x:x_d^*)$ that neither x_m^* or x_d^* is a good estimate for the joint response data, that is, $2I(x:x_m^*)$ ($2I(x:x_d^*)$) is a measure of the goodness-of-fit of the linear logit model (marginal logit model) to the joint response data. $2I(x_d^*:x_m^*)$ is a measure of the effect of the relationship among the parameters $\tau_{11}^{AB}, \tau_{21}^{AB}, \ldots, \tau_{81}^{AB}$ and $\tau_{11}^{AW}, {}_{21}^{AW}, \ldots, {}_{81}^{AW}$ of $x_d^*(ijk)$ implied by the hypothesis of logit linearity. We remark that x_m^* and x_d^* correspond respectively to model 3 and 8 of Mantel and Brown (1973).

We shall return to the question of finding a model providing an acceptable fit to the joint response data of Table 1 after considering data giving the prevalence of persistent cough and persistent phlegm amongst the same group of miners.

In Table 5 is given a 9x2x2 cross-classification of the same miners as in Table 1, but showing the combined prevalence of persistent cough and persistent phlegm. We denote the observed frequency in any cell by x(ijk) with

Variable		Index	1	2	3	4	...	9
Age Group	A	i	20-24	25-29	30-34	35-39	...	60-64
Cough	C	j	yes	no				
Phlegm	P	k	yes	no				

Since Table 5 has the same dimensions as Table 1 the design matrix and log-linear representation in Fig. 1 and the log-linear regression (1) for the x(ijk) values of Table 1 will be the same for the x(ijk) of Table 5 with the replacement of the superscripts B, W by C, P respectively.

To determine the significance of effects and whether or not there is second-order interaction we fit a sequence of nested models based on the marginals

H_a: $x(i..), x(.jk)$

H_b: $x(.jk), x(ij.)$

H_c: $x(.jk), x(ij.), x(i.k)$

and denote the corresponding m.d.i. estimates by x_a^*, x_b^*, x_c^* respectively. We note that x_a^* and x_b^* have the explicit form $x_a^*(ijk) = x(i..)\ x(.jk)/n$, $x_b^*(ijk) = x(ij.)\ x(.jk)/x(.j.)$ but x_c^* cannot be explicitly represented as a product of marginals. H_a is the null hypothesis that the incidence of cough and phlegm is homogeneous over the age groups. H_b is the null hypothesis that the incidence of phlegm is homogeneous over the age groups given the incidence of cough. H_c is the null hypothesis of no second-order interaction. The columns of Fig. 1 implied for the design matrix or log-linear representation of the three models are

H_a: 1-11, 28,

H_b: 1-19, 28,

H_c: 1-28 .

The hypotheses may also be stated as implying that the parameters corresponding to the columns of Fig. 1 not used in the design matrix or for the representation are zero. Analysis of information Table 6 summarizes the results.

Table 6

Analysis of Information

Component due to	Information	D.F.
a) x(i..), x(.jk)	$2I(x:x^*_a) = 1259.090$	24
b) x(.jk), x(ij.)	$2I(x^*_b:x^*_a) = 1180.385$	8
	$2I(x:x^*_b) = 78.705$	16
c) x(.jk), x(ij.), x(i.k)	$2I(x^*_c:x^*_b) = 72.009$	8
	$2I(x:x^*_c) = 6.696$	8

From Table 6 we infer that the 8 interaction parameters corresponding to columns 29-36 of Fig. 1 may be taken as zero. From Fig. 1 we see that the parametric representation of the log-odds or logits under the model of no second-order interaction are

$$\ln \frac{x^*_c(i11)}{x^*_c(i21)} = \tau^C_1 + \tau^{AC}_{i1} + \tau^{CP}_{11} ,$$

$$\ln \frac{x^*_c(i12)}{x^*_c(i22)} = \tau^C_1 + \tau^{AC}_{i1} ,$$

$$\ln \frac{x^*_c(i11)}{x^*_c(i12)} = \tau^P_1 + \tau^{AP}_{i1} + \tau^{CP}_{11} ,$$

$$\ln \frac{x^*_c(i21)}{x^*_c(i22)} = \tau^P_1 + \tau^{AP}_{i1} , \qquad i=1,2,\ldots,9.$$

The values of x^*_c are given in Table 8.

The values of the parameters in the parametric representation of the logits are

$$\tau_1^C = -2.0987,\ \tau_1^P = -2.4756,\ \tau_{11}^{CP} = 3.8500, \text{ and}$$

	τ_{i1}^{AC}	τ_{i1}^{AP}
1	-1.7955	-0.7132
2	-1.5083	-0.6904
3	-1.1155	-0.6729
4	-1.0052	-0.5734
i = 5	-0.5939	-0.5473
6	-0.3801	-0.4448
7	-0.1422	-0.3070
8	-0.1103	-0.0639
9	0	0

The covariance matrix of these 19 parameters has been computed, but is not given herein.

We mention however that the variance of τ_{11}^{CP} is 0.003116 so that

$$X^2 = (3.85)^2/0.003116 = 4756.90$$

is approximately a chi-squared with one degree of freedom. We see in Analysis of Information Table 7 a verification of the fact that the association parameter τ_{11}^{CP} is very significantly different from zero.

Table 7

Analysis of Information

Component due to	Information	D.F.
e) x(ij.), x(i.k)	$2I(x:x_e^*) = 6273.746$	9
c) x(ij.), x(i.k), x(.jk)	$2I(x_c^*:x_e^*) = 6267.050$	1
	$2I(x:x_c^*) = 6.696$	8

We remark that H_e: x(ij.), x(i.k) represents the model that cough and phlegm are not associated given the age grouping. The corresponding estimate may be explicitly represented as

$$x_e^*(ijk) = x(ij.)\,x(i.k)/x(i..).$$

$2I(x_c^*:x_e^*)$ tests the null hypothesis that $\tau_{11}^{CP} = 0$ and the value of $2I(x:x_c^*) = 6.696$, 8 D.F. implies that the association between cough and phlegm has the same value over all the age groupings.

We now examine the hypothesis that the logits of x_c^* vary linearly with age, that is, that successive differences of the logits are constant. As before we can express the parameters τ_{11}^{AC}, τ_{11}^{AP}, under this hypothesis in terms of τ_{11}^{AC} and τ_{11}^{AP} as

$$H_n: \quad \tau_{i1}^{AC} = \frac{9-i}{8}\,\tau_{11}^{AC}\,, \quad \tau_{i1}^{AP} = \frac{9-i}{8}\,\tau_{11}^{AP}\,, \qquad i=1,\ldots,8.$$

If we denote the estimate satisfying logit linearity within the model of no second-order interaction by x_n^*, then the design matrix or log-linear representation corresponding to H_n given by columns 1-11, 28, 37, 38 of Fig. 1, of course, with the replacement of the superscripts B, W by C, P respectively and the use of τ^{AC}, τ^{AP} instead of τ_{11}^{AC}, τ_{11}^{AP} respectively for convenience.

The values of x_n^* are given in Table 9. The values of the parameters in the logit representation under the logit linearity model,

$$\ell n \frac{x_n^*(i11)}{x_n^*(i21)} = \tau_1^C + \frac{9-i}{8}\tau^{AC} + \tau_{11}^{CP},$$

$$\ell n \frac{x_n^*(i12)}{x_n^*(i22)} = \tau_1^C + \frac{9-i}{8}\tau^{AC}.$$

$$\ell n \frac{x_n^*(i11)}{x_n^*(i12)} = \tau_1^P + \frac{9-i}{8}\tau^{AP} + \tau_{11}^{CP},$$

$$\ell n \frac{x_n^*(i21)}{x_n^*(i22)} = \tau_1^P + \frac{9-i}{8}\tau^{AP},$$

are

$$\tau_1^C = -1.8939,\ \tau_1^P = -2.5495,$$

$$\tau^{AC} = -1.8312,\ \tau^{AP} = -0.7646,\ \tau_{11}^{CP} = 3.8442.$$

The covariance matrix of these five parameters is given in Table 10. The associated analysis of information is given in Table 11.

Table 11

Analysis of Information

Component due to	Information	D.F.
H_n	$2I(x:x^*_n) = 28.831$	22
H_c	$2I(x^*_c:x^*_n) = 22.135$	14
	$2I(x:x^*_c) = 6.696$	8

The value $2I(x:x^*_n)$ is a measure of the goodness-of-fit of the logit linearity model and $2I(x^*_c:x^*_n)$ is a measure of the effect of replacing the common parameters τ^{AC}, τ^{AP} by τ^{AC}_{i1}, τ^{AP}_{i1}, $i=1,\ldots,8$. It is clear that x^*_c provides a better fit to the original data than x^*_n, using more parameters however, but at the 5% level of significance the logit linearity model provides an acceptable fit, with a simpler model.

In our analysis of the incidence of cough and phlegm over the age groups we concluded that the association of these factors was the same over all the age groupings. However, in multidimensional contingency tables in which, for example, time or age is one of the classifications, there may occur an age effect such that an hypothesis of interest may be rejected for the entire table, but an hypothesis taking the possible age effect into account may produce an acceptable partitioning. We now propose to illustrate techniques applicable to the solution of such problems by a further study of the 9x2x2 contingency Table 1, containing nine age groupings, for which the hypothesis of no second-order interaction is rejected. An acceptable partitioning is determined. Within the partitioned model we then consider a subhypothesis of logit linearity (Kullback and Fisher, 1973).

Let us now find the estimate under the classic null hypothesis of no second-order interaction. The minimum discrimination information estimate $x_2^*(ijk)$ under the hypothesis H_2 of no second-order interaction is obtained by iteratively fitting the marginals x(ij.), x(i.k), x(.jk) (see Ku et al., 1971, for example) and is given in Table 12. The design matrix or log-linear representation of $x_2^*(ijk)$ is given by the columns 1-28 in Fig. 1. Indeed, the no second-order interaction hypothesis is that the values of the last eight parameters in x(ijk) have the hypothetical values

$$\text{(2)} \qquad \tau_{111}^{ABW} = \tau_{211}^{ABW} = \ldots = \tau_{811}^{ABW} = 0 \ .$$

Computing the associated minimum discrimination information statistic we find

$$2I(x:x_2^*) = 2\sum\sum\sum x(ijk)\ \ell n(x(ijk)/x_2^*(ijk)) = 26.673,\ 8D.F.$$

We recall that this is the same as the log-likelihood ratio chi-squared statistic (see e.g. Darroch 1962). We reject the null hypothesis of no second-order interaction, that is, the hypothetical values in (2) are not acceptable parameters for x(ijk).

Among other properties the null hypothesis of no second-order interaction implies a common value for the association (measured by the logarithm of the cross-product ratio) between breathlessness and wheeze over all age-groups. In terms of the parameters defining $x_2^*(ijk)$ this common value as determined from columns 1-28 of Fig. 1 is

$$\ell n \frac{x_2^*(i11)x_2^*(i22)}{x_2^*(i12)x_2^*(i21)} = \tau_{11}^{BW} = 2.8348, \quad i=1,2,\ldots,9 \ .$$

We summarize the results and supplement analysis of information Table 4 by analysis of information Table 13.

Table 13

Analysis of Information

Component due to	Information	D.F.
d) x(ij.), x(i.k)	$2I(x:x^*_d) = 3051.854$	9
H_2: x(ij.), x(i.k), x(.jk)	$2I(x^*_2:x^*_d) = 3025.181$	1
	$2I(x:x^*_2) = 26.673$	8

The value of $2I(x^*_2:x^*_d)$ implies a significant (nonzero) association between breathlessness and wheeze but the value of $2I(x:x^*_2)$ leads me to conclude that there is not a common value of this association over all the age groups. We note that the estimate x^*_2 corresponds to model 9 of Mantel and Brown (1973).

It seems reasonable to conjecture that the presence of second-order interaction may be related to an age effect. That is, there may be a common value of the association between breathlessness and wheeze over some of the younger age groups and a common but different value of this association over the remaining age groups. We therefore re-examined the computer output for x^*_2. Among other items there was given for each cell a number called OUTLIER, the value of

$$2(x(ijk)\ \ell n(x(ijk)/x^*_2(ijk)) + (n-x(ijk)\ \ell n(n-x(ijk))/(n-x^*_2(ijk))).$$

Ireland (1972) has shown that large values of OUTLIER are effective in recognizing outliers under the estimation procedure in question. In the case at hand the value of OUTLIER for cell 812 was 4.959 with the next largest value 2.722 for cell 212.

Let us therefore consider a partitioning of the second-order interaction for the age groups under 55 and for the age groups 55 and over by computing the minimum discrimination information estimate $x_t^*(ijk)$ subject to the marginal restraints of $x_2^*(ijk)$ and also the restraints

$$(3) \qquad \tau_{111}^{ABW} = \tau_{211}^{ABW} = \ldots = \tau_{711}^{ABW},\ \tau_{811}^{ABW} = \tau_{911}^{ABW} = 0\ .$$

The design matrix or log-linear representation for $x_t^*(ijk)$ is given by columns 1-28, 39 in Fig. 1, that is, with the eight columns corresponding to $\tau_{111}^{ABW}, \tau_{211}^{ABW}, \ldots, \tau_{811}^{ABW}$ replaced by the one column labeled τ^{ABW}. The values of $x_t^*(ijk)$ are given in Table 14. In terms of the parameters defining $x_t^*(ijk)$, from columns 1-28, 39 in Fig. 1, it is found that

$$\ln \frac{x_t^*(i11)x_t^*(i22)}{x_t^*(i12)x_t^*(i21)} = \tau_{11}^{BW} + \tau^{ABW} = 3.0007, \quad i=1,\ldots,7$$

$$\ln \frac{x_t^*(i11)x_t^*(i22)}{x_t^*(i12)x_t^*(i21)} = \tau_{11}^{BW} = 2.5212, \quad i=8,9\ .$$

The associated analysis of information Table 15 summarizes results.

Table 15

Analysis of Information

Component due to	Information	D.F.
No second-order interaction	$2I(x:x_2^*) = 26.673$	8
Effect	$2I(x_t^*:x_2^*) = 16.700$	1
Interaction (partition)	$2I(x:x_t^*) = 9.973$	7

We note that $2I(x_t^*:x_2^*)$ which measures the effect of the hypothesis in (3) is very significant, and from the value of $2I(x:x_t^*)$ we may accept the inference that there is a common association between breathlessness and wheeze for the age groups under 55 and a different but common value for the age groups 55 and over and that in fact $x_t^*(ijk)$ is a good fit to the original data.

We remark that, as a matter of fact, the values of $x_t^*(ijk)$ were computed by iteratively fitting all the two-way marginals of the 7x2x2 table of the age groups under 55 and separately iteratively fitting all the two-way marginals of the 2x2x2 table of the age groups 55 and over.

To verify the indication given by OUTLIER we also examined the other possible "break points" with the following results

Partition	$2I(x:x_2^*)$	D.F.
Under 35	0.612	2
Over 35	15.990	5
Under 40	1.856	3
Over 40	11.541	4
Under 45	3.311	4
Over 45	8.373	3
Under 50	8.420	5
Over 50	7.861	2

These values confirm the inference suggested by OUTLIER.

If we now consider the logits for breathlessness and wheeze, respectively, for the age groups under 55, from the design matrix or log-linear representation for $x_t^*(ijk)$ in Fig. 1 (columns 1-28, 39) we see that

$$\ell n \frac{x_t^*(i11)}{x_t^*(i21)} = \tau_1^B + \tau_{i1}^{AB} + \tau_{11}^{BW} + \tau^{ABW}; \quad \ell n \frac{x_t^*(i12)}{x_t^*(i22)} = \tau_1^B + \tau_{i1}^{AB}, \quad i=1,\dots,7$$

$$\ell n \frac{x_t^*(i11)}{x_t^*(i12)} = \tau_1^W + \tau_{i1}^{AW} + \tau_{11}^{BW} + \tau^{ABW}; \quad \ell n \frac{x_t^*(i21)}{x_t^*(i22)} = \tau_1^W + \tau_{i1}^{AW}, \quad i=1,\dots,7 .$$

The corresponding logits for the age groups 55 and over are given by

$$\ell n \frac{x_t^*(811)}{x_t^*(821)} = \tau_1^B + \tau_{81}^{AB} + \tau_{11}^{BW}; \quad \ell n \frac{x_t^*(812)}{x_t^*(822)} = \tau_1^B + \tau_{81}^{AB}$$

$$\ell n \frac{x_t^*(911)}{x_t^*(921)} = \tau_1^B \qquad + \tau_{11}^{BW}; \quad \ell n \frac{x_t^*(912)}{x_t^*(922)} = \tau_1^B$$

$$\ell n \frac{x_t^*(811)}{x_t^*(812)} = \tau_1^W + \tau_{31}^{AW} + \tau_{11}^{BW}; \quad \ell n \frac{x_t^*(821)}{x_t^*(822)} = \tau_1^W + \tau_{81}^{AW}$$

$$\ell n \frac{x_t^*(911)}{x_t^*(912)} = \tau_1^W \qquad + \tau_{11}^{BW}; \quad \ell n \frac{x_t^*(921)}{x_t^*(922)} = \tau_1^W .$$

The numerical values of these logits are given in Table 16.

We now consider the hypothesis that within the partitioned no second-order hypothesis, that is, within the $x_t^*(ijk)$ model, the logits are linearly related for the age groups under 55, in other words, we consider the fitting of straight lines to the logits for the age groups under 55 by assuming that the differences of logits for successive age groups are constant.

Thus we shall consider a null hypothesis that

$$\tau^{AB}_{71} - \tau^{AB}_{61} = \tau^{AB}_{61} - \tau^{AB}_{51} = \tau^{AB}_{51} - \tau^{AB}_{41} = \ldots = \tau^{AB}_{21} - \tau^{AB}_{11} ,$$

$$\tau^{AW}_{71} - \tau^{AW}_{61} = \tau^{AW}_{61} - \tau^{AW}_{51} = \tau^{AW}_{51} - \tau^{AW}_{41} = \ldots = \tau^{AW}_{21} - \tau^{AW}_{11} .$$

If, as a matter of convenience, we consider the design matrix or log-linear representation of $x^*_t(ijk)$ as in Fig. 2, that is, a reparametrization of the log-linear representation in Fig. 1, then the chains of equalities yield the relations among the parameters

$$\tau^{AB}_{i1} = \frac{7-i}{6}\,\tau^{AB}_{11},\quad \tau^{AW}_{i1} = \frac{7-i}{6}\,\tau^{AW}_{11},\quad i=1,2,\ldots,7 .$$

The design matrix or log-linear representation for the linear logit model estimate $x^*_v(ijk)$, using τ^{AB} and τ^{AW} respectively, instead of τ^{AB}_{11} and τ^{AW}_{11} is given in columns 1-11, 28-31 of Fig. 2. The values in columns 30, 31 arise from the fact that in the log-linear representation as in (1) the terms

$$\tau^{AB}_{11}T^{AB}_{11}(ijk) + \tau^{AB}_{21}T^{AB}_{21}(ijk) +\ldots+ \tau^{AB}_{61}T^{AB}_{61}(ijk)$$

and the terms

$$\tau^{AW}_{11}T^{AW}_{11}(ijk) + \tau^{AW}_{21}T^{AW}_{21}(ijk) +\ldots+ \tau^{AW}_{61}T^{AW}_{61}(ijk)$$

because of the relations among the parameters reduce to

$$T^{AB}(T_{11}^{AB}(ijk) + (5/6)T_{21}^{AB}(ijk) + (4/6)T_{31}^{AB}(ijk) +\ldots+ (1/6)T_{61}^{AB}(ijk))$$

and

$$T^{AW}(T_{11}^{AW}(ijk) + (5/6)T_{21}^{AW}(ijk) + (4/6)T_{31}^{AW}(ijk) +\ldots+ (1/6)T_{61}^{AW}(ijk))$$

respectively.

The iteration used to compute $x_v^*(ijk)$ is (see Darroch and Ratcliff 1972)

$$x^{(5n+1)}(ijk) = \frac{x(i..)}{x^{(5n)}(i..)} x^{(5n)}(ijk)$$

$$x^{(5n+2)}(ijk) = \frac{x(.j.)}{x^{(5n+1)}(.j.)} x^{(5n+1)}(ijk)$$

$$x^{(5n+3)}(ijk) = \frac{x(..k)}{x^{(5n+2)}(..k)} x^{(5n+2)}(ijk)$$

$$x^{(5n+4)}(ijk) = \left(\frac{h_1}{h_1^{(5n+3)}}\right)^{a_1(ijk)} \left(\frac{h_2}{h_2^{(5n+3)}}\right)^{a_2(ijk)} \left(\frac{h_3}{h_3^{(5n+3)}}\right)^{a_3(ijk)}$$

$$\cdot\ x^{(5n+3)}(ijk)$$

$$x^{(5n+5)}(ijk) = \left(\frac{k_1}{k_1^{(5n+4)}}\right)^{b_1(ijk)} \left(\frac{k_2}{k_2^{(5n+4)}}\right)^{b_2(ijk)} \left(\frac{k_3}{k_3^{(5n+4)}}\right)^{b_3(ijk)} x^{(5n+4)}(ijk)$$

$$x^{(0)}(ijk) = n/28, \quad n = \sum_{i=1}^{7} \sum_{j=1}^{2} \sum_{k=1}^{2} x(ijk) .$$

All marginals refer to the 7x2x2 table and the values of $a_m(ijk)$, $b_m(ijk)$, m=1,2,3 and the restraints h_m, k_m, m=1,2,3 are given in Fig. 3. We remark that since $x_v^*(ijk)=x_t^*(ijk)$ for i=8,9, we can perform the iteration by consideration of the 7x2x2 table only. The values of $x_v^*(ijk)$ are given in Table 17.

Results are summarized in analysis of information Table 18.

Table 18

Analysis of Information

Component due to	Information	D.F.
Interaction (linear logits)	$2I(x:x_v^*) = 29.560$	17
Effect	$2I(x_t^*:x_v^*) = 19.587$	10
Interaction (partition)	$2I(x:x_t^*) = 9.973$	7

Since $2I(x:x_v^*)$ and $2I(x_t^*:x_v^*)$ fall between the 5% and 2% values of the tabulated chi-squared values with the appropriate degrees of freedom, we might accept the null hypothesis of linearity of the logits within the partitioned second-order interaction model, that is, infer from the value of $2I(x_t^*:x_v^*)$ that the parameters τ_{11}^{AB}, τ_{21}^{AB}, ..., τ_{71}^{AB} and τ_{11}^{AW}, τ_{21}^{AW}, ..., τ_{71}^{AW} of $x_t^*(ijk)$ satisfy the relations among the parameters implied by the logit linearity and that the estimate $x_v^*(ijk)$ under the logit linearity model is an acceptable estimate for the original observations.

Table 1: Number of subjects responding for the two symptoms in terms of age group

x(ijk)

Breathlessness			Yes, j=1		No, j=2		
Wheeze			Yes k=1	No k=2	Yes k=1	No k=2	Total
	1	20-24	9	7	95	1841	1952
	2	25-29	23	9	105	1654	1791
	3	30-34	54	19	177	1863	2113
Age groups	4	35-39	121	48	257	2357	2783
(years)	i=5	40-44	169	54	273	1778	2274
	6	45-49	269	88	324	1712	2393
	7	50-54	404	117	245	1324	2090
	8	55-59	406	152	225	967	1750
	9	60-64	372	106	132	526	1136

Data from Ashford and Sowden (1970).

Table 2

$x^*_d(ijk)$

	j=1		j=2	
	k=1	k=2	k=1	k=2
1	0.852	15.148	103.147	1832.851
2	2.287	29.713	125.713	1633.287
3	7.981	65.019	223.019	1816.980
4	22.954	146.046	355.045	2258.954
i=5	43.345	179.655	398.655	1652.344
6	88.467	268.533	504.533	1531.466
7	161.784	359.215	487.216	1081.784
8	201.199	356.801	429.801	762.198
9	212.070	265.929	291.929	366.070

Table 3

$x^*_m(ijk)$

	j=1		j=2	
	k=1	k=2	k=1	k=2
1	1.497	24.391	111.365	1814.747
2	3.079	36.225	137.235	1614.459
3	8.037	68.253	214.554	1822.152
4	22.967	140.816	367.283	2251.931
i=5	39.612	175.330	379.461	1679.595
6	84.650	270.466	485.742	1552.140
7	142.437	328.542	489.605	1129.413
8	214.641	357.415	441.955	735.987
9	230.975	277.656	284.834	342.486

Table 5: Combined prevalence of persistent cough and persistent phlegm in British coal miners in terms of age - all smokers without pneumoconiosis

x(ijk)

Cough			Yes, j=1		No, j=2		
Phlegm			Yes k=1	No k=2	Yes k=1	No k=2	Total
Age groups (years)	1	20-24	77	29	66	1780	1952
	2	25-29	89	40	64	1598	1791
	3	30-34	145	75	80	1813	2113
	4	35-39	237	101	107	2338	2783
	5	40-44	282	116	82	1794	2274
	i=6	45-49	373	152	99	1769	2393
	7	50-54	430	158	95	1407	2090
	8	55-59	445	122	88	1095	1750
	9	60-64	321	87	61	667	1136
			2399	880	742	14261	18282

Data from Ashford, Morgan et al. (1970).

Table 8.

x^*_c

i	j=1, k=1	j=1, k=2	j=2, k=1	j=2, k=2
1	69.919	36.096	73.078	1772.902
2	85.750	43.272	67.248	1594.727
3	147.105	72.942	77.893	1815.057
4	233.341	104.742	110.657	2334.258
i=5	276.957	121.121	87.043	1788.881
6	376.455	148.600	95.546	1772.402
7	437.482	150.460	87.521	1414.543
8	446.480	120.414	86.522	1096.588
9	325.511	82.354	56.491	671.648

Table 9.

x^*_n

i	j=1, k=1	j=1, k=2	j=2, k=1	j=2, k=2
1	72.602	42.728	64.450	1772.220
2	90.057	48.170	63.589	1589.185
3	142.725	69.383	80.161	1820.731
4	250.478	110.668	111.899	2309.956
i=5	269.945	108.401	95.926	1799.728
6	370.019	135.044	104.586	1733.352
7	414.629	137.531	93.218	1444.623
8	437.604	131.922	78.255	1102.219
9	350.939	96.152	49.918	638.992

Table 10.

Covariance matrix of τ_1^C, τ_1^P, τ^{AC}, τ^{AP}, τ_{11}^{CP}

values in x^*_n

τ_1^C	τ_1^P	τ^{AC}	τ^{AP}	τ_{11}^{CP}
0.0028	-0.0011	-0.0038	0.0024	-0.0011
	0.0037	0.0029	-0.0046	-0.0019
		0.0091	-0.0060	-0.0004
			0.0092	0.0010
				0.0031

Table 12: No second-order interaction estimate for the data of Table 1

$x_2^*(ijk)$

	j=1		j=2	
	k=1	k=2	k=1	k=2
1	7.547	8.454	96.448	1839.547
2	17.089	14.914	110.907	1648.087
3	45.954	27.054	185.040	1854.947
4	111.407	57.611	266.585	2347.390
i=5	162.527	60.504	279.467	1771.497
6	271.823	85.231	321.175	1714.769
7	398.159	122.871	250.848	1318.129
8	431.692	126.271	199.319	992.729
9	380.802	97.091	123.210	534.909

$$\ln \frac{x_2^*(i11)x_2^*(i22)}{x_2^*(i12)x_2^*(i21)} = \tau_{11}^{BW} = 2.8348$$

Table 14: Partitioned second-order interaction estimate

$x_t^*(ijk)$

	j=1		j=2	
	k=1	k=2	k=1	k=2
1	8.182	7.819	95.816	1840.183
2	18.306	13.695	109.692	1649.306
3	48.466	24.539	182.532	1857.463
4	116.719	52.292	261.279	2352.709
i=5	168.521	54.497	273.479	1777.504
6	280.217	76.810	312.784	1723.192
7	408.590	112.349	240.417	1328.652
8	411.545	146.550	219.454	972.450
9	366.455	111.450	137.546	520.550

$$\ln \frac{x_t^*(i11)x_t^*(i22)}{x_t^*(i12)x_t^*(i21)} = 3.0007, \quad i=1,\dots,7$$

$$\ln \frac{x_t^*(i11)x_t^*(i22)}{x_t^*(i12)x_t^*(i21)} = 2.5212, \quad i=8,9$$

Table 16: Logits

$\ell n \dfrac{x_t^*(i1k)}{x_t^*(i2k)}$

i	k=1	k=2
1	-2.4605	-5.4611
2	-1.7904	-4.7911
3	-1.3261	-4.3267
4	-0.8058	-3.8065
i=5	-0.4842	-3.4848
6	-0.1100	-3.1106
7	0.5303	-2.4703
8	0.6288	-1.8925
9	0.9799	-1.5413

Breathlessness

$\ell n \dfrac{x_t^*(ij1)}{x_t^*(ij2)}$

i	j=1	j=2
1	0.0455	-2.9552
2	0.2902	-2.7104
3	0.6806	-2.3200
4	0.8029	-2.1977
i=5	1.1289	-1.8717
6	1.2942	-1.7064
7	1.2911	-1.7095
8	1.0326	-1.4887
9	1.1903	-1.3309

Wheeze

Table 17: Linear logit estimate within partitioned second-order interaction model

$x_v^*(ijk)$

	j=1		j=2	
	k=1	k=2	k=1	k=2
1	11.860	9.990	108.934	1821.215
2	20.398	13.952	120.522	1636.127
3	44.705	24.830	169.946	1873.519
4	107.932	48.677	263.913	2362.476
i=5	158.232	57.944	248.880	1808.943
6	288.909	85.919	292.375	1725.797
7	416.964	100.688	271.429	1300.919
8	411.545	146.550	219.454	972.450
9	366.455	111.450	137.546	520.550

$$\ell n \frac{x_v^*(i11)x_v^*(i22)}{x_v^*(i12)x_v^*(i21)} = 2.9881, \qquad i=1,\dots,7$$

$$\ell n \frac{x_v^*(i11)x_v^*(i22)}{x_v^*(i12)x_v^*(i21)} = 2.5212, \qquad i=8,9$$

Figure 1: Log-linear representation

			1	2 τ^{A}_{1}	3 τ^{A}_{2}	4 τ^{A}_{3}	5 τ^{A}_{4}	6 τ^{A}_{5}	7 τ^{A}_{6}	8 τ^{A}_{7}	9 τ^{A}_{8}	10 τ^{B}_{1}	11 τ^{W}_{1}	12 τ^{AB}_{11}	13 τ^{AB}_{21}	14 τ^{AB}_{31}	15 τ^{AB}_{41}	16 τ^{AB}_{51}	17 τ^{AB}_{61}	18 τ^{AB}_{71}	19 τ^{AB}_{81}
[illegible]	[illegible]	[illegible]	1	1								1	1	1							
[illegible]	[illegible]	[illegible]	1	1								1		1							
[illegible]	[illegible]	[illegible]	1	1									1								
[illegible]	[illegible]	[illegible]	1	1																	
[illegible]	[illegible]	[illegible]	1		1							1	1		1						
[illegible]	[illegible]	[illegible]	1		1							1			1						
[illegible]	[illegible]	[illegible]	1		1								1								
[illegible]	[illegible]	[illegible]	1		1																
[illegible]	[illegible]	[illegible]	1			1						1	1			1					
[illegible]	[illegible]	[illegible]	1			1						1				1					
[illegible]	[illegible]	[illegible]	1			1							1								
[illegible]	[illegible]	[illegible]	1			1															
[illegible]	1	[illegible]	1				1					1	1				1				
[illegible]	1	2	1				1					1					1				
[illegible]	[illegible]	1	1				1						1								
[illegible]	[illegible]	[illegible]	1				1														
[illegible]	[illegible]	[illegible]	1					1				1	1					1			
[illegible]	[illegible]	[illegible]	1					1				1						1			
[illegible]	[illegible]	[illegible]	1					1					1								
[illegible]	[illegible]	[illegible]	1					1													
[illegible]	[illegible]	[illegible]	1						1			1	1						1		
[illegible]	[illegible]	2	1						1			1							1		
[illegible]	2	[illegible]	1						1				1								
[illegible]	[illegible]	2	1						1												
[illegible]	[illegible]	[illegible]	1							1		1	1							1	
[illegible]	[illegible]	[illegible]	1							1		1								1	
[illegible]	2	[illegible]	1							1			1								
[illegible]	2	2	1							1											
[illegible]	[illegible]	2	1								1	1	1								1
[illegible]	[illegible]	[illegible]	1								1	1									1
[illegible]	[illegible]	[illegible]	1								1		1								
[illegible]	2	2	1								1										
[illegible]	[illegible]	[illegible]	1									1	1								
[illegible]	[illegible]	[illegible]	1									1									
[illegible]	[illegible]	[illegible]	1										1								
[illegible]	2	[illegible]	1																		

			20 τ^{AW}_{11}	21 τ^{AW}_{21}	22 τ^{AW}_{31}	23 τ^{AW}_{41}	24 τ^{AW}_{51}	25 τ^{AW}_{61}	26 τ^{AW}_{71}	27 τ^{AW}_{81}	28 τ^{BW}_{11}	29 τ^{ABW}_{111}	30 τ^{ABW}_{211}	31 τ^{ABW}_{311}	32 τ^{ABW}_{411}	33 τ^{ABW}_{511}	34 τ^{ABW}_{611}	35 τ^{ABW}_{711}	36 τ^{ABW}_{811}	37 τ^{AB}	38 τ^{AW}	39 τ^{ABW}
[illegible]	[illegible]	[illegible]	1								1	1								1	1	1
[illegible]	[illegible]	[illegible]																		1		
[illegible]	[illegible]	[illegible]	1																		1	
[illegible]	[illegible]	[illegible]																				
[illegible]	[illegible]	[illegible]		1							1		1							7/8	7/8	1
[illegible]	[illegible]	[illegible]																		7/8		
[illegible]	[illegible]	[illegible]		1																	7/8	
[illegible]	[illegible]	[illegible]																				
[illegible]	[illegible]	[illegible]			1						1			1						6/8	6/8	1
[illegible]	[illegible]	[illegible]																		6/8		
[illegible]	[illegible]	[illegible]			1																6/8	
[illegible]	[illegible]	[illegible]																				
[illegible]	1	[illegible]				1					1				1					5/8	5/8	1
[illegible]	1	2																		5/8		
[illegible]	[illegible]	1				1															5/8	
[illegible]	[illegible]	[illegible]																				
[illegible]	[illegible]	[illegible]					1				1					1				4/8	4/8	1
[illegible]	[illegible]	[illegible]																		4/8		
[illegible]	[illegible]	[illegible]					1														4/8	
[illegible]	[illegible]	[illegible]																				
[illegible]	[illegible]	[illegible]						1			1						1			3/8	3/8	1
[illegible]	[illegible]	2																		3/8		
[illegible]	2	[illegible]						1													3/8	
[illegible]	[illegible]	2																				
[illegible]	[illegible]	[illegible]							1		1							1		2/8	2/8	1
[illegible]	[illegible]	[illegible]																		2/8		
[illegible]	2	[illegible]							1												2/8	
[illegible]	2	2																				
[illegible]	[illegible]	2								1	1								1	1/8	1/8	
[illegible]	[illegible]	[illegible]																		1/8		
[illegible]	[illegible]	[illegible]								1											1/8	
[illegible]	2	2																				
[illegible]	[illegible]	[illegible]									1											
[illegible]	[illegible]	[illegible]																				
[illegible]	[illegible]	[illegible]																				
[illegible]	2	[illegible]																				

Figure 2: Log-linear representation

			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
i	j	k	L	τ^A_1	τ^A_2	τ^A_3	τ^A_4	τ^A_5	τ^A_6	τ^A_8	τ^A_9	τ^B_1	τ^W_1	τ^{AB}_{11}	τ^{AB}_{21}	τ^{AB}_{31}	τ^{AB}_{41}	τ^{AB}_{51}	τ^{AB}_{61}	τ^{AB}_{81}	τ^{AB}_{91}	τ^{AW}_{11}	τ^{AW}_{21}	τ^{AW}_{31}	τ^{AW}_{41}	τ^{AW}_{51}	τ^{AW}_{61}	τ^{AW}_{81}	τ^{AW}_{91}	τ^{BW}_{11}	τ^{ABW}	τ^{AB}	τ^{AW}
1	1	1	1	1								1	1	1								1								1		1	1
1	1	2	1	1								1		1																		1	
1	2	1	1	1									1									1											1
1	2	2	1	1																													
2	1	1	1		1							1	1		1								1							1		5/6	5/6
2	1	2	1		1							1			1																	5/6	
2	2	1	1		1								1										1										5/6
2	2	2	1		1																												
3	1	1	1			1						1	1			1								1						1		4/6	4/6
3	1	2	1			1						1				1																4/6	
3	2	1	1			1							1											1									4/6
3	2	2	1			1																											
4	1	1	1				1					1	1				1								1					1		3/6	3/6
4	1	2	1				1					1					1															3/6	
4	2	1	1				1						1												1								3/6
4	2	2	1				1																										
5	1	1	1					1				1	1					1								1				1		2/6	2/6
5	1	2	1					1				1						1														2/6	
5	2	1	1					1					1													1							2/6
5	2	2	1					1																									
6	1	1	1						1			1	1						1								1			1		1/6	1/6
6	1	2	1						1			1							1													1/6	
6	2	1	1						1				1														1						1/6
6	2	2	1						1																								
7	1	1	1									1	1																	1			
7	1	2	1									1																					
7	2	1	1										1																				
7	2	2	1																														
8	1	1	1							1		1	1							1								1		1	1		
8	1	2	1							1		1								1													
8	2	1	1							1			1															1					
8	2	2	1							1																							
9	1	1	1								1	1	1								1								1	1	1		
9	1	2	1								1	1									1												
9	2	1	1								1		1																1				
9	2	2	1								1																						
			x^*_t	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
			x^*_v	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																	✓	✓	✓	✓

Preceding page blank

Figure 3

τ^{AB}

	111	112	121	122	211	212	221	222	311	312	321	322	411	412	421	422	511	512	521	522	611	612	621	622	711	712	721	722
$a_1(ijk)$	1	1			5/6	5/6			4/6	4/6			3/6	3/6			2/6	2/6			1/6	1/6						
$a_2(ijk)$			1	1			5/6	5/6			4/6	4/6			3/6	3/6			2/6	2/6			1/6	1/6				
$a_3(ijk)$					1/6	1/6	1/6	1/6	2/6	2/6	2/6	2/6	3/6	3/6	3/6	3/6	4/6	4/6	4/6	4/6	5/6	5/6	5/6	5/6	1	1	1	1

τ^{AW}

	111	112	121	122	211	212	221	222	311	312	321	322	411	412	421	422	511	512	521	522	611	612	621	622	711	712	721	722
$b_1(ijk)$	1		1		5/6		5/6		4/6		4/6		3/6		3/6		2/6		2/6		1/6		1/6					
$b_2(ijk)$		1		1		5/6		5/6		4/6		4/6		3/6		3/6		2/6		2/6		1/6		1/6				
$b_3(ijk)$					1/6	1/6	1/6	1/6	2/6	2/6	2/6	2/6	3/6	3/6	3/6	3/6	4/6	4/6	4/6	4/6	5/6	5/6	5/6	5/6	1	1	1	1

h_1: $x(11.) + \frac{5}{6}x(21.) + \frac{4}{6}x(31.) + \frac{3}{6}x(41.) + \frac{2}{6}x\ (51.) + \frac{1}{6}x(61.)$

h_2: $x(12.) + \frac{5}{6}x(22.) + \frac{4}{6}x(32.) + \frac{3}{6}x(42.) + \frac{2}{6}x(52.) + \frac{1}{6}x(62.)$

h_3: $\frac{1}{6}x(2..) + \frac{2}{6}x(3..) + \frac{3}{6}x(4..) + \frac{4}{6}x(5..) + \frac{5}{6}x(6..) + x(7..)$

k_1: $x(1.1) + \frac{5}{6}x(2.1) + \frac{4}{6}x(3.1) + \frac{3}{6}x(4.1) + \frac{2}{6}x(5.1) + \frac{1}{6}x(6.1)$

k_2: $x(1.2) + \frac{5}{6}x(2.2) + \frac{4}{6}x(3.2) + \frac{3}{6}x(4.2) + \frac{2}{6}x(5.2) + \frac{1}{6}x(6.2)$

k_3: $\frac{1}{6}x(2..) + \frac{2}{6}x(3..) + \frac{3}{6}x(4..) + \frac{4}{6}x(5..) + \frac{5}{6}x(6..) + x(7..)$

All marginals refer to the 7x2x2 table for age groups under 55.

5. The General Linear Hypothesis

1. Minimum Discrimination Information Estimation

In Chapter 3, Log-linear Representation the minimum discrimination information theorem was examined with particular emphasis on problems of fitting contingency tables based on a set of observed marginals. In such cases the $T(\omega)$ functions are indicator functions and hence take the values 0 or 1 only. In Kullback (1970) quadratic approximations to the minimum discrimination information statistics were considered and the relation of these quadratic approximations with K. Pearson's X^2 (Berkson, 1972).

We now propose to consider problems in which the $T(\omega)$ functions are general linear functions of the $p(\omega)$'s. In these problems the restraints are determined by hypotheses of interest and one is concerned whether the observed data are consistent therewith. Although these considerations really are part of the general theory already discussed it seems worthwhile to examine them in detail. We shall use the notation, terminology, and concepts of the preceding chapters with some slight modifications. Appropriate computer programs have been prepared to make application feasible.

Preceding page blank

As in Chapter 3, we want the value of $p(\omega)$ which minimizes the discrimination information

$$(1.1)\quad I(p:\pi) = \sum_{\Omega} p(\omega)\, \ln \frac{p(\omega)}{\pi(\omega)}$$

over the family of p- distributions which satisfy the restraints (using matrix notation)

$$(1.2)\quad \underline{Cp} = \underline{\theta}$$

where

$\underline{C}$ is $(r+1) \times \Omega$, $\underline{p}$ is $\Omega \times 1$, $\underline{\theta}$ is $(r+1) \times 1$, and the rank of $\underline{C}$ is $r + 1 \leq \Omega$.

If we denote the elements of the matrix $\underline{C}$ by $c_i(\omega)$, $i = 1,\ldots,(r+1)$, $\omega = 1,\ldots, \Omega$, then (1.2) is

$$(1.3)\quad \sum_{\Omega} c_i(\omega)p(\omega) = \theta_i\ ,\ i = 1,\ldots,r+1.$$

We shall usually assume $c_1(\omega) = 1$, all ω, and $\theta_1 = 1$.

In accordance with the minimum discrimination information theorem, or by differentiation of (1.1) with respect to $p(\omega)$ and using Lagrange multipliers, the minimizing distribution has the form

$$(1.4)\quad p^*(\omega) = \exp\{\lambda_1 c_1(\omega) + \lambda_2 c_2(\omega) + \ldots + \lambda_{r+1} c_{r+1}(\omega)\}\, \pi(\omega)$$

or

$$(1.5)\quad \ln \frac{p^*(\omega)}{\pi(\omega)} = \lambda_1 c_1(\omega) + \lambda_2 c_2(\omega) + \ldots + \lambda_{r+1} c_{r+1}(\omega),\ \omega=1,\ldots,\Omega .$$

This is equivalent to the version

$$(1.6)\quad \ln \frac{p^*(\omega)}{\pi(\omega)} = L + \tau_1 T_1(\omega) + \ldots + \tau_r T_r(\omega)$$

as used in Chapter 3, Log-linear Representation with

$$(1.7)\quad \lambda_1 = L,\ \lambda_{i+1} = \tau_i,\ c_1(\omega) \equiv 1,\ c_{i+1}(\omega) = T_i(\omega),\ \theta_{i+1} = \theta_i^*,$$

$$i = 1,\ldots,r ,$$

with the restraints

$$(1.8)\quad \sum_\Omega p^*(\omega) = 1,\ \sum_\Omega T_\alpha(\omega) p^*(\omega) = \theta_\alpha^*,\ \alpha = 1, 2, \ldots r.$$

In accordance with (1.3) - (1.8) we consider the partitioning of the matrices as follows:

$$\underline{C} = \begin{pmatrix} \underline{C}_1 \\ \underline{C}_2 \end{pmatrix} \text{ where } \underline{C}_1 \text{ is } 1 \times \Omega,\ \underline{C}_2 \text{ is } r \times \Omega ,$$

$$\underline{\theta} = \begin{pmatrix} 1 \\ \underline{\theta}^* \end{pmatrix} \text{ where } \underline{\theta}^* \text{ is } r \times 1 ,$$

that is $\underline{C}_1 \underline{p} = 1$, $\underline{C}_2 \underline{p} = \underline{\theta}^*$.

In the applications we take $\pi(\omega) = x(\omega)/N$, where $N = \sum_\Omega x(\omega)$. Setting $x^*(\omega) = Np^*(\omega)$, the minimum discrimination information statistic is

(1.9) $2I(x^*:x) = 2\sum_{\Omega} x^*(\omega)\ln \frac{x^*(\omega)}{x(\omega)}$,

which is asymptotically distributed as χ^2 with r degrees of freedom if the observed table x(ω) satisfies the null hypothesis or model implied by (1.2).

In accordance with the discussion in Kullback (1970, section 4, and 7) the quadratic approximation to $2I(x^*:x)$ is (see Chapter 3, section 6 herein)

(1.10) $2I(x^*:x) \approx (N\underline{\theta}^* - N\hat{\underline{\theta}})'\underline{S}_{22.1}^{-1}(N\underline{\theta}^* - N\hat{\underline{\theta}})$,

where $\underline{C}_1\underline{\pi} = 1$, $\underline{C}_2\underline{\pi} = \hat{\underline{\theta}}$, $\underline{S} = \begin{pmatrix}\underline{C}_1\\ \underline{C}_2\end{pmatrix} \underline{D}_x(\underline{C}_1', \underline{C}_2')$

$$= \begin{pmatrix} \underline{C}_1\underline{D}_x\underline{C}_1' & \underline{C}_1\underline{D}_x\underline{C}_2' \\ \underline{C}_2\underline{D}_x\underline{C}_1' & \underline{C}_2\underline{D}_x\underline{C}_2' \end{pmatrix} = \begin{pmatrix} \underline{S}_{11} & \underline{S}_{12} \\ \underline{S}_{21} & \underline{S}_{22} \end{pmatrix}, \quad \underline{S}_{22.1} = \underline{S}_{22} - \underline{S}_{21}\underline{S}_{11}^{-1}\underline{S}_{12} ,$$

and $\underline{D}_x$ is the Ω x Ω diagonal matrix with main diagonal entries x(ω). We shall see that the right-hand side of (1.10) is the minimum modified χ^2.

Some examples of the matrix $\underline{C}$

Consider a 3 x 3 contingency table

Table

11	12	13
21	22	23
31	32	33

	11	12	13	21	22	23	31	32	33	
ω	1	2	3	4	5	6	7	8	9	θ
	1	1	1	1	1	1	1	1	1	1
	0	1	0	-1	0	0	0	0	0	0
	0	0	1	0	0	0	-1	0	0	0
	0	0	0	0	0	1	0	-1	0	0

Hyp. of symmetry

$p(12) = p(21);\ p(13) = p(31);$
$p(23) = p(32);$

	11	12	13	21	22	23	31	32	33	
ω	1	2	3	4	5	6	7	8	9	θ
	1	1	1	1	1	1	1	1	1	1
	0	1	1	-1	0	0	-1	0	0	0
	0	-1	0	1	0	1	0	-1	0	0

Hyp. of marginal homogeneity

$p(11)+p(12)+p(13)=p(11)+p(21)+p(31)$,
$p(21)+p(22)+p(23)=p(12)+p(22)+p(32)$.

Consider a 2 x 2 contingency table

Table

11	12
21	22

	11	12	21	22	
ω	1	2	3	4	θ
	1	1	1	1	1
	1	1	0	0	3/4
	1	0	1	0	3/4

Hyp of specified marginals

$p(11)+p(12) = 3/4$;
$p(11)+p(21) = 3/4$.

Implies $p(21)+p(22)=1/4$, $p(12)+p(22)=1/4$.

Consider a 2 x 2 x 2 contingency table

Table

1		2	
11	12	11	12
21	22	21	22

	111	112	121	122	211	212	221	222	
ω	1	2	3	4	5	6	7	8	θ
	1	1	1	1	1	1	1	1	1
	1	1	1	1	0	0	0	0	1/2
	1	1	0	0	1	1	0	0	1/2
	1	0	1	0	1	0	1	0	1/2

Hyp of specified marginals

$p(1..)=p(111)+p(112)+p(121)+p(122)=1/2$;
$p(.1.)=p(111)+p(112)+p(211)+p(212)=1/2$;
$p(..1)=p(111)+p(121)+p(211)+p(221)=1/2$.
Implies $p(2..)=p(.2.)=p(..2)=1/2$.

2. Minimum Modified χ^2 Estimation

We shall use the same notation as before. For minimum modified χ^2 estimation we want the value of $p(\omega)$ which minimizes the modified χ^2,

$$\frac{1}{N}\chi'^2 = \sum_{\Omega} \frac{(p(\omega) - \pi(\omega))^2}{\pi(\omega)} \tag{2.1}$$

subject to the constraints (1.2) or (1.3).

Differentiating χ'^2 with respect to $p(\omega)$ and using Lagrange multipliers we have

$$(2.2) \quad \frac{\tilde{p}(\omega) - \pi(\omega)}{\pi(\omega)} - \lambda_1 c_1(\omega) - \dots - \lambda_{r+1} c_{r+1}(\omega) = 0, \quad \omega = 1, \dots, \Omega.$$

If we set $\xi(\omega) = (\tilde{p}(\omega) - \pi(\omega)/\pi(\omega)$, $\underline{\xi}' = (\xi(1), \dots, \xi(\Omega))$, $\underline{\lambda}' = (\lambda_1, \dots, \lambda_{r+1})$, then (2.2) may be written as (matrix notation)

$$(2.3) \quad \underline{\xi} = \underline{C}'\underline{\lambda} ,$$

or

$$(2.4) \quad \underline{\tilde{p}} = \underline{\pi} + \underline{D}_\pi \underline{C}'\underline{\lambda} ,$$

where $\underline{\tilde{p}}' = (\tilde{p}(1), \dots, \tilde{p}(\Omega))$, $\underline{\pi}' = (\pi(1), \dots, \pi(\Omega))$, and $\underline{D}_\pi$ is the $\Omega \times \Omega$ diagonal matrix with main diagonal $\pi(1), \dots, \pi(\Omega)$. If we set (see (1.10))

$$(2.5) \quad \underline{C\pi} = \underline{\phi}, \quad \underline{\phi}' = (1, \underline{\hat{\theta}}'),$$

then from (2.4) we get

$$(2.6) \quad \underline{C}(\underline{\tilde{p}} - \underline{\pi}) = \underline{\theta} - \underline{\phi} = \underline{C}\underline{D}_\pi \underline{C}'\underline{\lambda} ,$$

or

$$(2.7) \quad \underline{\lambda} = (\underline{C}\underline{D}_\pi \underline{C}')^{-1}(\underline{\theta} - \underline{\phi}) ,$$

that is

$$(2.8) \quad \underline{\tilde{p}} = \underline{\pi} + \underline{D}_\pi \underline{C}' (\underline{C}\underline{D}_\pi \underline{C}')^{-1}(\underline{\theta} - \underline{\phi}),$$

or

with $\tilde{\underline{x}} = N\tilde{\underline{p}}$, $\underline{x} = N\underline{\pi}$,

$$(2.9) \quad \tilde{\underline{x}} = \underline{x} + \underline{D}_x\underline{C}'(\underline{CD}_x\underline{C}')^{-1}(N\underline{\theta} - N\underline{\phi}),$$

where $\underline{D}_x = N\underline{D}_\pi$. Since

$$(2.10) \quad \min \chi'^2 = \sum_\Omega \frac{(\tilde{x}(\omega)-x(\omega))^2}{x(\omega)} = (\underline{D}_x^{-1/2}(\tilde{\underline{x}}-\underline{x}))'(\underline{D}_x^{-1/2}(\tilde{\underline{x}}-\underline{x}))$$

and from (2.9)

$$(2.11) \quad \underline{D}_x^{-1/2}(\tilde{\underline{x}} - \underline{x}) = \underline{D}_x^{1/2}\underline{C}'(\underline{CD}_x\underline{C}')^{-1}(N\underline{\theta} - N\underline{\phi}),$$

we have

$$(2.12) \quad \min \chi'^2 = (N\underline{\theta} - N\underline{\phi})'(\underline{CD}_x\underline{C}')^{-1}\underline{CD}_x^{1/2}\underline{D}_x^{1/2}\underline{C}'(\underline{CD}_x\underline{C}')^{-1}(N\underline{\theta}-N\underline{\phi})$$

$$= (N\underline{\theta} - N\underline{\phi})'(\underline{CD}_x\underline{C}')^{-1}(N\underline{\theta} - N\underline{\phi}).$$

Using the notation of (1.10)

$$(2.13) \quad (\underline{CD}_x\underline{C}')^{-1} = \underline{S}^{-1} = \begin{pmatrix} \underline{S}^{11} & \underline{S}^{12} \\ \underline{S}^{21} & \underline{S}_{22.1}^{-1} \end{pmatrix},$$

$$(2.14) \quad (N\underline{\theta} - N\underline{\phi})' = (0, N\underline{\theta}^* - N\hat{\underline{\theta}})',$$

hence

$$(2.15) \quad \min \chi'^2 = (0, N\underline{\theta}^* - N\hat{\underline{\theta}})'\begin{pmatrix} \underline{S}^{11} & \underline{S}^{12} \\ \underline{S}^{21} & \underline{S}^{-1}_{22.1} \end{pmatrix}\begin{pmatrix} 0 \\ N\underline{\theta}^* - N\hat{\underline{\theta}} \end{pmatrix}$$

$$= (N\underline{\theta}^* - N\hat{\underline{\theta}})'\underline{S}_{22.1}^{-1}(N\underline{\theta}^* - N\hat{\underline{\theta}}),$$

that is, the right-hand side of (1.10). Note that if we use the approximation

$$(2.16) \quad \ell n\frac{p(\omega)}{\pi(\omega)} = \ell n\left(1 + \frac{p(\omega) - \pi(\omega)}{\pi(\omega)}\right) \approx \frac{p(\omega) - \pi(\omega)}{\pi(\omega)}$$

then (2.2) is an approximation to (1.5).

3. An iterative computer algorithm - Single Sample

For convenience in discussion let us call the preceding discussion the single sample case. We shall consider an extension of the concepts to the k-sample case but it will be helpful not to go to the k-sample case directly. We now consider an iterative computer algorithm which will provide the minimum discrimination information estimate with the minimum modified χ^2 estimate as a by product. The single sample algorithm is a special case of the k-sample algorithm, but it will be helpful to consider the single sample case in detail (see Dempster, 1971).

(3.1) $\underline{C}\,\underline{p} = \underline{\theta}$, $\underline{C} = \begin{pmatrix} \underline{C}_1 \\ \underline{C}_2 \end{pmatrix}$, $\underline{C}_1$ is $1 \times \Omega$, $\underline{C}_2$ is $r \times \Omega$,

$\underline{\theta} = \begin{pmatrix} 1 \\ \underline{\theta}^* \end{pmatrix}$, $\underline{\theta}^*$ is $r \times 1$.

(3.2) $\underline{C}\,\underline{x} = N\,\underline{\phi}$, $\underline{\phi} = \begin{pmatrix} 1 \\ \hat{\underline{\theta}} \end{pmatrix}$, $\hat{\underline{\theta}}$ is $r \times 1$, $\underline{x}$ is $\Omega \times 1$

matrix of observations, $N = \sum_{\Omega} x(\omega)$,

(3.3) $\underline{D}_x$ is $\Omega \times \Omega$ diagonal matrix of observations,

(3.4) $\underline{S} = \underline{C}\,\underline{D}_x\,\underline{C}' = \begin{pmatrix} \underline{S}_{11} & \underline{S}_{12} \\ \underline{S}_{21} & \underline{S}_{22} \end{pmatrix}$, $\underline{S}_{11}$ is 1×1, $\underline{S}'_{21} = \underline{S}_{12}$

is $1 \times r$, $\underline{S}_{22}$ is $r \times r$,

(3.5) $\underline{S}_{22.1} = \underline{S}_{22} - \underline{S}_{21}\,\underline{S}_{11}^{-1}\,\underline{S}_{12}$,

$$(3.6)\quad \underline{\Delta} = N\underline{\theta} - N\underline{\phi} = \begin{pmatrix} 0 \\ N\underline{\theta}^* - N\hat{\underline{\theta}} \end{pmatrix} = \begin{pmatrix} 0 \\ \underline{d} \end{pmatrix},$$

$$\underline{d} = N\underline{\theta}^* - N\hat{\underline{\theta}} \text{ is } r \times 1 ,$$

$$(3.7)\quad \underline{t}^{(j)} = \underline{S}_{22.1}^{-1(j)}\, \underline{d}^{(j)},\ j=0,1,2,\ldots.$$

Let $\underline{\ell n}\ y$ denote an $\Omega \times 1$ matrix and $\underline{\ell n}\ x$ the $\Omega \times 1$ matrix of $\ell n\ x(1),\ldots,\ell n\ x(\Omega)$, where $x(1),\ldots,x(\Omega)$ are the original observations.

$$(3.8)\quad (\underline{tau})^{(j+1)} = (\underline{tau})^{(j)} + \underline{t}^{(j)},\ (\underline{tau})^{(j)} \equiv 0 \text{ for } j=0,$$

$$j=0,1,2,\ldots ,$$

$$(3.9)\quad \underline{\ell n}\ y^{(j)} = \underline{\ell n}\ x + \underline{C}_2'(tau)^{(j)},\ j=1,2,\ldots ,$$

$$(3.10)\quad y^{(j)}(1),\ldots,y^{(j)}(\Omega),\ j=1,2,\ldots ,$$

$$(3.11)\quad L^{(j)} = \ell n \frac{N}{y^{(j)}(1)+\ldots+y^{(j)}(\Omega)},\ j=1,2,\ldots ,$$

$$(3.12)\quad \left\{\begin{array}{l} \ell n\ x^{(j)}(1) = L^{(j)} + \ell n\ y^{(j)}(1) \\ \qquad\vdots \\ \ell n\ x^{(j)}(\Omega) = L^{(j)} + \ell n\ y^{(j)}(\Omega) \end{array}\right.\ j=1,2,\ldots ,$$

$$(3.13)\quad x^{(j)}(1),\ldots,x^{(j)}(\Omega),\ j=1,2,\ldots .$$

In step (3.7), j=0 corresponds to the values computed in steps (3.1) to (3.6) using the original observations, and j=1,2,... corresponds to the procedures in steps (3.1) to (3.6) however using the values

$x^{(j)}(1),\ldots,x^{(j)}(\Omega)$ in step (3.13).

Note that in step (3.9) $\underline{\ell n}$ x is <u>always</u> composed of the original observations.

The iteration is continued until the maximum value of the absolute values of the differences between successive iterates is less than a specified small value.

The final iterated value $x^{(j)}$ is the m.d.i. estimate x* and 2I(x*:x) is computed and is asymptotically a chi-square with r degrees of freedom.

The matrix $\underline{S}_{22.1}^{-1}$ for the last iterate is the covariance matrix of the taus which are the parameter values of x*.

If the min. mod. χ^2 estimates and the min. mod. χ^2 value are desired the program continues and computes

$$(3.14) \quad \underline{\lambda} = (\underline{C}\ \underline{D}_x\ \underline{C}')^{-1}\ \underline{\Delta} = \underline{S}^{-1}\ \underline{\Delta} = \begin{pmatrix} \underline{S}^{12} & \underline{d} \\ \underline{S}_{22.1}^{-1} & \underline{d} \end{pmatrix},$$

$$(3.15) \quad \underline{\mu} = \underline{C}'\ \underline{\lambda} = \underline{C}'(\underline{C}\ \underline{D}_x\ \underline{C}')^{-1}\ \underline{\Delta},$$

$$(3.16) \quad \tilde{x} = x + \underline{D}_x\ \underline{\mu} = x + \underline{D}_x\ \underline{C}'(\underline{C}\ \underline{D}_x\ \underline{C}')^{-1}\ \underline{\Delta},$$

$$(3.17) \quad X^2 = \underline{\Delta}'\ \underline{\lambda} = \underline{\Delta}'(\underline{C}\ \underline{D}_x\ \underline{C}')^{-1}\ \underline{\Delta} = (0,\underline{d}')\begin{pmatrix} \underline{S}^{11} & \underline{S}^{12} \\ \underline{S}^{21} & \underline{S}_{22.1}^{-1} \end{pmatrix}\begin{pmatrix} 0 \\ \underline{d} \end{pmatrix}$$

$$= \underline{d}'\ \underline{S}_{22.1}^{-1}\ \underline{d}.$$

The $\tilde{x}$ in (3.16) are the minimum modified χ^2 estimates and X^2 in (3.17) is the value of the minimum modified χ^2 with r degrees of freedom and is the quadratic approximation to 2I(x*:x).

Note that X^2 in (3.17) can be calculated without first getting $\tilde{x}$.

To illustrate the single sample algorithm let us consider a 2 x 2 contingency table discussed by R. A. Fisher, Statistical Methods for Research Workers, 7th Ed. p. 314 and also considered in

Ireland and Kullback, (1968b)

using a different algorithm, viz, adjustments of the marginals.

The 2 x 2 contingency table gives seedling counts on self-fertilised heterozygotes for two factors in maize, Starch v. Sugary and Green v. White base leaf.

Table

	Green	White	
Starchy	1997	906	2903
Sugary	904	32	936
	2901	938	3839

In accordance with genetic theory, the marginals should occur in the ratio 3 to 1 and it is desired to calculate an estimate consistent with the genetic theory and test whether the observed values are consistent therewith.

The $\underline{C}$ matrix and $\underline{\theta}$ are

	11	12	21	22	
ω	1	2	3	4	θ
	1	1	1	1	1
	1	1	0	0	.75
	1	0	1	0	.75

$$\underline{x} = \begin{pmatrix} 1997 \\ 906 \\ 904 \\ 32 \end{pmatrix}, \qquad \underline{D}_x = \begin{pmatrix} 1997 & 0 & 0 & 0 \\ 0 & 906 & 0 & 0 \\ 0 & 0 & 904 & 0 \\ 0 & 0 & 0 & 32 \end{pmatrix},$$

$$\underline{\ell n\ x} = \begin{pmatrix} 7.599401 \\ 6.809039 \\ 6.806829 \\ 3.465736 \end{pmatrix},$$

$$\begin{array}{l} \underline{C}\ \underline{x} = N\underline{\phi} \\ N = 3839 \end{array} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1997 \\ 906 \\ 904 \\ 32 \end{pmatrix} = \begin{pmatrix} 3839 \\ 2903 \\ 2901 \end{pmatrix},$$

$$\underline{S} = \underline{C}\ \underline{D}_x\ \underline{C}' = \begin{pmatrix} 3839 & 2903 & 2901 \\ 2903 & 2903 & 1997 \\ 2901 & 1997 & 2901 \end{pmatrix},$$

$$\underline{S}_{22.1} = \begin{pmatrix} 2903 & 1997 \\ 1997 & 2901 \end{pmatrix} - \begin{pmatrix} 2903 \\ 2901 \end{pmatrix} \frac{1}{3839} (2903 \quad 2901)$$

$$= \begin{pmatrix} 2903 & 1997 \\ 1997 & 2901 \end{pmatrix} - \begin{pmatrix} 2195.2094 & 2193.6971 \\ 2193.6971 & 2192.1857 \end{pmatrix}$$

$$= \begin{pmatrix} 707.7906 & -196.6971 \\ -196.6971 & 708.8143 \end{pmatrix},$$

$$\underline{\Delta} = \begin{pmatrix} 0 \\ \underline{d} \end{pmatrix}, \quad \underline{d} = N\underline{\theta}^{*} - N\underline{\hat{\theta}}$$

$$= \begin{pmatrix} 2879.25 \\ 2879.25 \end{pmatrix} - \begin{pmatrix} 2903 \\ 2901 \end{pmatrix}$$

$$= \begin{pmatrix} -23.75 \\ -21.75 \end{pmatrix},$$

$$\underline{S}_{22.1}^{-1} = \frac{1}{\text{Det}} \begin{pmatrix} 708.8143 & +196.6971 \\ +196.6971 & 707.7906 \end{pmatrix}, \ \text{Det} = 463002.3496$$

$$= \begin{pmatrix} .00153 & .00042 \\ .00042 & .00153 \end{pmatrix},$$

$$\underline{t}^{(0)} = \begin{pmatrix} .00153 & .00042 \\ .00042 & .00153 \end{pmatrix} \begin{pmatrix} -23.75 \\ -21.75 \end{pmatrix} = \begin{pmatrix} -.0454725 \\ -.0432525 \end{pmatrix},$$

$$(\underline{tau})^{(1)} = \begin{pmatrix} -.0454725 \\ -.0432525 \end{pmatrix},$$

$$\underline{C}_2'(\underline{tau})^{(1)} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} -.0454725 \\ -.0432525 \end{pmatrix} = \begin{pmatrix} -.088725 \\ -.0454725 \\ -.0432525 \\ 0 \end{pmatrix},$$

$$\ln y^{(1)}(1) = 7.599401 - .088725 = 7.510676,$$
$$\ln y^{(1)}(2) = 6.809039 - .0454725 = 6.763567,$$
$$\ln y^{(1)}(3) = 6.806829 - .0432525 = 6.763576,$$
$$\ln y^{(1)}(4) = 3.465736 - 0 = 3.465736,$$

$$y^{(1)}(1) = 1827.448, \quad y^{(1)}(3) = 865.733,$$
$$y^{(1)}(2) = 865.733, \quad y^{(1)}(4) = 32.000,$$

$$y^{(1)}(1)+\ldots+y^{(1)}(4) = 3590.914,$$

$$L^{(1)} = \ln \frac{3839}{3590.914} = 0.066805,$$

$$\ln x^{(1)}(1) = 0.066805 + 7.510676 = 7.577481, \quad x^{(1)}(1) = 1953.701,$$
$$\ln x^{(1)}(2) = 0.066805 + 6.763567 = 6.830372, \quad x^{(1)}(2) = 925.535,$$
$$\ln x^{(1)}(3) = 0.066805 + 6.763576 = 6.830381, \quad x^{(1)}(3) = 925.543,$$
$$\ln x^{(1)}(4) = 0.066805 + 3.465736 = 3.532541, \quad x^{(1)}(4) = 34.211,$$

$$X^2 = (-23.75, -21.75) \begin{pmatrix} -.0454725 \\ -.0432525 \end{pmatrix} = 2.021.$$

Retaining two decimal places we take

$$x^*(1) = 1953.71 = x^*(11),$$

$$x^*(2) = 925.54 = x^*(12),\quad x^*(1.) = 2879.25,$$

$$x^*(3) = 925.54 = x^*(21),\quad x^*(.1) = 2879.25,$$

$$x^*(4) = \underline{34.21} = x^*(22).$$

$$3839.00$$

Since $\underline{d}^{(1)} = \begin{pmatrix}0\\0\end{pmatrix}$,

$$\underline{t}^{(1)} = \underline{S}_{22.1}^{-1(1)}\ \underline{d}^{(1)} = (\underline{0}),$$

and there will be no change in the estimate by further iteration.

$$2I(x^*:x) = 2\left(1953.71\ln\frac{1953.71}{1997} + 925.54\ln\frac{925.54}{906}\right.$$

$$\left.+ 925.54\ln\frac{925.54}{904} + 34.21\ln\frac{34.21}{32}\right)$$

$$= 2(-42.8174 + 19.7492 + 21.7946 + 2.2846)$$

$$= 2(1.011) = 2.022\ .$$

4. k-samples

The extension of the previous single sample discussion to the case of k-samples makes use of an approach due to

Gokhale (1973).

Consider the k discrete spaces Ω_i, $i=1,2,\ldots,k$, where we designate the "points" or "cells" of Ω_i by $\omega_i(j)$, $j=1,2,\ldots,\Omega_i$ $\underline{\omega}_i = (\omega_i(1),\ldots,\omega_i(\Omega_i))$. We use Ω_i to represent both the space and the number of "cells" in it. Let $\underline{p}_i = (p_i(\omega_i(1)),\ldots,p_i(\omega_i(\Omega_i)))$, $i=1,\ldots,k$, be k sets of probability distributions defined respectively over Ω_i, $i=1,2,\ldots,k$. Let $\underline{p}' = (\underline{p}_1,\ldots,\underline{p}_k)$ be a $1 \times \Omega$ matrix, where $\Omega = \Omega_1+\Omega_2+\ldots+\Omega_k$. Let P be the collection of all such matrices (vectors) $\underline{p}$. For a given $\underline{\pi}' = (\underline{\pi}_1,\ldots,\underline{\pi}_k)\varepsilon P$ and $\underline{p}\varepsilon P$ the generalized discrimination information is given by

$$(4.1) \quad I(\underline{p}:\underline{\pi}) = \sum_i w_i \sum_{j=1}^{\Omega_i} p_i(\omega_i(j))\ln(p_i(\omega_i(j)/\pi_i(\omega_i(j))),$$

where the constants w_i are known and are such that $\sum_i w_i = 1$, $0<w_i<1$. Let us denote the elements ("points" or "cells") of $\Omega = \Omega_1+\Omega_2+\ldots+\Omega_k$ by $\omega(ij)$, $i=1,\ldots,k$, $j=1,\ldots,\Omega_i$, so that $\omega(i1),\ldots,\omega(i\Omega_i)$ are the components of Ω belonging to Ω_i.

The minimum discrimination information estimate is the value of $\underline{p}$ which minimizes the generalized discrimination information in (4.1) over the family of $\underline{p}$'s which satisfy the restraints

$$\underline{B}\,\underline{p} = \underline{\theta}\,, \tag{4.2}$$

where $\underline{B}$ is $(k+r)x\Omega$, $\underline{p}$ is $\Omega x1$, $\underline{\theta}$ is $(k+r)x1$ and the rank of $\underline{B}$ is $k+r<\Omega$. We shall now transform the problem to a canonical form similar to that of the single sample case.

Let

(4.3) $\underline{W}_i$ be an $\Omega_i x \Omega_i$ diagonal matrix with diagonal elements w_i,

and

$$\underline{W} = \begin{pmatrix} \underline{W}_1 & 0 & . & . & . & 0 \\ 0 & \underline{W}_2 & . & . & . & 0 \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \underline{W}_k \end{pmatrix}, \quad \Omega \times \Omega\,, \tag{4.4}$$

$$\underline{P} = \underline{W}\,\underline{p},\ \underline{P}'=(P(\omega(11),\ldots,P(\omega(1\Omega_1),\ldots,P(\omega(k1)),\ldots,P(\omega(k,\Omega_k))) \tag{4.5}$$

$$\underline{\Pi} = \underline{W}\,\underline{\pi},\ \underline{\Pi}' = (\Pi(\omega(11),\ldots,\Pi(\omega(k\Omega_k))), \tag{4.6}$$

$$\underline{C} = \underline{B}\,\underline{W}^{-1},\ \underline{C} \text{ is } (k+r)x\Omega,\ \underline{W}^{-1} = \underline{V}\,. \tag{4.7}$$

We note that

$$\sum_{\Omega} P(\omega) = \sum_{j=1}^{\Omega_1} w_1 p_1(\omega_1(j))+\ldots+\sum_{j=1}^{\Omega_k} w_k p_k(\omega_k(j)) \tag{4.8}$$

$$= w_1 +\ldots+ w_k = 1\,,$$

$$(4.9)\quad \sum_{\Omega} \Pi(\omega) = \sum_{j=1}^{\Omega_1} w_1 \pi_1(\omega_1(j)) + \dots + \sum_{j=1}^{\Omega_k} w_k \pi_k(\omega_k(j))$$

$$= w_1 + \dots + w_k = 1 \ ,$$

$$(4.10)\quad I(\underline{p}:\underline{\pi}) = \sum_{i=1}^{k} \sum_{j=1}^{\Omega_i} w_i p_i(\omega_i(j)) \ln \frac{w_i p_i(\omega_i(j))}{w_i \pi_i(\omega_i(j))}$$

$$= \sum_{\Omega} P(\omega) \ln \frac{P(\omega)}{\Pi(\omega)} = I(P:\Pi),$$

$$(4.11)\quad \underline{B}\ \underline{p} = \underline{B}\ \underline{W}^{-1}\ \underline{W}\ \underline{p} = \underline{C}\ \underline{P} = \underline{\theta}\ .$$

In terms of the canonical transformation the k-sample problem may now be formulated as finding the m.d.i. estimate $P^*(\omega)$ minimizing

$$(4.12)\quad I(P:\Pi) = \sum_{\Omega} P(\omega) \ln \frac{P(\omega)}{\Pi(\omega)} \ ,$$

subject to

$$(4.13)\quad \underline{C}\ \underline{P} = \underline{\theta}\ ,$$

where $\underline{C}$ is $(k+r)\mathrm{x}\Omega$, $\underline{P}$ is $\Omega\mathrm{x}1$, $\underline{\theta}$ is $(k+r)\mathrm{x}1$ and the rank of $\underline{C}$ is $k+r<\Omega$. Paralleling the discussion of the single sample case, with appropriate modifications, we denote the elements of the matrix $\underline{C}$ by $c_i(\omega)$, $i=1,\dots,k,k+1,\dots,k+r$, $\omega=11,\dots,1\Omega_1,\dots,k1,\dots,k\Omega_k$. We may write (4.13) as

$$(4.14)\quad \sum_{\Omega} c_i(\omega) P(\omega) = \theta_i,\ i=1,\dots,k,k+1,\dots,k+r.$$

We shall usually assume $b_i(w_i(j))=1, j=1,\dots,\Omega_i, i=1,\dots,k,$

and zero otherwise, that is,

(4.15) $c_i(\omega) = v_i$ for $\omega = i1, \ldots, i\Omega_i$, $v_i = 1/w_i$,

$= 0$ otherwise, $i=1,2,\ldots,k$,

$\theta_i = 1$, $i=1,2,\ldots,k$.

In accordance with the m.d.i. theorem we have

(4.16) $$\ell n \frac{P^*(\omega)}{\Pi(\omega)} = \lambda_1 c_1(\omega)+\ldots+\lambda_k c_k(\omega)+\lambda_{k+1}c_{k+1}(\omega)+\ldots+\lambda_{k+r}c_{k+r}(\omega),$$

$$\omega = 11,\ldots,k\Omega_k.$$

We now partition the matrices as follows:

(4.17) $\underline{C} = \begin{pmatrix} \underline{C}_1 \\ \underline{C}_2 \end{pmatrix}$, where $\underline{C}_1$ is $k\times\Omega$, $\underline{C}_2$ is $r\times\Omega$,

(4.18) $\underline{\theta} = \begin{pmatrix} \underline{1} \\ \underline{\theta}^* \end{pmatrix}$, where $\underline{1}$ is a kx1 matrix of 1's, $\underline{\theta}^*$ is

rx1 that is, $\underline{C}_1 \underline{P} = \underline{1}$, $\underline{C}_2 \underline{P} = \underline{\theta}^*$.

If we have k samples corresponding to $\Omega_1,\ldots,\Omega_k$, where the sum of the observations in the i-th sample is N_i and $N = N_1 + N_2+\ldots+N_k$, then $w_i = N_i/N$,

(4.19) $x^*(\omega) = N\, P^*(\omega)$,

(4.20) $x(\omega) = N\, \Pi(\omega)$,

(4.21) $$\sum_{\Omega} x(\omega) = \sum_{j=1}^{\Omega_1} N_1 \frac{x_1(j)}{N_1} +\ldots+ \sum_{j=1}^{\Omega_k} N_k \frac{x_k(j)}{N_k}$$

$$= N_1+\ldots+N_k = N.$$

The minimum discrimination information statistic is

$$(4.22)\quad 2I(x^*:x) = 2NI(P^*:\Pi) = 2\sum_{\Omega} x^*(\omega)\ln\frac{x^*(\omega)}{x(\omega)},$$

which is asymptotically distributed as χ^2 with r degrees of freedom if the observed values satisfy the hypothesis or model implied by (4.2) or (4.13). If we set

$$(4.23)\quad \underline{C}\,\underline{N}\,\underline{\Pi} = \underline{C}\,\underline{x} = N\,\underline{\phi},\ \underline{\phi} = \begin{pmatrix} \underline{1} \\ \underline{\theta} \end{pmatrix},$$ where $\underline{x}$ is Ωx1, $\underline{1}$ is a kx1 matrix of 1's, $\hat{\underline{\theta}}$ is rx1,

then the quadratic approximation to $2I(x^*:x)$ is given by the minimum modified χ^2 with rD.F.,

$$(4.24)\quad X^2 = (N\underline{\theta}^*-N\hat{\underline{\theta}})'S^{-1}_{22.1}(N\underline{\theta}^*-N\hat{\underline{\theta}}),$$

where

$$(4.25)\quad \underline{S} = \underline{C}\,\underline{D}_x\,\underline{C}' = \begin{pmatrix} \underline{C}_1\underline{D}_x\underline{C}_1' & \underline{C}_1\underline{D}_x\underline{C}_2' \\ \underline{C}_2\underline{D}_x\underline{C}_1' & \underline{C}_2\underline{D}_x\underline{C}_2' \end{pmatrix}$$

$$= \begin{pmatrix} \underline{S}_{11} & \underline{S}_{12} \\ \underline{S}_{21} & \underline{S}_{22} \end{pmatrix},$$

where $\underline{S}_{11}$ is k x k, $\underline{S}_{21}' = \underline{S}_{21}$ is k x r, $\underline{S}_{22}$ is r x r and $\underline{S}_{22.1} = \underline{S}_{22}-\underline{S}_{21}\underline{S}_{11}^{-1}\underline{S}_{12}$.

An elementary example illustrating the $2I(x^*:x)$ quadratic approximation using the several sample approach.

Suppose we have observed two binomial samples

$$x(11),\ x(12),\ x(11) + x(12) = N_1,$$

$$x(21),\ x(22),\ x(21) + x(22) = N_2,$$

and we want to test the null hypothesis that $p(11) = p(21)$.

The set up corresponding to $\underline{Bp} = \underline{\theta}$ is

	11	12	21	22	
ω :	1	2	3	4	θ
	1	1	0	0	1
	0	0	1	1	1
	1	0	-1	0	0

Using $v_1 = \frac{1}{w_1} = N/N_1$, $V_2 = \frac{1}{w_2} = N/N_2$, $N = N_1 + N_2$,

the transformation to $\underline{CP} = \underline{\theta}$ is

	11	12	21	22	
ω :	1	2	3	4	θ
	v_1	v_1	0	0	1
	0	0	v_2	v_2	1
	v_1	0	$-v_2$	0	0

We must compute $\underline{CD_xC'}$, that is

$$\begin{pmatrix} v_1 & v_1 & 0 & 0 \\ 0 & 0 & v_2 & v_2 \\ v_1 & 0 & -v_2 & 0 \end{pmatrix} \begin{pmatrix} x(11) & 0 & 0 & 0 \\ 0 & x(12) & 0 & 0 \\ 0 & 0 & x(21) & 0 \\ 0 & 0 & 0 & x(22) \end{pmatrix} \begin{pmatrix} v_1 & 0 & v_1 \\ v_1 & 0 & 0 \\ 0 & v_2 & -v_2 \\ 0 & v_2 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} v_1^2(x(11) + x(12)) & 0 & v_1^2 x(11) \\ 0 & v_2^2(x(21) + x(22)) & -v_2^2 x(21) \\ v_1^2 x(11) & -v_2^2 x(21) & v_1^2 x(11) + v_2^2 x(21) \end{pmatrix}.$$

We now find

$$\underline{S}_{22.1} = v_1^2 x(11) + v_2^2 x(21) - (v_1^2 x(11), -v_2^2 x(21)) \begin{pmatrix} v_1^2 N_1 & 0 \\ 0 & v_2^2 N_2 \end{pmatrix}^{-1} \begin{pmatrix} v_1^2 x(11) \\ -v_2^2 x(21) \end{pmatrix}$$

$$= v_1^2 x(11) + v_2^2 x(21) - \frac{v_1^2 x^2(11)}{N_1} - v_2^2 \frac{x^2(21)}{N_2}$$

$$= v_1^2 x(11) \left(1 - \frac{x(11)}{N_1}\right) + v_2^2 x(21) \left(1 - \frac{x(21)}{N_2}\right).$$

But

$$\underline{d} = 0 - (v_1 x(11) - v_2 x(21)),$$

hence $X^2 = \underline{d}'\underline{S}_{22.1}^{-1}\underline{d}$ is

$$X^2 = \frac{(v_1 x(11) - v_2 x(21))^2}{v_1^2 x(11)\left(1 - \frac{x(11)}{N_1}\right) + v_2^2 x(21)\left(1 - \frac{x(21)}{N_2}\right)} =$$

$$= \frac{\left(\frac{Nx(11)}{N_1} - \frac{NX(21)}{N_2}\right)^2}{\frac{N^2}{N_1^2}\, x(11)\left(1 - \frac{x(11)}{N_1}\right) + \frac{N^2}{N_2^2}\, x(21)\left(1 - \frac{x(21)}{N_2}\right)}$$

$$= \frac{(\hat{p}(11) - \hat{p}(21))^2}{\frac{\hat{p}(11)\hat{q}(11)}{N_1} + \frac{\hat{p}(21)\hat{q}(21)}{N_2}}, \quad \hat{p}(11) = \frac{x(11)}{N_1}, \ \hat{q}(11) = 1 - \hat{p}(11),$$

$$\hat{p}(21) = \frac{x(21)}{N_2}, \ \hat{q}(21) = 1 - \hat{p}(21).$$

See Kullback (1959, p. 311), Snedecor and Cochran (1967, p. 496).

5. An iterative computer algorithm - k-samples

For convenience (computer-wise) we shall use n_i for Ω_i and n for Ω, that is, $n = n_1+n_2+\ldots+n_k$, where n_i is the number of "cells" in the i-th sample whose total number of observations is N_i.

(5.1) $\underline{C}\,\underline{P} = \underline{\theta}$, $\underline{C} = \begin{pmatrix} \underline{C}_1 \\ \underline{C}_2 \end{pmatrix}$, $\underline{C}_1$ is k x n, $\underline{C}_2$ is r x n,

$\underline{\theta} = \begin{pmatrix} \underline{1} \\ \underline{\theta}^* \end{pmatrix}$, $\underline{1}$ is a k x 1 matrix of ones, $\underline{\theta}^*$ is r x 1,

(5.2) $\underline{C}\,\underline{x} = N\,\underline{\phi}$, $\underline{\phi} = \begin{pmatrix} \underline{1} \\ \hat{\underline{\theta}} \end{pmatrix}$, $\underline{1}$ is a k x 1 matrix of ones,

$\hat{\underline{\theta}}$ is r x 1,

(5.3) $\underline{D}_x$ is n x n diagonal matrix of observations,

(5.4) $\underline{S} = \underline{C}\,\underline{D}_x\,\underline{C}' = \begin{pmatrix} \underline{S}_{11} & \underline{S}_{12} \\ \underline{S}_{21} & \underline{S}_{22} \end{pmatrix}$, $\underline{S}_{11}$ is k x k, $\underline{S}_{21}' = \underline{S}_{12}$ is k x r,

$\underline{S}_{22}$ is r x r,

(5.5) $\underline{S}_{22.1} = \underline{S}_{22} - \underline{S}_{21}\underline{S}_{11}^{-1}\underline{S}_{12}$,

(5.6) $\underline{\Delta} = N\underline{\theta} - N\underline{\phi} = \begin{pmatrix} \underline{0} \\ \underline{d} \end{pmatrix}$, $\underline{0}$ is a k x 1 matrix of zeros,

$\underline{d} = N\underline{\theta}^* - N\hat{\underline{\theta}}$ is r x 1,

(5.7) $\underline{t}^{(j)} = \underline{S}_{22.1}^{-1(j)}\,\underline{d}^{(j)}$, j=0,1,2,...

Let $\underline{\ln}\ y$ denote an n x 1 matrix and $\underline{\ln}\ x$ the n x 1 matrix of $\ln x(1),\ldots,\ln x(n)$,

(5.8) $\underline{(tau)}^{(j+1)} = \underline{(tau)}^{(j)} + \underline{t}^{(j)}$, $\underline{(tau)}^{(j)} \equiv 0$ for j=0,

j=0,1,2,... ,

(5.9) $\ell n\ y^{(j)} = \ell n\ x + C_2'(tau)^{(j)},\ j=1,2,\ldots,$

(5.10) $y^{(j)}(1),\ldots,y^{(j)}(n),\ j=1,2,\ldots,$

$$(5.11)\quad \begin{cases} S_1^{(j)} = \sum_{\Omega_1} y^{(j)}(\omega) & \text{for } \omega \text{ the } n_1 \text{ values in the first set,} \\ \vdots & \vdots \\ S_k^{(j)} = \sum_{\Omega_k} y^{(j)}(\omega) & \text{for } \omega \text{ the } n_k \text{ values in the } k^{\underline{th}} \text{ set,} \end{cases}$$

(5.12) $v_h\ L_h^{(j)} = M_h^{(j)} = \ell n \dfrac{N_h}{S_h^{(j)}},\ h=1,2,\ldots,k,$

(5.13) $\ell n\ x^{(j)}(\omega) = M_h^{(j)} + \ell n\ y^{(j)}(\omega)$, for ω in set $h=1,2,\ldots,k$,
$j=1,2,\ldots,$

(5.14) $x^{(j)}(1),\ldots,x^{(j)}(n),\ j=1,2,\ldots$

In step (5.7), $j=0$ corresponds to the values computed in steps (5.1) to (5.6) using $\underline{x}$ and $j=1,2,\ldots$ corresponds to the procedures in steps (5.1) to (5.6) however using the values $x^{(j)}(1),\ldots,x^{(j)}(n)$ in step (5.14). Note that in step (5.9) $\underline{\ell n}$ x is <u>always</u> composed of the initial values x.

The iteration is continued until the maximum value of the absolute values of the differences between successive iterates is less than a specified small value.

The final iterated value $x^{(j)}$ is the m.d.i. estimate x^* and $2I(x^*:x)$ is computed with r degrees of freedom.

If the min. mod. χ^2 estimates and the min. mod. χ^2

value are desired the program continues and computes ,

$$(5.15)\quad \underline{\lambda} = (\underline{C}\ \underline{D}_x\ \underline{C}')^{-1}\underline{\Delta} = \underline{S}^{-1}\ \underline{\Delta} = \begin{pmatrix} \underline{S}^{12}\ \underline{d} \\ \underline{S}^{-1}_{22.1}\ \underline{d} \end{pmatrix},$$

$$(5.16)\quad \underline{\mu} = \underline{C}'\ \underline{\lambda} = \underline{C}'(\underline{C}\ \underline{D}_x\ \underline{C}')^{-1}\ \underline{\Delta},$$

$$(5.17)\quad \underline{\tilde{x}} = \underline{x} + \underline{D}_x\ \underline{\mu} = \underline{x} + \underline{D}_x\ \underline{C}'(\underline{C}\ \underline{D}_x\ \underline{C}')^{-1}\ \underline{\Delta},$$

$$(5.18)\quad X^2 = \underline{\Delta}'\ \underline{\lambda} = \underline{\Delta}'(\underline{C}\ \underline{D}_x\ \underline{C}')^{-1}\ \underline{\Delta} = (\underline{0},\underline{d}')\begin{pmatrix} \underline{S}^{11} & \underline{S}^{12} \\ \underline{S}^{21} & \underline{S}^{-1}_{22.1} \end{pmatrix}\begin{pmatrix} \underline{0} \\ \underline{d} \end{pmatrix}$$

$$= \underline{d}'\ \underline{S}^{-1}_{22.1}\ \underline{d}.$$

The $\tilde{x}$ in (5.17) are the minimum modified χ^2 estimates and X^2 in (5.18) is the value of the minimum modified χ^2 with r degrees of freedom and is the quadratic approximation to $2I(x^*:x)$. Note that X^2 in (5.18) can be computed without getting $\tilde{x}$.

6. Computer Programs

A basic program using the marginal fitting technique was prepared by Professor C. T. Ireland of The George Washington University. The current version in The George Washington University Computer Center is CONTAB III.

A modification of CONTAB was prepared by Marian Fisher. This program is in The George Washington University Computer Center as CONTABMOD. It provides as output, in addition to the estimates and their logarithms, the design matrices, values of the taus, and the covariance matrix of the taus.

The following programs are applicable to problems as described in the preceding chapter, as well as the "smoothing" or fitting problems. These programs were compiled by John C. Keegel and are in The George Washington University Computer Center.

For its interest we first illustrate the marginal fitting algorithm for the two-way marginals of a three-way table.

1. Iteration, marginal fitting algorithm.

The values of the p^*-table can be computed by an iterative scheme which adjusts the π-table to satisfy successively the given marginal restraints. For a three-way table when all two-way marginals $p(ij.)$, $p(i.k)$, $p(.jk)$ are given, the iteration cycles through

$$
\begin{aligned}
&\overset{(3n+1)}{p}(ijk) = \frac{p(ij.)}{\overset{(3n)}{p}(ij.)}\,\overset{(3n)}{p}(ijk)\\
(1)\quad &\overset{(3n+2)}{p}(ijk) = \frac{p(i.k)}{\overset{(3n+1)}{p}(i.k)}\,\overset{(3n+1)}{p}(ijk)\\
&\overset{(3n+3)}{p}(ijk) = \frac{p(.jk)}{\overset{(3n+2)}{p}(.jk)}\,\overset{(3n+2)}{p}(ijk), \qquad n = 0,1,\ldots
\end{aligned}
$$

where $\overset{(0)}{p}(ijk)$ may be $1/rcd$ or $p_1^*(ijk)$. For a four-way table when all three-way marginals $p(ijk.)$, $p(ij.\ell)$, $p(i.k\ell)$, $p(.jk\ell)$ are given the iteration cycles through

$$p^{(4n+1)}(ijk\ell) = \frac{p(ijk.)}{p^{(4n)}(ijk.)}\, p^{(4n)}(ijk\ell)$$

$$p^{(4n+2)}(ijk\ell) = \frac{p(ij.\ell)}{p^{(4n+1)}(ij.\ell)}\, p^{(4n+1)}(ijk\ell)$$

(2)
$$p^{(4n+3)}(ijk\ell) = \frac{p(i.k\ell)}{p^{(4n+2)}(i.k\ell)}\, p^{(4n+2)}(ijk\ell)$$

$$p^{(4n+4)}(ijk\ell) = \frac{p(.jk\ell)}{p^{(4n+3)}(.jk\ell)}\, p^{(4n+3)}(ijk\ell)$$

where $p^{(o)}(ijk\ell)$ may be $1/rstu$ or $p_1^*(ijk\ell)$ or $p_2^*(ijk\ell)$. It can be shown that the iteration converges to p^* and p^* is unique

Although the above iteration has been in terms of probabilities, in practice it has been found more convenient not to divide everything by n and the iterations are carried out using observed or estimated occurrences $n\pi(ijk\ell)=n/rstu, x(i...), x(ij..)$, etc., $x^*(ijk\ell)=np^*(ijk\ell)$, and in fact our subsequent discussions will be in terms of observed or estimated occurrences. In certain cases when the estimates can be given explicitly in terms of specified marginals the iteration is completed after the first cycle, for example, given the observed one-way marginals $x_1^*(ijk\ell)=$ $x(i...)x(.j..)x(..k.)x(...\ell)/n^3$.

Usually 5 to 7 cycles have been found to be sufficient to obtain agreement between marginals to within 0.001 when more than one cycle is required.

It may be helpful to elaborate somewhat the iterative algorithm given in (1) in terms of occurrences as follows:

1. Start with $x^{(o)}(ijk) = n/r.c.d.$
2. Compute the marginals $x^{(o)}(ij.)$.

3. Adjust $x^{(0)}(ijk)$ by the ratios of the observed marginals $x(ij.)$ to computed marginals $x^{(0)}(ij.)$. The adjusted entries are $x^{(1)}(ijk)$.

4. Compute the marginals $x^{(1)}(i.k)$.

5. Adjust $x^{(1)}(ijk)$ by the ratios of the observed marginals $x(i.k)$ to the computed marginals $x^{(1)}(i.k)$. The adjusted entries are $x^{(2)}(ijk)$.

6. Compute the marginals $x^{(2)}(.jk)$.

7. Adjust $x^{(2)}(ijk)$ by the ratios of the observed marginals $x(.jk)$ to the computed marginals $x^{(2)}(.jk)$. The adjusted entries are $x^{(3)}(ijk)$ and one cycle is completed.

8. Continue the procedure from steps (2) through (7) above using $x^{(3)}(ijk)$ as the starting entries.

9. Continue the process until the three sets of observed marginals agree to within the specified tolerance.

We shall illustrate the iterative algorithm (1) with Cochran's data (1954) for the 2 x 2 x 3 Table 1.

TABLE 1

Data on number of mothers with previous infant losses

Birth Order		Number of mothers with losses	no losses
2	Problem	$x(111) = 20$	$x(121) = 82$
	Control	$x(211) = 10$	$x(221) = 54$
3-4	Problem	$x(112) = 26$	$x(122) = 41$
	Control	$x(212) = 16$	$x(222) = 30$
5+	Problem	$x(113) = 27$	$x(123) = 22$
	Control	$x(213) = 14$	$x(223) = 23$

The sets of observed marginals are

$x(ij.)$	
73	145
40	107

$x(.jk)$		
30	42	41
136	71	45

$x(i.k)$		
102	67	49
64	46	37

We shall find the values of $x_2^*(ijk)$ fitting these marginals.

Using $x^{(0)}(ijk) = 365/(2 \times 2 \times 3) = 30.416$ the sequence of values in Table 2 is obtained. After the first cycle, the "resemblance" between $x^{(3)}(ijk)$ and the final values $x^*(ijk)$ is already evident, and the tolerance requirement of 0.001 is met after 5 cycles.

TABLE 2

ijk	$x^{(1)}(ijk)$	$x^{(2)}(ijk)$	$x^{(3)}(ijk)$	$x^{(4)}(ijk)$	$x^{(5)}(ijk)$	$x^{(6)}(ijk)$	$x^{(14)}(ijk)$	Fitted Data $x^{(15)}(ijk)$	Original Data $x(ijk)$
111	24.333	34.156	19.869	20.079	20.427	20.355....	20.503	20.503	20
211	13.333	17.415	10.130	9.941	9.679	9.645....	9.497	9.497	10
121	48.333	67.844	80.633	80.184	81.573	81.637....	81.497	81.497	82
221	35.667	46.585	55.367	55.791	54.321	54.363....	54.503	54.503	54
112	24.333	22.435	26.959	27.244	27.019	27.047....	27.213	27.213	26
212	13.333	12.517	15.041	14.759	14.937	14.953....	14.787	14.787	16
122	48.333	44.564	40.540	40.314	39.981	39.957....	39.787	39.787	41
222	35.667	33.483	30.369	30.693	31.063	31.043....	31.213	31.213	30
113	24.333	16.408	25.410	25.677	25.074	25.148....	25.284	25.284	27
213	13.333	10.067	15.590	15.299	15.805	15.852....	15.716	15.716	14
123	48.333	32.592	24.639	24.502	23.926	23.862....	23.716	23.716	22
223	35.667	26.932	20.361	20.516	21.195	21.138....	21.284	21.284	23

2. KULLITR 2

KULLITR 2 is the computer program that performs the steps and procedures described in Chapter 5, sections

4. k - samples, 5. An iterative computer algorithm - k - samples.

The program is flexible and can accomodate a variety of experimental situations. In some problems the value of $N\underline{\theta}$ may be determined from some known distribution $\underline{x}$ by $N\underline{\theta} = \underline{C}\ \underline{x}$. In such cases it is not necessary to supply $N\underline{\theta}$ but furnish $\underline{x}$ and the program computes $N\underline{\theta} = \underline{C}\ \underline{x}$. For k - samples it is not necessary for the analyst to compute the appropriate weights c and the matrix $\underline{W}$, since if the user provides the $\underline{B}$ matrix the program computes $\underline{C} = \underline{B}\ \underline{W}^{-1}$. Of course if the user desires to use arbitrary weights not related to the sample sizes one may have to supply the $\underline{C}$ matrix since in such cases the program cannot compute it. In those cases where $N\underline{\theta}$ is provided by "external" hypotheses the program will also compute the minimum modified chi-squared estimates unless the user specifies otherwise. By properly setting appropriate parameters, in the case of complete contingency tables, cells will be coded lexicographically as in other programs for contingency table analysis.

The information that must be supplied to the program is divided into three segments:

(1) Parameters

(2) Factor names

(3) Table data and constraints

The parameter list (1) must be followed by ; . The factor names (2) must be followed by ; .

For segments (1) and (2) the parameter name followed by = followed by the parameter value must be punched on the cards. The parameters must be separated by a blank; however the order of punching the parameters within segment (1) is not important. In segment (3), only numerical values are punched, and the numbers must be separated by blanks. Observed values of zero are punched as 0 but the program treats them automatically as 0.000001.

JCL Instruction

1. // Standard Job Card
2. // EXEC PL1XG, DSN = 'U.ST6630.IRELAND; PROG=KULLITR2
3. //GO.PUNCH DD SYSOUT=B, DCB=(RECFM=F, BLKSIZE=80)
4. //GO.SYSIN DD*

   (1) Parameters

   (2) Factor names

   (3) Table information

5. /*

The cards numbered 2,3,4,5 above make up the EXEC program. Card 3 is necessary only if punched output is desired and may otherwise be omitted. Card 5 follows the parameters, factor names and data and indicates the end of the run. If several jobs are to be run, the parameters , factor names and table information for each may be separated by a blank card and card 5 of the EXEC program placed at the very end.

(1) Parameters - * items are mandatory

PARAMETER	DEFAULT	EFFECT
TITLE = 'NAME'		Identifies the run by name.
(Title name must be in apostrophes)		The RHS must be in ' '.
*OBS = n	0	The number of different "cells"
*CNSTRNT = m	0	All the constraints imposed on the final distribution. If C is an mxn matrix then OBS=n and CNSTRNT = m
CARDS = ' ' B	'0'B	'1'B causes the final distribution to be punched on cards and included as part of the output.
FACTORS = number	1	The number specifies the dimensions of a contingency table and causes the cells to be coded lexicographically.

PARAMETER	DEFAULT	EFFECT
NUMSET = k	1	The number k is the number of samples in the k - sample problem.
INTERNAL = ' 'B	'1'B	'1'B causes $N\underline{\theta}$ to be calculated as $\underline{C}\ \underline{x}$ from a user supplied distribution $\underline{x}$. '0'B implies that $N\underline{\theta}$ will be supplied.
MATDIF = ' 'B	'0'B	'1'B implies that ill conditioned matrices appear and inverts with special procedures. '0'B uses standard procedures and will apply in most cases.
TOL 1 =	.01	TOL 1 is the maximum absolute difference allowed for $N\theta-N\hat{\theta}$ for the first k constraints in a k - sample problem. The tolerance value should not involve more than 6 digits.
TOL 2 =	.01	TOL 2 is the maximum absolute difference allowed for the last r components of $N\theta-N\hat{\theta}$. (See TOL 1)
TOPCOUNT =	15	If the program does not converge (satisfy TOL 1 and TOL 2) after the number of iterations specified

PARAMETER	DEFAULT	EFFECT
		by TOPCOUNT, the tolerances are relaxed by moving the offending tolerance one decimal place to the left in steps of 5 iterations.
BMAT = ' 'B	'0'B	If BMAT = '1'B the program expects only the $\underline{B}$ matrix to be supplied and will compute $\underline{C} = \underline{B}\ \underline{W}^{-1}$. If BMAT = '0'B the $\underline{C}$ matrix must be supplied.
AOK = ' 'B	'1'B	If AOK = '1'B the program computes the minimum modified chi-squared estimate. In this case INTERNAL = '0'B. AOK = '0'B suppresses the minimum modified chi-squared estimate. Should be used if the matrix $\underline{S} = \underline{C}\ \underline{D}_x\ \underline{C}'$ will cause problems in the attempt to invert it.
UNIF = ' 'B	'1'B	This parameter applies only when INTERNAL = '1'B. If UNIF = '1'B, the initial distribution in the iteration will be the uniform distribution and need not be supplied, the program computes it. If UNIF='0'B the initial distribution for the iteration must be supplied.

PARAMETER	DEFAULT	EFFECT
CONDIF = ' 'B	'0'B	CONDIF = '1'B is used if there will be difficulty in convergence particularly when initial distribution is uniform and table is large or cell entries have a wide range. Make TOPCOUNT large if used.
LISTS = ' ' B	'0'B	'1'B lists the S matrix '0'B suppresses the listing of the S matrix.
FIRSTEST = ' 'B	'1'B	'0'B suppresses listing first estimate, '1'B lists the first estimate.

THE PARAMETER LIST MUST BE FOLLOWED BY ;

(2) Factor names

This segment is used only if FACTORS > 1. Each factor name in ' ' is preceded by FACNAME (f) = where f is the factor number. For example, for a 2x2x2 table where the first factor is time, the second factor is cutting and the third factor is mortality we have

FACNAME (1) = 'TIME'

FACNAME (2) = 'CUTTING'

FACNAME (3) = 'MORTALITY'

This segment is optional and if used must terminate with ; . If not used ; must still be supplied only if FACTORS > 1.

(3) Table data and constraints

In this segment only the numerical values must be supplied following the indicated sequence.

Levels. If FACTORS > 1 and we have a 5x6x2 contingency table then the numbers 5 6 2 are punched. If we had a 4x3x2x2x2 contingency table then the numbers 4 3 2 2 2 are punched. If we had a 12x2x2 contingency table then the numbers 12 2 2 are punched. If FACTORS = 1 no values are punched.

PARTITION NUMBERS. If NUMSET > 1, that is, k - samples, then the number of distinct observations or cells in each set must appear. These will add to the number of columns of the

C matrix. For example if NUMSET = 3 with 16 observations in sample 1, 4 observations in sample 2 and 4 observations in sample 3 then the numbers 16 4 4 are punched. (The C matrix has 24 columns). If NUMSET = 4 with two observations in each set then the numbers 2 2 2 2 are punched (the C matrix has 8 columns).

The B or C matrix by rows. The B matrix if BMAT = '1'B, and the C matrix if BMAT = '0'B.

The observed values must be punched in lexicographic order corresponding to the columns of the C matrix. Observed values of zero are punched as 0 but the program automatically treats them as 0.000001.

Nθ. This is supplied only if INTERNAL = '0'B. The number of values must be the same as CNSTRNT = m, that is, the number of rows of the C matrix.

The initial distribution for the iteration. To be supplied only if INTERNAL = '1'B and UNIF = '0'B.

Remarks. In the cases when INTERNAL = 'O'B, the output includes X^2 the minimum modified chi-squared value (the quadratic approximation to 2I(x*:x)) and 2I(x*:x) where x* is the minimum discrimination information estimate and x the observed values. Both X^2 and 2I(x*:x) are asymptotically distributed as chi-squared with r = m-k degrees of freedom.

In the cases when INTERNAL = '1'B, the output includes X^2, the chi-squared approximation to 2I(x*:x) where now x is the initial distribution of the iteration, and also 2I(Z:x*) where Z is the observed distribution. The degrees of freedom for X^2 and 2I(x*:x) are (m-k)-(m'-k) = m-m' where the $\underline{C}$ matrix for the determination of the initial distribution is m' x n. The degrees of freedom for 2I(Z:x*) are n-m where the $\underline{C}$ matrix is m x n. In this case we also have the analysis of information relation

$$\underset{n-m'}{2I(Z:x)} = \underset{m-m'}{2I(x^*:x)} + \underset{n-m}{2I(Z:x^*)}$$

with the associated degrees of freedom. The use of x for the initial and Z for the observed distribution should cause no difficulty in this case as the output specifies "Z IS OBSERVED TABLE AND X IS INITIAL DIST."

3. DARRAT

The generalized iterative scaling procedure described by J. N. Darroch and D. Ratcliff (1972), Generalized iterative scaling for log-linear models, Annals Math. Statist. 43, No. 5, 1470-1480, extends the Deming-Stephan algorithm to cases in which the "design matrix" does not consist only of zeros and ones. A discussion of the procedure and the proof of the convergence of the iteration are to be found in the cited reference. We shall present an exposition of the iteration and a user's guide to the related computer program DARRAT similar to that for KULLITR 2. The basic concepts discussed for the analysis of k-samples are applicable here too. The basic difference with KULLITR 2 is the iterative algorithm used.

For convenience as a frame of reference we give the generalized iterative scaling algorithm as given by Darroch and Ratcliff. Let I be a finite set and let $\underset{\sim}{p} = [p(i); i\varepsilon I, p(i) \geq 0, \sum_{i\varepsilon I} p(i)=1]$ be a probability function on I. Suppose that p is a member of a family of distributions satisfying the constraints.

$$\sum_{i\varepsilon I} b_{si}p(i) = k_s, s = 1,2...d \quad \sum_{i\varepsilon I} p(i) = 1 \qquad (1)$$

where for all s there exist $i\varepsilon I$ such that $b_{si} \neq 0$. The constraints in (1) may be reformulated into the equivalent canonical form

$$\sum_{i\epsilon I} a_{ri}p(i) = h_r, r = 1,2,\ldots c \;, \qquad (2)$$

$$a_{ri} \geqq 0, \quad \sum_{r=1}^{c} a_{ri} = 1, h_r > 0, \quad \sum_{r=1} h_r = 1 \;,$$

by defining

$$a_{si} = t_s(u_s+b_{si}), \text{ all } i, \qquad (3)$$

$$h_s = t_s(u_s+k_s), \; s=1,2,\ldots,d \;,$$

where $u_s \geqq 0$, $t_s > 0$ are chosen to make

$$a_{si} \geqq 0 \text{ and } \sum_{s=1}^{d} a_{si} \leqq 1 \text{ for all } i\epsilon I.$$

If $\sum_{s=1}^{d} a_{si}=1$ for all i define c=d, otherwise define c=d+1 and

let $a_{ci} = 1 - \sum_{s=1}^{d} a_{si}$, $h_c = 1 - \sum_{s=1}^{d} h_s$.

Now let $\underset{\sim}{\pi} = [\pi(i), i\epsilon I, \pi(i)>0, \sum_{i\epsilon I} \pi(i) \leqq 1]$ be a subprobability function on I. The minimum discrimination information estimate $p^*(i)$, $i\epsilon I$, is that member of the family $\underset{\sim}{p}$ satisfying the restraints (2) and minimizing

$$I(\underset{\sim}{p};\underset{\sim}{\pi}) = \sum_{i\epsilon I} p(i) \; \ell n\frac{p(i)}{\pi(i)} \qquad (4)$$

and is given by

$$\ell n \frac{p^*(i)}{\pi(i)} = \sum_{r=1}^{c} a_{ri}\tau_r \;, \qquad (5)$$

where the τ_r are parameters to be determined so that $p^*(i)$ satisfies the constraints (2). The values of $p^*(i)$ may be determined by the convergent iteration

$$p^{(n+1)}(i) = p^{(n)}(i) \prod_{r=1}^{c} \left(\frac{h_r}{h_r^{(n)}} \right)^{a_{ri}}, \qquad n=0,1,2,... \tag{6}$$

$$p^{(0)}(i) = \pi(i) \ , \ h_r^{(n)} = \sum_{i \in I} a_{ri} p^{(n)}(i).$$

We remark that if we use the relations $i=\omega$, $I=\Omega$, $k=\theta$, $b_{si}=b_s(\omega)$, then the constraints in (1) above are the same as the constraints (1.2) in Chapter 5, section 1 or (4.2) in Chapter 5, section 4.

DARRAT is a computer program that performs the steps and procedures of the Darroch-Ratcliff generalized iterative scaling procedure. The iteration will converge at a faster rate if instead of modifying the appropriate design matrix as a unit into the canonical form as above the design matrix is subdivided into blocks of related rows (similar to the notion of marginals) and each block reduced to the canonical form. The user must decide which rows of the design matrix are to be put into a common block and the program then converts these blocks to canonical form for cycles within an iteration. As in KULLITR 2 the program is flexible and can accomodate a variety of experimental situations. In some problems the value of $N\underline{\theta}$ may be determined from some known distribution $\underline{Z}$ by $N\underline{\theta}=\underline{CZ}$. In such cases it is not necessary to supply $N\underline{\theta}$ but furnish $\underline{Z}$ and the program computes the restraints $N\underline{\theta}=\underline{CZ}$. For k-samples it is not necessary for the analyst to compute the appropriate weights and the matrix $\underline{W}$, since if the user provides the $\underline{B}$ matrix the program computes $\underline{C}=\underline{BW}^{-1}$. Of course if the user desires to use arbitrary weights not related to the sample sizes one may have to supply the $\underline{C}$ matrix since in such cases the program cannot compute it. By properly setting appropriate parameters, in the case of complete contingency tables, cells will be coded lexicographically as in other programs for contingency table analysis.

The information that must be supplied to the program is divided into three segments.

(1) Parameters

(2) Factor names

(3) Table data and constraints

The parameter list (1) must be followed by ; . The factor names (2) must be followed by ;. Segment (2) is only used when the parameter FACTORS is > 1. In case FACTORS > 1 and factor names are not used the ; must still be used. In case FACTORS=1 the ; must not be used. For segments (1) and (2) the parameter name followed by = followed by the parameter value must be punched on cards. The parameters must be separated by a blank. However the order of punching the parameters within segment (1) is not important.

In segment (3), only numerical values are punched, and the numbers must be separated by blanks. Observed values of zero are punched as 0 but the program treats them automatically as 0.000001.

JCL Instructions

1. // Standard Job Card
2. // EXEC PL1X6, DSN='U.ST6630. IRELAND', PROG=DARRAT
3. // GO.PUNCH DD SYSOUT = B,DCB = (RECFM = F, BLKSIZE = 80)
4. // GO.SYSIN DD *
   - (1) Parameters
   - (2) Factor names
   - (3) Table data and constraints
5. /*

The cards numbered 2, 3, 4, 5 above make up the EXEC program. Card 3 is necessary only if punched out put is desired and may otherwise be omitted. Card 5 follows the parameters, factor names and table data and constraints and indicates the end of the run. If several jobs are to be run with one execution of DARRAT, the parameters, factor names table data and constraints for each may be separated by a blank card and card 5 of the EXEC program placed at the very end.

(1) Parameters - * items are mandatory

PARAMETER	DEFAULT	EFFECT
TITLE = 'NAME' (Title name must be in apostrophes)		Identifies the run by name. The RHS must be in ' '
*OBS = n	0	The number of different "cells"
*CNSTRNT = m	0	All the constraints imposed on the final distribution. If $\underline{C}$ is an m x n matrix then OBS = n and CNSTRNT = m.
CARDS = ' 'B	'0'B	'1'B causes the final distribution to be punched on cards and included as part of the output.
FACTORS = number	1	The number specifies the dimensions of a contingency table and causes the cells to be coded lexicographically

PARAMETER	DEFAULT	EFFECT
NUMSET = k	1	The number k is the number of samples in the k-sample problem.
INTERNAL = ' 'B	'1'B	'1'B causes the restraints $N\underline{\theta}$ to be calculated as $\underline{C}\ \underline{Z}$ from a user supplied distribution $\underline{Z}$. '0'B implies that $N\underline{\theta}$ will be supplied
TOL 1 =	.01	TOL 1 is the maximum absolute difference allowed for $N\underline{\theta} - N\hat{\underline{\theta}}$ for the first k constraints in a k-sample problem. The tolerance value should not involve more than 6 digits.
TOL 2 =	.01	TOL 2 is the maximum absolute difference allowed for the last r components of $N\underline{\theta} - N\hat{\underline{\theta}}$ (See TOL 1).

PARAMETER	DEFAULT	EFFECT
TOPCOUNT =	50	If the program does not converge (satisfy TOL 1 and TOL 2) after the number of iterations specified by TOPCOUNT, the tolerances are relaxed by moving the offending tolerance one decimal place to the left.
BMAT = ' 'B	'0'B	If BMAT = '1'B the program expects only the $\underline{B}$ matrix to be supplied and will compute $\underline{C} = \underline{B}\,\underline{W}^{-1}$. If BMAT = '0'B the $\underline{C}$ matrix must be supplied.
UNIF = ' 'B	'1'B	This parameter applies only when INTERNAL = '1'B. If UNIF = '1'B, the initial distribution in the iteration will be the uniform distribution and need not be supplied, the program computes it. If UNIF = '0'B the initial distribution for the iteration must be supplied.

PARAMETER	DEFAULT	EFFECT
BLOCKS =	1	Specifies the number of sets of rows of C to be put into canonical form for cycling through the iteration.

THE PARAMETER LIST MUST BE FOLLOWED BY ;

(2) Factor names.

This segment is used only if FACTORS > 1. Each factor name in ' ' is preceded by FACNAME(f) = where f is the factor number. For example, for a 2x2x2 table where the first factor is time, the second factor is cutting, and the third factor is mortality, we have

FACNAME(1) = 'TIME'

FACNAME(2) = 'CUTTING'

FACNAME(3) = 'MORTALITY'

This segment is optional, and if used must terminate with ; If factor names are not used and FACTORS > 1 ; must still be supplied.

(3) Table data and constraints

In this segment only the numerical values must be supplied following the indicated sequence.

a) Levels If FACTORS > 1 and we have a 5x6x2 contingency table, for example, then the numbers 5 6 2 are punched. If we had a 4x3x2x2x2 contingency table, for example, then the numbers 4 3 2 2 2 are punched. If we had a 12x2x2 contingency table, for example, then the numbers 12 2 2 are punched. If FACTORS = 1, no values are punched.

b) <u>BLOCK numbers</u> Omit if BLOCKS = 1. The matrix $\underline{C}$ is divided into a number of sets of rows specified by the parameter BLOCKS in segment (1). The number of rows of $\underline{C}$ in each set (or block) must be specified. These numbers must add to the number of rows in $\underline{C}$ (the value of CNSTRNT). For example if

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

we might specify BLOCKS = 3, treating each row as a unit and punch 1 1 1. There will be three cycles in the iteration. For example if

$$\underline{B} = \begin{matrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \end{matrix}$$

we would specify BLOCKS = 2, treating the first four normalizing restraints as one block and the last row as another block and punch 4 1. For example if

$$\underline{B} = \begin{matrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 & 3 & 0 \end{matrix}$$

we would specify BLOCKS = 3, treating the first four normalizing restraints as one block and each of the fifth and sixth rows as other blocks and we punch 4 1 1. The iteration would proceed through three cycles.

c) Partition numbers. If NUMSET > 1, that is, k-samples, then the number of distinct observations or cells in each set must appear. These will add to the number of columns of the $\underline{C}$ matrix. For example if NUMSET = 3 with 16 observations in sample 1, 4 observations in sample 2 and 4 observations in sample 3 then the numbers 16 4 4 are punched. (The $\underline{C}$ matrix has 24 columns). If NUMSET = 4 with two observations in each set then the numbers 2 2 2 2 are punched. (The $\underline{C}$ matrix has 8 columns).

d) the $\underline{B}$ or $\underline{C}$ matrix by rows. The $\underline{B}$ matrix if BMAT = '1'B, and the $\underline{C}$ matrix if BMAT = '0'B.

e) The observed values must be punched in lexicographic order corresponding to the columns of the $\underline{C}$ matrix. Observed values of zero are punched as 0 but the program automatically treats them as 0.000001.

f) $N\underline{\theta}$. This is supplied only if INTERNAL = '0'B. The number of values must be the same as CNSTRNT = m, that is, the number of rows of the $\underline{C}$ matrix.

g) The initial distribution for the iteration. To be supplied only if INTERNAL = '1'B and UNIF = '0'B.

In the cases when INTERNAL = '0'B, the output includes $2I(x^*:x)$ where x^* is the minimum discrimination information estimate and x the observed values (also the initial distribution of the iteration). $2I(x^*:x)$ has $r = m-k$ D.F.

In the cases when INTERNAL = '1'B, the output includes $2I(x^*:x)$ where x^* is the minimum discrimination information estimate and x is the initial distribution of the iteration and also $2I(z:x^*)$ where z is the observed distribution. The degrees of freedom for $2I(z:x^*)$ are n-m where the $\underline{C}$ matrix is m x n and the degrees of freedom for $2I(x^*:x)$ are $(m-k)-(m'-k) = m-m'$ where the $\underline{C}$ matrix for the determination of the initial distribution is m'xn. In this case we also have the analysis of information relation

$$\underset{n-m'}{2I(z:x)} = \underset{m-m'}{2I(x^*:x)} + \underset{n-m}{2I(z:x^*)}$$

with the associated degrees of freedom. The output carries the statement "Z IS OBSERVED TABLE AND X IS INITIAL DISTRIBUTION."

4. GOKHALE

GOKHALE is a computer program that implements an algorithm presented by D.V. Gokhale (1972), Analysis of Log-linear Models, Journal Royal Statist. Soc. Ser. B. 34, 3, 371-376. The algorithm may be characterized as a method of steepest descent. The algorithm calculates the minimum discrimination information (MDI) estimate that minimizes

(1) $I = \Sigma p_t \ln (p_t/\pi_t)$

subject to the restraints

(2) $\underline{Cp} = \underline{\theta}$.

This is achieved by examining only estimates that satisfy the restraints (2) and following the gradient of (1) in the direction of steepest descent. The procedure converges to the MDI estimate.

The program is designed to be as flexible as possible. It accepts either complete or partial tables and weights the design matrix in the latter case if the user so indicates (similar to KULLITR2 and DARRAT). Constraints are either supplied or the program will calculate them from a user supplied distribution.

In the output are listed the values of the MDI estimate, the values of the parameters in the log-linear model, and the covariance matrix of the values of the parameters. By properly setting appropriate parameters in the program, in the case of complete contingency tables, cells will be coded lexicographically as in other programs for contingency table analysis.

The information that must be supplied to the program is divided into three segments:

(1) Parameters

(2) Factor names

(3) Table data and constraints.

The parameter list (1) must be followed by ;. The factor names (2) must be followed by ;. Segment (2) is used only when the parameter FACTORS is greater than 1. In cases FACTORS >1 and factor names are not used the ; must still be used. In case FACTORS=1 the ; must not be used. For segments (1) and (2) the parameter name followed by = followed by the parameter value must be punched on cards. The parameters must be separated by a blank space. The order of punching the parameters within segment (1) is not important. In segment (3) only numerical values are punched, and the numbers must be separated by blank spaces. Observed values of zero are punched as 0 but the program treats them automatically as 0.000001.

JCL Instructions

1. // Standard Job Card
2. // EXEC PL1X6, DSN='U.ST6630.IRELAND', PROG=GOKHALE
3. //GO.PUNCH DD SYSOUT=B,DCB=(RECFM=F,BLKSIZE=80)
4. // GO.SYSIN DD *

    (1) Parameters

    (2) Factor names

    (3) Table data and constraints

5. /*

The cards numbered 2,3,4,5 above make up the EXEC program. Card 3 is necessary only if punched output is desired and may otherwise be omitted. Card 5 follows the parameters, factor names and the data and constraints and indicates the end of the run. If several jobs are to be run with one execution of GOKHALE, the parameters, factor names, table data and constraints for each may be separated by a blank card and card 5 of the EXEC program placed at the very end.

(1) Parameters-- * items are mandatory

PARAMETER	DEFAULT	EFFECT
TITLE='NAME' (Title name must be in apostrophes)		Identifies the run by name. The RHS must be in ' '.
*OBS=n	0	The number of different "cells."
*CNSTRNT=m	0	All the constraints imposed on the final distribution. If $\underline{C}$ is an m x n matrix then OBS=n and CNSTRNT=m.
EPZ=number	.0001	When the length of the gradient becomes smaller than EPZ, the algorithm is deemed to have converged.
CARDS=' 'B	'0'B	'1'B causes the final distribution to be punched on cards and included as part of the output.
FACTORS=number	1	The number specifies the dimensions of a contingency table and causes the cells to be coded lexicographically.
NUMSET=k	1	The number k is the number of samples in the k-sample problem.

INTERNAL=' 'B	'1'B	'1'B causes the restraints $\underline{N\theta}$ to be calculated as $\underline{CZ}$ from a user supplied distribution $\underline{Z}$. '0'B implies that $\underline{N\theta}$ will be supplied.
TOPCOUNT=	36	If the program does not converge (satisfy EPZ) after the number of iterations specified by TOPCOUNT, then EPZ is multiplied by 10.
BMAT=' 'B	'0'B	If BMAT='1'B the program expects only the $\underline{B}$ matrix to be supplied and will compute the $\underline{C}$ matrix by weighting the $\underline{B}$-matrix properly. If BMAT='0'B the $\underline{C}$-matrix must be supplied.
UNIF=' 'B	'1'B	This parameter applies only when INTERNAL='1'B. If UNIF='1'B, the initial distribution in the iteration will be the uniform distribution and need not be supplied, the program computes it. If UNIF='0'B the initial distribution for the iteration must be supplied.
MATDIF=' 'B	'0'B	'1'B implies that ill conditioned matrices may appear and inverts with special procedures. '0'B uses standard procedured and will apply in most cases.

THE PARAMETER LIST MUST BE FOLLOWED BY A SEMI_COLON ;

(2) Factor Names

This segment is used only if FACTORS>1. Each factor name in ' ' is preceded by FACNAME(f)= where f is the factor number. For example, for a 2x2x2 table where the first factor is time, the second factor is cutting and the third factor is mortality we have

FACNAME(1)='TIME' FACNAME(2)='CUTTING' FACNAME(3)='MORTALITY'

This segment is optional and if used must terminate with ;. If not used ; must still be supplied only if FACTORS>1. If FACTORS=1 no factor names are given and no semi-colon is punched.

(3) Table data and constraints

In this segment only the numerical values must be supplied following the indicated sequence.

LEVELS. If FACTORS>1 and we have a 5x6x2 contingency table then the numbers 5 6 2 are punched. If we had a 4x3x2x2x2 contingency table then the numbers 4 3 2 2 2 are punched. If we had a 12x2x2 contingency table then the numbers 12 2 2 are punched. If FACTORS=1 no values are punched.

PARTITION NUMBERS If NUMSET>1, that is, k-samples, then the number of distinct observations or cells in each set must appear. These will add to the number of columns of the C-matrix. For example, if NUMSET=3 with 16 observations in sample 1, 4 observations in sample 2, and 4 observations in sample 3 then the numbers 16 4 4 are punched. (The C-matrix has 24 columns). If NUMSET=4 with two observations in each set then the numbers 2 2 2 2 are punched (the C-matrix has 8 columns).

The $\underline{B}$ or $\underline{C}$ matrix by rows. The $\underline{B}$-matrix if BMAT='1'B, and the $\underline{C}$-matrix if BMAT='0'B.

The observed values must be punched in lexicographic order corresponding to the columns of the $\underline{C}$-matrix. Observed values of zero are punched as 0 but the program automatically treats them as 0.000001.

N$\underline{\theta}$. This is supplied only if INTERNAL='0'B. The number of values must be the same as CNSTRNT=m, that is, the number of rows of the $\underline{C}$-matrix.

The initial distribution for the iteration. To be supplied only if INTERNAL='1'B and UNIF='0'B.

Remarks In the cases when INTERNAL='0'B, the output includes X^2, the minimum modified chi-squared value (the quadratic approximation to $2I(x^*:x)$) and the minimum modified chi-squared estimates which are used as the initial values in the iteration, since they satisfy the constraints. Both X^2 and $2I(x^*:x)$ where x^* is the MDI estimate and x the observed values are asymptotically distributed as chi-squared with r=m-k degrees of freedom.

5. MATGEN

MATGEN is a computer program that generates and provides punched card output of design matrices, the B or C matrices, for use as input for the programs KULLITR2, DARRAT, GOKHALE. We recall that the program CONTABMOD generates the design matrices for models fitting various sets of observed marginals for use in computing the tau parameters and their covariance matrix as part of the program output.

By considering the string of the successive rows of the matrix as made up of vectors of appropriate sizes it will usually be found that a relatively small number of different vectors have to be assembled to compose the matrix.

The input to MATGEN consists of two segments. The first contains parameter values and these must include parameter name followed by =. The second segment consists of a set of numerical values that must be entered in a prescribed order.

(1) Parameter List

PARAMETER	DEFAULT	EFFECT
ROWS=m	1	m is the number of rows of the m x n matrix
COLS=n	1	n is the number of columns of the m x n matrix
VECTSIZES=k	1	k is the number of different size basic generating vectors

THIS PARAMETER LIST MUST TERMINATE WITH ;

(2) Numerical Values

NUMBER SIZE LIST. This is a list of ordered pairs of numbers. The first of the pair is the number of basic vectors whose size (length) is given by the second of the pair. For example

2 4 3 2

means two basic vectors of length four and three basic vectors of length two. For this case VECTSIZES=2.

BASIC VECTOR LIST. The vectors must be entered according to the lengths specified in the NUMBER SIZE LIST. All vectors of length four would be entered first followed by the vectors of length two.

GENERATION LIST. This list consists of pairs of numbers. The first component of the pair is the number of successive occurrences of the vector whose ordinal number in the basic vector list is the second component of the pair.

JCL Instructions

1. // Standard Job Card
2. //#EXEC#PL1X6,DSN='U.ST6630.IRELAND',PROG=MATGEN
3. //GO.PUNCH#DD#SYSOUT=B,DCB=(RECFM=FB,BLKSIZE=80)
4. //GO.SYSIN#DD#*
5. /*

Note that # represents a blank space. Card 5 follows the numerical values and terminates the program.

Example. Suppose we want to generate the following matrix (Of course we would not use the program for such a matrix but would punch it directly. However, it will illustrate the procedure.)

```
1 1 0 0|1 1 0 0|
1 0 1 0|1 0 1 0|
0 0|1 0 1 0|0 0|
0 0|0 0|1 1 0 0|
1 0|1 1 0 0|1 1|
1 1|1 1|0 0|0 0|
0 0|0 0|1 1|1 1|
```

EXEC Cards

```
ROWS=7  COLS=8  VECTSIZES=2 ;
2 4   3 2
1 1 0 0   1 0 1 0
1 1   0 0   1 0


2 1
2 2
1 4   1 2 3 4   1 1
1 5  1 1   3 3
4 4  2 3
/*
```

Note that the vecters in ordinal number are

1st	1 1 0 0
2nd	1 0 1 0
3rd	1 1
4th	0 0
5th	1 0

It is not necessary that the elements of the matrix consist only of 0's and 1's. Negative values may occur also. A vector may be

.833333 .833333 0 0

or

0 -1 0 -1

or

0 1 2

etc. depending on the problem requirement.

7. No Interaction on a Linear Scale in a 2 x 2 x 2 Contingency Table.

1. Minimum discrimination information estimation.

Consider the population 2 x 2 x 2 contingency table 1

Table 1

	B j=1		β j=2	
	C k=1	γ k=2	C k=1	γ k=2
i=1 A	P(111)	P(112)	P(121)	P(122)
i=2 α	P(211)	P(212)	P(221)	P(222)

The experimental procedure selects a fixed number of observations under the four possible combinations of the factors (B,β), (C,γ) and determines the number of occurrences of (A,α) for each case. In effect then the procedure is examining four binomials with

(1) $P(1jk) + P(2jk) = 1,\ j=1,2, k = 1,2.$

The corresponding observed values are shown in table 2. It is desired to test whether the observed values are consistent with a null hypothesis of no interaction on a linear scale,

Table 2

	j=1		j=2	
	k=1	k=2	k=1	k=2
i=1	x(111)	x(112)	x(121)	x(122)
i=2	x(211)	x(212)	x(221)	x(222)
	x(.11)	x(.12)	x(.21)	x(.22)

that is

(2) H_0: $P(111) - P(112) = P(121) - P(122)$

or $P(111) - P(112) - P(121) + P(122) = 0.$

We shall determine estimates for the cell entries subject to the null hypothesis and compare the estimated and observed values. The estimated table is given in table 3 where the λ's are to be determined.

Table 3

	j=1		j=2	
	k=1	k=2	k=1	k=2
i=1	$x(111)+\lambda_1$	$x(112)+\lambda_2$	$x(121)+\lambda_3$	$x(122)+\lambda_4$
i=2	$x(211)-\lambda_1$	$x(212)-\lambda_2$	$x(221)-\lambda_3$	$x(222)-\lambda_4$
	x(.11)	x(.12)	x(.21)	x(.22)

We shall use the principle of minimum discrimination information estimation and thus determine the λ's which minimize

$$
(3)\begin{cases}
(x(111)+\lambda_1)\,\ell n\dfrac{x(111)+\lambda_1}{x(111)}+(x(211)-\lambda_1)\,\ell n\dfrac{x(211)-\lambda_1}{x(211)} \\
+(x(112)+\lambda_2)\,\ell n\dfrac{x(112)+\lambda_2}{x(112)}+(x(212)-\lambda_2)\,\ell n\dfrac{x(212)-\lambda_2}{x(212)} \\
+(x(121)+\lambda_3)\,\ell n\dfrac{x(121)+\lambda_3}{(121)}+(x(212)-\lambda_3)\,\ell n\dfrac{x(212)-\lambda_3}{x(212)} \\
+(x(122)+\lambda_4)\,\ell n\dfrac{x(122)+\lambda_4}{x(122)}+(x(222)-\lambda_4)\,\ell n\dfrac{x(222)-\lambda_4}{x(222)} \\
+\tau\left(\dfrac{x(111)+\lambda_1}{x(.11)}-\dfrac{x(112)+\lambda_2}{x(.12)}-\dfrac{x(121)+\lambda_3}{x(.21)}+\dfrac{x(122)+\lambda_4}{x(122)}\right)
\end{cases},
$$

where τ is a Lagrange undetermined multiplier and (2) reflected by the condition

$$
(4)\quad \frac{x(111)+\lambda_1}{x(.11)}-\frac{x(112)+\lambda_2}{x(.12)}-\frac{x(121)+\lambda_3}{x(.21)}+\frac{x(122)+\lambda_4}{x(.22)}=0.
$$

Differentiating (3) with respect to $\lambda_1,\ldots,\lambda_4$ leads to the "normal" equations

$$
(5)\begin{cases}
\ell n\dfrac{x(111)+\lambda_1}{x(111)}-\ell n\dfrac{x(211)-\lambda_1}{x(211)}+\dfrac{\tau}{x(.11)}=0\,, \\
\ell n\dfrac{x(112)+\lambda_2}{x(112)}-\ell n\dfrac{x(212)-\lambda_2}{x(212)}-\dfrac{\tau}{x(.12)}=0\,, \\
\ell n\dfrac{x(121)+\lambda_3}{x(121)}-\ell n\dfrac{x(221)-\lambda_3}{x(221)}-\dfrac{\tau}{x(.21)}=0\,, \\
\ell n\dfrac{x(122)+\lambda_4}{x(122)}-\ell n\dfrac{x(222)-\lambda_4}{x(222)}+\dfrac{\tau}{x(.22)}=0.
\end{cases}
$$

There are a number of different iterative approaches to determine the solution to (5) but our interest here is to examine the relation of an approximate solution to other proposed methods.

Assuming that the ratio of the λ's to the observed values are small, we use the approximations

$$\ell n \frac{x(111)+\lambda_1}{x(111)} \approx \frac{\lambda_1}{x(111)}, \quad \ell n \frac{x(211)-\lambda_1}{x(211)} \approx -\frac{\lambda_1}{x(211)}, \text{ etc.}$$

in (5) and get

$$(6)\quad \begin{cases} \dfrac{\lambda_1}{x(111)} + \dfrac{\lambda_1}{x(211)} + \dfrac{\tau}{x(.11)} = 0 = \lambda_1 \dfrac{x(.11)}{x(111)\ x(211)} + \dfrac{\tau}{x(.11)}, \\[2ex] \dfrac{\lambda_2}{x(112)} + \dfrac{\lambda_2}{x(212)} - \dfrac{\tau}{x(.12)} = 0 = \lambda_2 \dfrac{x(.12)}{x(112)\ x(212)} - \dfrac{\tau}{x(.12)}, \\[2ex] \dfrac{\lambda_3}{x(121)} + \dfrac{\lambda_3}{x(221)} - \dfrac{\tau}{x(.21)} = 0 = \lambda_3 \dfrac{x(.21)}{x(121)\ x(221)} - \dfrac{\tau}{x(.21)}, \\[2ex] \dfrac{\lambda_4}{x(122)} + \dfrac{\lambda_4}{x(222)} + \dfrac{\tau}{x(.22)} = 0 = \lambda_4 \dfrac{x(.22)}{x(122)\ x(222)} + \dfrac{\tau}{x(.22)}. \end{cases}$$

From (6) and (4) we have, introducing the notation $x(1ij) = x(.ij)p(ij)$, $x(2ij)= x(.ij)q(ij)$, $p(ij) + q(ij) = 1$,

$$(7)\quad \begin{cases} \lambda_1 = -\dfrac{x(111)x(211)}{(x(.11))^2}\tau = -\ p(11)q(11)\tau, \\[2ex] \lambda_2 = \dfrac{x(112)x(212)}{(x(.12))^2}\tau = p(12)q(12)\tau, \\[2ex] \lambda_3 = \dfrac{x(121)x(221)}{(x(.21))^2}\tau = p(21)q(21)\tau, \\[2ex] \lambda_4 = -\dfrac{x(122)x(222)}{(x(.22))^2}\tau = -\ p(22)q(22)\tau, \\[2ex] \tau = \dfrac{p(11) - p(12) - p(21) + p(22)}{\dfrac{p(11)q(11)}{x(.11)} + \dfrac{p(12)q(12)}{x(.12)} + \dfrac{p(21)q(21)}{x(.21)} + \dfrac{p(22)q(22)}{x(.22)}}. \end{cases}$$

Let us write

$$\text{(8)}\quad \begin{aligned} x^*(111) &= x(111) + \lambda_1,\ x^*(211) = x(211) - \lambda_1 , \\ x^*(112) &= x(112) + \lambda_2,\ x^*(212) = x(212) - \lambda_2 , \end{aligned}$$

etc.

where the λ's satisfy (5).

If we also use the approximations

$$\text{(9)}\quad 2\{(x(111)+\lambda_1)\ \ln \frac{x(111)+\lambda_1}{x(111)} + (x(211)-\lambda_1)\ \ln \frac{x(211)-\lambda_1}{x(211)}\}$$

$$=\lambda_1^2\left(\frac{1}{x(111)} + \frac{1}{x(211)}\right) = \lambda_1^2 \frac{x(.11)}{x(111)x(211)} = \frac{\lambda_1^2}{x(.11)p(11)q(11)},$$

then we get for the minimum discrimination information statistic

$$\text{(10)}\quad 2I(x^*:x)=2\sum\sum\sum x^*(ijk)\ln \frac{x^*(ijk)}{x(ijk)}$$

$$\simeq\tau^2\left\{\frac{p(11)q(12)}{x(.11)} + \frac{p(12)q(12)}{x(.12)} + \frac{p(21)q(21)}{x(.21)} + \frac{p(22)q(22)}{x(.22)}\right\}$$

$$= \frac{(p(11)-p(12)-p(21)+p(22))^2}{\frac{p(11)q(11)}{x(.11)} + \frac{p(12)q(12)}{x(.12)} + \frac{p(21)q(21)}{x(.21)} + \frac{p(22)q(22)}{x(.22)}}$$

$$=\lambda_1^2\left(\frac{1}{x(111)} + \frac{1}{x(211)}\right)+\lambda_2^2\left(\frac{1}{x(112)} + \frac{1}{x(212)}\right)+\ldots+\lambda_4^2\left(\frac{1}{x(122)} + \frac{1}{x(222)}\right).$$

Note that the last value in (10) is the modified Neyman χ^2

$$\text{(11)}\quad \chi^2 = \sum \frac{(\text{obs}-\text{exp})^2}{\text{obs}}$$

and indeed the equations in (6) are those to determine the minimum modified χ^2 estimates. The next to last value in (10) is the statistic given by Bhapkar and Koch (1968, p. 116) based on a criterion due to Wald. The square root of this value is the statistic used by Snedecor and Cochran (1967, p. 496).

In accordance with the minimum discrimination information theorem the log-linear representation for x*(ijk) is given graphically as in figure 1 where the interpretation is

$$(12)\begin{cases} \ln \dfrac{x^*(111)}{x(111)} = L_1 + \tau/x(.11) \ , \\[2ex] \ln \dfrac{x^*(211)}{x(111)} = L_1 \ , \\[2ex] \ln \dfrac{x^*(112)}{x(112)} = L_2 - \tau/x(.12) \ , \\[2ex] \ln \dfrac{x^*(212)}{x(212)} = L_2 \ , \\ \quad \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \\ \ln \dfrac{x^*(222)}{x(222)} = L_4 \ . \end{cases}$$

Recalling (8) we see that (12) in fact leads to (5). If we write

$$(13)\begin{cases} \theta^* = \dfrac{x^*(111)}{x(.11)} - \dfrac{x^*(112)}{x(.12)} - \dfrac{x^*(121)}{x(.21)} + \dfrac{x^*(122)}{x(.21)} = p^*(11) - p^*(12) - p^*(21) + p^*(22) \ , \\[3ex] \theta = \dfrac{x(111)}{x(.11)} - \dfrac{x(112)}{x(.12)} - \dfrac{x(121)}{x(.21)} + \dfrac{x(122)}{x(.21)} = p(11) - p(12) - p(21) + p(22) \ , \end{cases}$$

then as shown in Kullback (1959, p. 101-106)

(14) $2I(x^*;x) \approx (\theta^*-\theta)^2/\sigma^2$,

where σ^2 is determined as follows. Let $\underline{T}$ denote the 8 x 5 matrix in figure 1, that is,

$$(15)\quad \underline{T} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1/x(.11) \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1/x(.12) \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1/x(.21) \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1/x(.22) \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

and $\underline{D}_x$ the 8 x 8 diagonal matrix with entries x(ijk), that is,

$$(16)\quad \underline{D}_x = \begin{pmatrix} x(111) & 0 & . & . & . & . & . & 0 \\ 0 & x(211) & & & & & & . \\ . & & x(112) & & & & & . \\ . & & & x(212) & & & & . \\ . & & & & x(121) & & & . \\ . & & & & & x(221) & & . \\ . & & & & & & x(122) & . \\ 0 & . & . & . & . & . & . & x(222) \end{pmatrix} .$$

Compute the 5 x 5 matrix $\underline{S} = \underline{T}'\underline{D}_x\underline{T}$ and partition it as follows

$$(17)\quad \underline{S} = \begin{pmatrix} \underline{S}_{11} & \underline{S}_{12} \\ \underline{S}_{21} & \underline{S}_{22} \end{pmatrix} , \ \underline{S}_{11} \text{ is } 4 \times 4,\ \underline{S}_{22} \text{ is } 1 \times 1,$$
$$\underline{S}_{21} = \underline{S}_{12}' \text{ is } 1 \times 4 ,$$

then σ^2 in (14) is given by

$$(18) \quad \sigma^2 = \underline{S}_{22} - \underline{S}_{21}\underline{S}_{11}^{-1}\underline{S}_{12} .$$

It may be verified that this results in

$$(19) \quad \sigma^2 = \frac{x(111)x(211)}{(x(.11))^3} + \frac{x(112)x(212)}{(x(.12))^3} + \frac{x(121)x(221)}{(x(.21))^3} + \frac{x(122)x(222)}{(x(.22))^3}$$

$$= \frac{p(11)q(11)}{x(.11)} + \frac{p(12)q(12)}{x(.12)} + \frac{p(21)q(21)}{x(.21)} + \frac{p(22)q(22)}{x(.22)} .$$

But θ^* in (13) is zero and we see that (14) is indeed the next-to-last value in (10). It is interesting to note that $2I(x^*:x)$ can be approximated without necessarily computing the values of $x^*(ijk)$.

Figure 1

i	j	k	L_1	L_2	L_3	L_4	τ
1	1	1	1				1/x(.11)
2	1	1	1				
1	1	2		1			-1/x(.12)
2	1	2		1			
1	2	1			1		-1/x(.21)
2	2	1			1		
1	2	2				1	1/x(.22)
2	2	2				1	

2. Example, Root ~~EXAMPLE~~ cuttings,

We shall illustrate the preceding discussion by Bartlett's data on root cuttings used also as an example by Snedecor and Cochran (1967), Bhapkar and Koch (1968), Berkson (1972).

The following from Bartlett (1935), ~~Contingency table interactions, J.Roy Statist Soc. Suppl., 2, 248-252,~~ who refers to data from Hoblyn and Palmer, is the result of an experiment designed to investigate the propogation of plum root stocks from root cuttings. There were 240 cuttings for each of the four treatments.

	At Once j=1		In Spring j=2	
	Long k=1	Short k=2	Long k=1	Short k=2
Dead i=1	84	133	156	209
Alive i=2	156	107	84	31
	240	240	240	240

From (7) it is found that $\tau=4(240)^2/46918$, $\lambda_1=-1.117183$, $\lambda_2=1.213266$, $\lambda_3=1.117183$, $\lambda_4=-0.552368$, and hence the minimum modified χ^2 estimates are:

	j=1		j=2	
	k=1	k=2	k=1	k=2
i=1	82.882817	134.213266	157.117183	208.447632
i=2	157.117183	105.786734	82.882817	31.552368

From (10) it is found that $2I(x^*:x)$ is approximately 0.08184492, 1 degree of freedom.

Bartlett's root cutting data was also used to illustrate other computer programs. The input cards for KULLITR2 were

```
TITLE = 'BARTLETT''S ROOT CUTTINGS'
TOL1 = .001   TOL2 = .001   CNSTRNTS = 5   OBS = 8
BMAT = '1'B   INTERNAL = '0'B   NUMSET = 4   FACTORS = 3 ;
FACNAME(1) = 'TIME'  FACNAME(2) = 'CUTTING'  FACNAME(3) = 'MORTALITY' ;
 2   2   2
 2   2   2   2
 1   1   0   0   0   0   0   0
 0   0   1   1   0   0   0   0
 0   0   0   0   1   1   0   0
 0   0   0   0   0   0   1   1
 1   0  -1   0  -1   0   1   0
84  156  133  107  156  84  209  31
960  960  960  960   0
```

Note that the computer output gives $2I(x^*:x) = 0.080972$ and the minimum modified chi-squared as $X^2 = 0.081845$. The computer output follows.

BARTLETT'S ROOT CUTTINGS

3 FACTOR TABLE:TIME*CUTTING*MORTALITY

B MATRIX

	1	2	3	4	5	6	7	8
1	1	1	0	0	0	0	0	0
2	0	0	1	1	0	0	0	0
3	0	0	0	0	1	1	0	0
4	0	0	0	0	0	0	1	1
5	1	0	-1	0	-1	0	1	0

WEIGHT(1)= 0.250000
WEIGHT(2)= 0.250000
WEIGHT(3)= 0.250000
WEIGHT(4)= 0.250000

INV WEIGHT(1)= 4.000000
INV_WEIGHT(2)= 4.000000
INV_WEIGHT(3)= 4.000000
INV_WEIGHT(4)= 4.000000

C DESIGN MATRIX

	1	2	3	4	5	6	7	8
1	4	4	0	0	0	0	0	0
2	0	0	4	4	0	0	0	0
3	0	0	0	0	4	4	0	0
4	0	0	0	0	0	0	4	4
5	4	0	-4	0	-4	0	4	0

OBSERVED VALUES

1	1	1	X(1)= 84.000000	LN_X(1)= 4.430817
1	1	2	X(2)= 156.000000	LN_X(2)= 5.049856
1	2	1	X(3)= 133.000000	LN_X(3)= 4.890349
1	2	2	X(4)= 107.000000	LN_X(4)= 4.672829
2	1	1	X(5)= 156.000000	LN_X(5)= 5.049856
2	1	2	X(6)= 84.000000	LN_X(6)= 4.430817
2	2	1	X(7)= 209.000000	LN_X(7)= 5.342334
2	2	2	X(8)= 31.000000	LN_X(8)= 3.433987

CONSTRAINTS
NTHETA(1)= 960.000000
NTHETA(2)= 960.000000
NTHETA(3)= 960.000000
NTHETA(4)= 960.000000
NTHETA(5)= 0.000000

```
ESTIMATE OF NTHETA AT COUNT=    1
NTHAT(1)= 960.000000
NTHAT(2)= 960.000000
NTHAT(3)= 960.000000
NTHAT(4)= 960.000000
NTHAT(5)=  16.000000
```

S

	1	2	3	4	5
1	3840	0	0	0	1344
2	0	3840	0	0	-2128
3	0	0	3840	0	-2496
4	0	0	0	3840	3344
5	1344	-2128	-2496	3344	9312

S22.1

	1
1	3127.866943

S22.1_INV

	1
1	0.000320

```
DELTA(1)=  0.000000
DELTA(2)=  0.000000
DELTA(3)=  0.000000
DELTA(4)=  0.000000
DELTA(5)=-16.000000
```

```
XSQ=     0.081845
```

ESTIMATE OF X AT COUNT=   1

1	1	1	XSTAR(1)=	82.886185	LN_XSTAR(1)=	4.417468
1	1	2	XSTAR(2)=	157.113739	LN_XSTAR(2)=	5.056970
1	2	1	XSTAR(3)=	134.211699	LN_XSTAR(3)=	4.899420
1	2	2	XSTAR(4)=	105.788101	LN_XSTAR(4)=	4.661438
2	1	1	XSTAR(5)=	157.113739	LN_XSTAR(5)=	5.056970
2	1	2	XSTAR(6)=	82.886261	LN_XSTAR(6)=	4.417469
2	2	1	XSTAR(7)=	208.443344	LN_XSTAR(7)=	5.339667
2	2	2	XSTAR(8)=	31.556580	LN_XSTAR(8)=	3.451762

ZI(XSTAR:X)= 0.081143

TAU(1)=-0.005115

ESTIMATE OF X AT COUNT= 3

1	1	1	XSTAR(1)=	82.885071	LN_XSTAR(1)=	4.417455
1	1	2	XSTAR(2)=	157.114777	LN_XSTAR(2)=	5.056976
1	2	1	XSTAR(3)=	134.213089	LN_XSTAR(3)=	4.899425
1	2	2	XSTAR(4)=	105.786911	LN_XSTAR(4)=	4.661427
2	1	1	XSTAR(5)=	157.114822	LN_XSTAR(5)=	5.056977
2	1	2	XSTAR(6)=	82.885178	LN_XSTAR(6)=	4.417457
2	2	1	XSTAR(7)=	208.442088	LN_XSTAR(7)=	5.339664
2	2	2	XSTAR(8)=	31.557114	LN_XSTAR(8)=	3.451796

ZI(XSTAR:X)= 0.080572

TAU(1)=-0.005120

ESTIMATE OF NTHETA AT COUNT= 3

NTHAT(1)= 959.999512
NTHAT(2)= 960.000000
NTHAT(3)= 960.000000
NTHAT(4)= 960.000000
NTHAT(5)= 0.000488

S

	1	2	3	4	
1	3839.998779	0.000000	0.000000	0.000000	1326.16235
2	0.000000	3840.000000	0.000000	0.000000	-2147.40942
3	0.000000	0.000000	3840.000000	0.000000	-2513.83715
4	0.000000	0.000000	0.000000	3840.000000	3335.08018
5	1326.162354	-2147.409424	-2513.837158	3335.080162	9322.49218

S22.1

	1
1	3121.385498

S22.1.INV

	1
1	0.000320

DELTA(1)= 0.000488
DELTA(2)= 0.000000
DELTA(3)= 0.000000
DELTA(4)= 0.000000
DELTA(5)=-0.000488

1	1	1	OUTLIER(1)= 0.014898
1	1	2	OUTLIER(2)= 0.007938
1	2	1	OUTLIER(3)= 0.011014
1	2	2	OUTLIER(4)= 0.013832
2	1	1	OUTLIER(5)= 0.007938
2	1	2	OUTLIER(6)= 0.014895
2	2	1	OUTLIER(7)= 0.001488
2	2	2	OUTLIER(8)= 0.009923

ITERATIONS= 3

TOL1=0.0010   TOL2=0.0010

S_INV

	1	2	3	4	5
1	0.000300	-0.000062	-0.000073	0.000097	-0.000112
2	-0.000062	0.000359	0.000115	-0.000154	0.000177
3	-0.000073	0.000115	0.000395	-0.000181	0.000208
4	0.000097	-0.000154	-0.000181	0.000503	-0.000276
5	-0.000112	0.000177	0.000208	-0.000276	0.000320

LAMBDA(1)= 0.001790
LAMBDA(2)=-0.002835
LAMBDA(3)=-0.003325
LAMBDA(4)= 0.004455
LAMBDA(5)=-0.005115

MU(1)=-0.013300
MU(2)= 0.007161
MU(3)= 0.009122
MU(4)=-0.011339
MU(5)= 0.007161
MU(6)=-0.013300
MU(7)=-0.002643

MU(8)= 0.017818

XSQ= 0.081845

MINIMUM MODIFIED CHI SQ ESTIMATE

1	1	1	XHAT(1)=	82.882813	LN_XHAT(1)=	4.417428
1	1	2	XHAT(2)=	157.117172	LN_XHAT(2)=	5.056992
1	2	1	XHAT(3)=	134.213257	LN_XHAT(3)=	4.899430
1	2	2	XHAT(4)=	105.786728	LN_XHAT(4)=	4.661425
2	1	1	XHAT(5)=	157.117172	LN_XHAT(5)=	5.056992
2	1	2	XHAT(6)=	82.882813	LN_XHAT(6)=	4.417428
2	2	1	XHAT(7)=	208.447632	LN_XHAT(7)=	5.339687
2	2	2	XHAT(8)=	31.552353	LN_XHAT(8)=	3.451648

2I(XHAT:X)= 0.081507

We also illustrate the first two iterative steps in the Darroch-Ratcliff iterative procedure applied to Bartlett's root cutting data.

$\underline{B}$ =

111	211	112	212	121	221	122	222	θ
1	1							1
		1	1					1
				1	1			1
						1	1	1
1		-1		-1		1		0

$\underline{Bp} = \underline{\theta}$

$\underline{BW}^{-1}$ = C =

111	211	112	212	121	221	122	222	θ
4	4							1
		4	4					1
				4	4			1
						4	4	1
4		-4		-4		4		0

$N_1 = N_2 = N_3 = N_4 = 240$

$N = 960$

$w_1 = w_2 = w_3 = w_4 = 240/960 = 1/4$

$v_1 = v_2 = v_3 = v_4 = 4$

$C = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} \begin{matrix} 4\text{x}8 \\ 1\text{x}8 \end{matrix}$

ω	1	2	3	4	5	6	7	8	
	111	211	112	212	121	221	122	222	
a_1	1	1							$1/4 = h_1$
a_2			1	1					$1/4 = h_2$
a_3					1	1			$1/4 = h_3$
a_4							1	1	$1/4 = h_4$

4		-4		-4		4		0
	4		-4		-4		4	0
8	4	0	4	0	4	8	4	4
4	8	4	0	4	0	4	8	4
0	0	8	8	8	8	0	0	4

ω	1	2	3	4	5	6	7	8	
	111	211	112	212	121	221	122	222	
α_1	2/3	1/3	0	1/3	0	1/3	2/3	1/3	$1/3 = k_1$
α_2	1/3	2/3	1/3	0	1/3	0	1/3	2/3	$1/3 = k_2$
α_3	0	0	2/3	2/3	2/3	2/3	0	0	$1/3 = k_3$

BARTLETT'S ROOT CUTTING DATA - NO INTERATION LINEAR SCALE

DARROCH - RATCLIFF ITERATION - INITIAL VALUE OBSERVED

[illegible]	$x^{(0)}$	α_1	α_2	α_3	$k_1^{(0)}$	$k_2^{(0)}$	$k_3^{(0)}$	a_1	a_2	a_3	a_4
1	84	2	1	0	168	84	0	1	0	0	0
2	156	1	2	0	156	312	0	1	0	0	0
3	133	0	1	2	0	133	266	0	1	0	0
4	107	1	0	2	107	0	214	0	1	0	0
5	156	0	1	2	0	156	312	0	0	1	0
6	84	1	0	2	84	0	168	0	0	1	0
7	209	2	1	0	418	209	0	0	0	0	1
8	31	1	2	0	31	62	0	0	0	0	1
	all divided by 3				964	956	960				

$$\ln x^{(1)}_{(1)} = (2/3)\ln(960/964)+(1/3)\ln(960/956)+\ln x^{(0)}_{(1)} = -.0027720067+.0013917903+4.430816799$$
$$= 4.429436582\ ,\ x^{(1)}_{(1)} = 83.8841$$

$$\ln x^{(1)}_{(2)} = (1/3)\ln(960/964)+(2/3)\ln(960/956)+\ln x^{(0)}_{(2)} = -.001386003367+.002783580667+5.049856007$$
$$= 5.051253585,\ x^{(1)}_{(2)} = 156.2182$$

$$\ln x^{(1)}_{(3)} = (1/3)\ln(960/956)+(2/3)\ln(960/960)+\ln x^{(0)}_{(3)} = .001391790333+4.890349128=4.891740918,$$
$$x^{(1)}_{(3)} = 133.1852$$

$$\ell n\, x^{1}_{(4)} = (1/3)\,\ell n(960/964) + (2/3)\,\ell n(960/960) + \ell n\, x^{0}_{(4)} = -.001386003367 + 4.672828835 = 4.671442832,\quad x^{1}_{(4)} = 106.8518$$

etc.

$$h^{(1)}_{1} = 83.8841 + 156.2182 = 240.1023,\quad h^{(1)}_{2} = 133.1852 + 106.8518 = 240.037,\ \text{etc.}$$

$$\ell n\, x^{(2)}_{(1)} = \ell n(240/240.1023) + \ell n\, x^{(1)}_{(1)} = -.0004261591 + 4.429436582 = 4.429010423,\quad x^{(2)}_{(1)} = 83.8484$$

$$\ell n\, x^{(2)}_{(2)} = \ell n(240/240.1023) + \ell n\, x^{(1)}_{(2)} = -.0004261591 + 5.051253585 = 5.050827426,\quad x^{(2)}_{(2)} = 156.1516$$

$$\ell n\, x^{(2)}_{(3)} = \ell n(240/240.037) + \ell n\, x^{(1)}_{(3)} = -.0001541547 + 4.891740918 = 4.891586763,\quad x^{(2)}_{(3)} = 133.1647$$

$$\ell n\, x^{(2)}_{(4)} = \ell n(240/240.037) + \ell n\, x^{(1)}_{(4)} = -.0001541547 + 4.671442832 = 4.671288677,\quad x^{(2)}_{(4)} = 106.8354$$

BARTLETT'S ROOT CUTTING DATA - NO INTERACTION LINEAR SCALE

DARROCH - RATCLIFF ITERATION - INITIAL VALUE FIRST ITERATE KULLITR.

ω	$x^{(0)}$	α_1	α_2	α_3	$K_1^{(0)}$	$K_2^{(0)}$	$K_3^{(0)}$	a_1	a_2	a_3	a_4
1	82.8862	2	1	0	165.7724	82.8862	0	1	0	0	0
2	157.1137	1	2	0	157.1137	314.2274	0	1	0	0	0
3	134.2119	0	1	2	0	134.2119	268.4238	0	1	0	0
4	105.7881	1	0	2	105.7881	0	211.5762	0	1	0	0
5	157.1137	0	1	2	0	157.1137	314.2274	0	0	1	0
6	82.8862	1	0	2	82.8862	0	165.7724	0	0	1	0
7	208.4432	2	1	0	416.8864	208.4432	0	0	0	0	1
8	31.5565	1	2	0	31.5565	63.1130	0	0	0	0	1
	all div.by 3				960.0033	959.9954	959.9998				

$$\ell n\ x^{(1)}_{(1)} = (2/3)\ell n(960/960.0033)+(1/3)\ell n(960/959.9954)+\ell n\ x^{(0)}_{(1)} = -.00000229167+.000001597+4.417468583$$
$$= 4.417467888,\quad x^{(1)}_{(1)} = 82.8861$$

$$\ell n\ x^{(1)}_{(2)} = (1/3)\ell n(960/960.0033)+(2/3)\ell n(960/959.9954)+\ell n\ x^{(0)}_{(2)} = -.0000011458+.0000031947+5.0569747$$
$$= 5.056971796,\quad x^{(1)}_{(2)} = 157.1140$$

$$\ell n\ x^{(1)}_{(3)} = (1/3)\ell n(960/959.9954)+(2/3)\ell n(960/959.9998)+\ell n\ x^{(0)}_{(3)} = .000001597+.0000013867+4.899419894$$
$$= 4.89942163,\quad x^{(1)}_{(3)} = 134.2121$$

$$\ell n\ x^{(1)}_{(4)} = (1/3)\ell n(960/960.0033)+(2/3)\ell n(960/959.9998)+\ell n\ x^{(0)}_{(4)} = -.0000011458+.0000013867+4.661438037$$
$$= 4.66143703,\quad x^{(1)}_{(4)} = 105.7880$$

etc.

$$h^{(1)}_{(1)} = 82.8861+157.1140 = 240.0001,\quad h^{(1)}_{(2)} = 134.2121+105.7880 = 240.0001 \text{ etc.}$$

$$\ell n\ x^{(2)}_{(1)} = \ell n(240/240.0001)+\ell n\ x^{(1)}_{(1)} = -.0000004167+4.417467888 = 4.417467471,\quad x^{(2)}_{(1)} = 82.8861$$

$$\ell n\ x^{(2)}_{(2)} = \ell n(240/240.0001)+\ell n\ x^{(1)}_{(2)} = -.0000004167+5.056971796 = 5.056971379,\quad x^{(2)}_{(2)} = 157.1140$$

$$\ell n\ x^{(2)}_{(3)} = \ell n(240/240.0001)+\ell n\ x^{(1)}_{(3)} = -.0000004167+4.89942163 = 4.899421213,\quad x^{(2)}_{(3)} = 134.2121$$

$$\ell n\ x^{(2)}_{(4)} = \ell n(240/240.0001)+\ell n\ x^{(1)}_{(4)} = \quad \text{etc.} \quad ,\ x^{(2)}_{(4)} = 105.7880$$

The DARRAT computer program using the initial distribution as the uniform after 31 iterations yielded the minimum discrimination information estimates

ω	$x^*(\omega)$	
111	82.886	
112	157.114	
121	134.212	
122	105.788	$2I(x^*:x) = 0.082$
211	157.114	
212	82.886	
221	208.443	
222	31.557	

The computer output using the GOKHALE program on Bartlett's root cutting data follows.

BARTLETT'S ROOT CUTTINGS

3 FACTOR TABLE:TIME*CUTTING*MORTALITY

GOKHALE PROGRAM

B_MATRIX

	1	2	3	4	5	6	7	8
1	1	1	0	0	0	0	0	0
2	0	0	1	1	0	0	0	0
3	0	0	0	0	1	1	0	0
4	0	0	0	0	0	0	1	1
5	1	0	-1	0	-1	0	1	0

WEIGHT(1)= 0.250000
WEIGHT(2)= 0.250000
WEIGHT(3)= 0.250000
WEIGHT(4)= 0.250000

INV_WEIGHT(1)= 4.000000
INV_WEIGHT(2)= 4.000000
INV_WEIGHT(3)= 4.000000
INV_WEIGHT(4)= 4.000000

C_DESIGN MATRIX

	1	2	3	4	5	6	7	8
1	4	4	0	0	0	0	0	0
2	0	0	4	4	0	0	0	0
3	0	0	0	0	4	4	0	0
4	0	0	0	0	0	0	4	4
5	4	0	-4	0	-4	0	4	0

OBSERVED VALUES

1	1	1	X(1)= 84.000000	LN_X(1)= 4.430817
1	1	2	X(2)= 156.000000	LN_X(2)= 5.049856
1	2	1	X(3)= 133.000000	LN_X(3)= 4.890349
1	2	2	X(4)= 107.000000	LN_X(4)= 4.672829
2	1	1	X(5)= 156.000000	LN_X(5)= 5.049856
2	1	2	X(6)= 84.000000	LN_X(6)= 4.430817
2	2	1	X(7)= 209.000000	LN_X(7)= 5.342334
2	2	2	X(8)= 31.000000	LN_X(8)= 3.433987

CONSTRAINTS
NTHETA(1)= 960.000000
NTHETA(2)= 960.000000

The DARRAT computer program using the initial distribution as the uniform after 31 iterations yielded the minimum discrimination information estimates

ω	$x^*(\omega)$	
111	82.886	
112	157.114	
121	134.212	
122	105.788	$2I(x^*:x) = 0.082$
211	157.114	
212	82.886	
221	208.443	
222	31.557	

BARTLETT'S ROOT CUTTINGS

3 FACTOR TABLE:TIME*CUTTING*MORTALITY

GOKHALE PROGRAM

B_MATRIX

	1	2	3	4	5	6	7	8
1	1	1	0	0	0	0	0	0
2	0	0	1	1	0	0	0	0
3	0	0	0	0	1	1	0	0
4	0	0	0	0	0	0	1	1
5	1	0	-1	0	-1	0	1	0

WEIGHT(1)= 0.250000
WEIGHT(2)= 0.250000
WEIGHT(3)= 0.250000
WEIGHT(4)= 0.250000

INV_WEIGHT(1)= 4.000000
INV_WEIGHT(2)= 4.000000
INV_WEIGHT(3)= 4.000000
INV_WEIGHT(4)= 4.000000

C_DESIGN MATRIX

	1	2	3	4	5	6	7	8
1	4	4	0	0	0	0	0	0
2	0	0	4	4	0	0	0	0
3	0	0	0	0	4	4	0	0
4	0	0	0	0	0	0	4	4
5	4	0	-4	0	-4	0	4	0

OBSERVED VALUES

1	1	1	X(1)= 84.000000	LN_X(1)= 4.430817
1	1	2	X(2)= 156.000000	LN_X(2)= 5.049856
1	2	1	X(3)= 133.000000	LN_X(3)= 4.890349
1	2	2	X(4)= 107.000000	LN_X(4)= 4.672829
2	1	1	X(5)= 156.000000	LN_X(5)= 5.049856
2	1	2	X(6)= 84.000000	LN_X(6)= 4.430817
2	2	1	X(7)= 209.000000	LN_X(7)= 5.342334
2	2	2	X(8)= 31.000000	LN_X(8)= 3.433987

CONSTRAINTS
NTHETA(1)= 960.000000
NTHETA(2)= 960.000000

```
NTHETA(3)= 960.000000
NTHETA(4)= 960.000000
NTHETA(5)=    0.000000



MIN. MOD. CHI SQ. EST.
XHAT(1)=  82.882813
XHAT(2)= 157.117172
XHAT(3)= 134.213257
XHAT(4)= 105.786728
XHAT(5)= 157.117172
XHAT(6)=  82.882813
XHAT(7)= 208.447632
XHAT(8)=  31.552353

ITERATIONS=  2
ESTIMATED DISTRIBUTION
  1   1   1        XSTAR(1)=  82.885101       LN_XSTAR(1)=   4.417456
  1   1   2        XSTAR(2)= 157.114883       LN_XSTAR(2)=   5.056977
  1   2   1        XSTAR(3)= 134.213211       LN_XSTAR(3)=   4.899429
  1   2   2        XSTAR(4)= 105.786774       LN_XSTAR(4)=   4.661426
  2   1   1        XSTAR(5)= 157.114883       LN_XSTAR(5)=   5.056977
  2   1   2        XSTAR(6)=  82.885101       LN_XSTAR(6)=   4.417456
  2   2   1        XSTAR(7)= 208.442993       LN_XSTAR(7)=   5.339665
  2   2   2        XSTAR(8)=  31.556992       LN_XSTAR(8)=   3.451795


Z1(XSTAR:X)=          0.081901

  1   1   1        OUTLIER(1)= 0.008141
  1   1   2        OUTLIER(2)= 0.004748
  1   2   1        OUTLIER(3)= 0.006407
  1   2   2        OUTLIER(4)= 0.007766
  2   1   1        OUTLIER(5)= 0.004748
  2   1   2        OUTLIER(6)= 0.008141
  2   2   1        OUTLIER(7)= 0.000949
  2   2   2        OUTLIER(8)= 0.005141
ESTIMATED CONSTRAINTS
NTHAT(1)= 960.000000
NTHAT(2)= 960.000000
NTHAT(3)= 959.999756
NTHAT(4)= 959.999756
NTHAT(5)=  -0.000000



S22.1

                    1

  1        3121.385742


S22.1_INV
```

1

1 0.000320

XSO= 0.082015

TAU(1)= 0.001780
TAU(2)=-0.002851
TAU(3)=-0.003340
TAU(4)= 0.004453
TAU(5)=-0.005120

8. Further Applications

In this chapter we consider six examples illustrating the application of the k-sample and the general linear hypothesis techniques.

Example 1. Gail's data. This example illustrates the procedure for getting m.d.i. estimates under hypotheses about the underlying probabilities of two contingency tables and testing the null hypothesis. An analysis of information table is also given in this case, including a subhypothesis. Note the difference in the analysis of information from those for the fitting problems.

Example

Gail's Data

As an illustration of the k-sample approach consider the following two contingency tables (artificial data) considered by Gail (1974, p. 97).

a)

20	5	5	30
6	4	2	12
26	9	7	42

b)

15	15	2	32
10	5	5	20
25	20	7	52

Table 1

The problem of interest was whether the underlying probabilities for the two tables were such that the respective marginal probabilities of the two tables were the same. If so, could it be a consequence of the fact that the tables were homogeneous?

Let us denote the observed values in the two tables as in Table 2

a)

$x(111)$	$x(112)$	$x(113)$	$x(11.)$
$x(121)$	$x(122)$	$x(123)$	$x(12.)$
$x(1.1)$	$x(1.2)$	$x(1.3)$	N_1

b)

$x(211)$	$x(212)$	$x(213)$	$x(21.)$
$x(221)$	$x(222)$	$x(223)$	$x(22.)$
$x(2.1)$	$x(2.2)$	$x(2.3)$	N_2

Table 2

For the hypothesis H_1 that the respective marginal probabilities are the same the basic values for the k-sample approach follow.

	ω	$x(\omega)$	$\ln x(\omega)$
111	1	20	2.995732
112	2	5	1.609438
113	3	5	1.609438
121	4	6	1.791759
122	5	4	1.386294
123	6	2	0.693147
211	7	15	2.708050
212	8	15	2.708050
213	9	2	0.693147
221	10	10	2.302585
222	11	5	1.609438
223	12	5	1.609438

$N_1 = 42$
$N_2 = 52$
$N = 94$
$w_1 = 42/94 = 0.446808$
$w_2 = 52/94 = 0.553191$
$v_1 = 1/w_1 = 2.238094$
$v_2 = 1/w_2 = 1.807692$

The $\underline{B}$ matrix for H_1 and the values of $\underline{\theta}$ and $N\underline{\theta}$ are given in Table 3.

$$\underline{W}_1 = \begin{pmatrix} w_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & w_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & w_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & w_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & w_1 \end{pmatrix} \qquad \underline{W}_2 = \begin{pmatrix} w_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & w_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & w_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & w_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & w_2 \end{pmatrix}$$

$$\underline{W} = \begin{pmatrix} \underline{W}_1 & \underline{O} \\ \underline{O} & \underline{W}_2 \end{pmatrix}$$

$\underline{C} = \underline{BW}^{-1}$, $\underline{C} = \begin{pmatrix} \underline{C}_1 \\ \underline{C}_2 \end{pmatrix}$, $\underline{C}_1$ is 2 x 12, $\underline{C}_2$ is 3 x 12

The $\underline{C}$ matrix is obtained by multiplying all the elements in the first 6 columns of the $\underline{B}$ matrix by 2.238094 and by multiplying all the elements in the last 6 columns of the $\underline{B}$ matrix by 1.807692.

$$\underline{C}\,\underline{x} = N\,\underline{\phi} = \begin{pmatrix} 94 \\ 94 \\ 9.296700 \\ 12.998163 \\ -16.010971 \end{pmatrix},$$

$$\underline{S} = \underline{C}\,\underline{D}_{\underline{x}}\underline{C}' = \begin{pmatrix} \underline{S}_{11} & \underline{S}_{12} \\ \underline{S}_{21} & \underline{S}_{22} \end{pmatrix},$$

$$\underline{S}_{22.1}^{-1} = \left(\underline{S}_{22} - \underline{S}_{21}\underline{S}_{11}^{-1}\underline{S}_{12}\right)^{-1} = \begin{pmatrix} 0.012198 & -0.001927 & -0.001776 \\ -0.001927 & 0.022284 & 0.017520 \\ -0.001776 & 0.017520 & 0.027\text{[illegible]}01 \end{pmatrix},$$

$$\underline{d} = N\,\underline{\theta}^{*} - N\,\hat{\underline{\theta}} = \begin{pmatrix} -9.296700 \\ -12.998163 \\ 16.010971 \end{pmatrix}.$$

The minimum modified χ^2 value, the quadratic approximation to $2I(x^*:x)$ is $\chi^2 = \underline{d}'\underline{S}_{22.1}^{-1}\underline{d} = 4.512$, 3 D.F.

After 3 iterations the values of the minimum discrimination information estimates are as follows.

	ω	$x^*(\omega)$	$\ell n\ x^*(\omega)$
111	1	16.504	2.803630
112	2	6.466	1.866627
113	3	4.042	1.396620
121	4	6.320	1.843785
122	5	6.604	1.887611
123	6	2.064	0.724458
211	7	18.263	2.904873
212	8	12.705	2.541984
213	9	2.476	0.906704
221	10	9.996	2.302228
222	11	3.477	1.246192
223	12	5.083	1.625813

It is found that $2I(x^*:x) = 4.333$, 3D.F. We now proceed to test the hypothesis H_2 that the two contingency tables are homogeneous. The $\underline{B}$ matrix, $\underline{\theta}$, and $N\underline{\nu}$ for H_2 are given in Table 4.

Using the $\underline{B}$ matrix of Table 4 we have

$$\underline{C} = \underline{B}\underline{W}^{-1}, \quad \underline{C} = \begin{pmatrix} \underline{C}_1 \\ \underline{C}_2 \end{pmatrix}, \quad \underline{C}_1 \text{ is } 2 \times 12, \; \underline{C}_2 \text{ is } 5 \times 12,$$

$$\underline{C}\,\underline{x} = N\underline{\nu} = \begin{pmatrix} 94 \\ 94 \\ 17.646500 \\ -15.924902 \\ 7.575089 \\ -4.648350 \\ -0.086081 \end{pmatrix}.$$

$\underline{S}_{22.1}^{-1}$ is now a 5×5 matrix, we omit the detailed values and $X^2 = \underline{d}'\underline{S}_{22.1}^{-1}\underline{d} = 9.300$, 5 D.F.

After 3 iterations the values of the minimum discrimination information estimates are:

	ω	$x^*(\omega)$	$\ln x^*(\omega)$
111	1	15.901	2.766356
112	2	8.559	2.146948
113	3	2.808	1.032321
121	4	7.419	2.004110
122	5	4.219	1.439503
123	6	3.095	1.129803
211	7	19.686	2.979930
212	8	10.596	2.360522
213	9	3.476	1.245895
221	10	9.186	2.217686
222	11	5.723	1.653077
223	12	3.832	1.343370

It is found that under H_2 $2I(x^*:x) = 9.008$ 5 D.F. If we denote the m.d.i. estimate under the marginal homogeneity hypothesis H_1 by x^*_M and under the homogeneity hypothesis H_2 by x^*_H, then we may summarize the results in the Analysis of Information Table 5.

Analysis of Information

Component due to	Information	D.F.
H_2	$2I(x^*_H : x) = 9.008$	5
H_1	$2I(x^*_H : x^*_M) = 4.675$	2
	$2I(x^*_M : x) = 4.333$	3

Table 5

We see that the tables are homogeneous; hence the marginals are also homogeneous.

Note that

$$2\ \Sigma x^*_H\ \ell n\ \frac{x^*_H}{x} = 2\ \Sigma x^*_H\ \ell n\ \frac{x^*_H}{x^*_M} + 2\ \Sigma x^*_H\ \ell n\ \frac{x^*_M}{x}\ .$$

But x^*_H also satisfies the restraints for x^*_M (homogeneity implies marginal homogeneity) hence

$$2\ \Sigma x^*_H\ \ell n\ \frac{x^*_M}{x} = 2\ \Sigma x^*_M\ \ell n\ \frac{x^*_M}{x}$$

and we have the analysis as in Table 5.

The statistics given by Gail (1974) are the same as the X^2 values given above.

	111	112	113	121	122	123	211	212	213	221	222	223		
ω	1	2	3	4	5	6	7	8	9	10	11	12	θ	Nθ
	1	1	1	1	1	1	0	0	0	0	0	0	1	94
	0	0	0	0	0	0	1	1	1	1	1	1	1	94
	1	1	1	0	0	0	-1	-1	-1	0	0	0	0	0
	1	0	0	1	0	0	-1	0	0	-1	0	0	0	0
	0	1	0	0	1	0	0	-1	0	0	-1	0	0	0

Table 3

ω	1	2	3	4	5	6	7	8	9	10	11	12	θ	Nθ
	1	1	1	1	1	1	0	0	0	0	0	0	1	94
	0	0	0	0	0	0	1	1	1	1	1	1	1	94
	1	0	0	0	0	0	-1	0	0	0	0	0	0	0
	0	1	0	0	0	0	0	-1	0	0	0	0	0	0
	0	0	1	0	0	0	0	0	-1	0	0	0	0	0
	0	0	0	1	0	0	0	0	0	-1	0	0	0	0
	0	0	0	0	1	0	0	0	0	0	-1	0	0	0

Table 4

Example 2. Gokhale discrete distributions. This example illustrates the application of the k-sample procedure to test hypotheses about the means and variances of two discrete distributions, not in the form of contingency tables. An analysis of information table is given.

$$\underline{W} = \begin{pmatrix} \underline{W}_1 & \underline{0} \\ \underline{0} & \underline{W}_2 \end{pmatrix},$$

$$\underline{C} = \underline{B}\underline{W}^{-1} = \begin{pmatrix} 3 & 3 & 3 & 3 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.5 & 1.5 \\ -6 & -3 & 0 & 3 & 6 & 2.25 & -2.25 \end{pmatrix},$$

$$\underline{x} = \begin{matrix} x(1) = 6 \\ x(2) = 18 \\ x(3) = 9 \\ x(4) = 24 \\ x(5) = 3 \\ x(6) = 72 \\ x(7) = 48 \end{matrix}, \qquad \begin{matrix} \ell n\ x(1) = 1.791759 \\ \ell n\ x(2) = 2.890371 \\ \ell n\ x(3) = 2.197225 \\ \ell n\ x(4) = 3.178054 \\ \ell n\ x(5) = 1.098612 \\ \ell n\ x(6) = 4.276666 \\ \ell n\ x(7) = 3.871201 \end{matrix},$$

$$N\underline{\theta} = \begin{pmatrix} 180 \\ 180 \\ 0 \end{pmatrix}, \quad N\underline{\phi} = \underline{C}\underline{x} = \begin{pmatrix} 180 \\ 180 \\ 54 \end{pmatrix},$$

$$\underline{S} = \underline{C}\,\underline{D}_x\,\underline{C}' = \begin{pmatrix} 540 & 0 & 0 \\ 0 & 270 & 81 \\ 0 & 81 & 1309.5 \end{pmatrix},$$

$$\underline{S}_{22.1} = 1309.5 - (0 \quad 81) \begin{pmatrix} 1/540 & 0 \\ 0 & 1/270 \end{pmatrix} \begin{pmatrix} 0 \\ 81 \end{pmatrix}$$

$$= 1285.199951,$$

$$S_{22.1}^{-1} = 0.000778, \quad \underline{\Delta} = \begin{pmatrix} 0 \\ 0 \\ -54 \end{pmatrix}, \quad d = -54,$$

$$X^2 = (-54)^2(.000778) = 2.269, \text{ 1 D.F.}$$

After two iterations there is obtained

$$\begin{matrix} X^*(1) = 7.618 \\ X^*(2) = 20.180 \\ X^*(3) = 8.909 \\ X^*(4) = 20.978 \\ X^*(5) = 2.315 \end{matrix} \qquad \begin{matrix} \ell n\ X^*(1) = 2.030505 \\ \ell n\ X^*(2) = 3.004673 \\ \ell n\ X^*(3) = 2.187082 \\ \ell n\ X^*(4) = 3.043467 \\ \ell n\ X^*(5) = 0.839582 \end{matrix} \qquad \begin{matrix} -2 \times 7.618 = -15.236 \\ -1 \times 20.180 = -20.180 \\ 0 \times 8.909 = 0 \\ 1 \times 20.978 = 20.978 \\ 2 \times 2.315 = 4.630 \end{matrix}$$

$X^*(6) = 66.538$ $\quad \ell n\, X^*(6) - 4.197771$ $\quad -1.5 \times 66.538 = -99.807$

$X^*(7) = 53.462$ $\quad \ell n\, X^*(7) = 3.978971$ $\quad -1.5 \times 53.462 = \ 80.193$

$$2I(X^*:X) = 2.248,\ 1\text{ D.F.}$$

$$(-15.236 - 20.180 + 0 + 20.978 + 4.630)/60 = -0.1635$$

$$(-99.807 + 80.193)/120 = -0.1635.$$

Under H_2 the restraints are $\underline{Bp} = \underline{\theta}$ with

$$\underline{B} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 & 1.5 & -1.5 \\ 4 & 1 & 0 & 1 & 4 & -2.25 & -2.25 \end{pmatrix},\quad \underline{\theta} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

Note that the last row of the matrix derives from

$$(-2)^2P_1(-2) + (-1)^2P_1(-1) + 0^2P_1(0) + 1^2P_1(1) + 2^2P_1(2) - \Big((-1.5)^2P_2(-1.5) + (1.5)^2P_2(1.5)\Big).$$

$$\underline{C}=\underline{BW}^{-1} = \begin{pmatrix} 3 & 3 & 3 & 3 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.5 & 1.5 \\ -6 & -3 & 0 & 3 & 6 & 2.25 & -2.25 \\ 12 & 3 & 0 & 3 & 12 & -3.375 & -3.375 \end{pmatrix},$$

$$N\underline{\theta} = \begin{pmatrix} 180 \\ 180 \\ 0 \\ 0 \end{pmatrix},\quad \underline{CX} = N\underline{\phi} = \begin{pmatrix} 180 \\ 180 \\ 54 \\ -171 \end{pmatrix},$$

$$\underline{S} = \underline{CD_xC'} = \begin{pmatrix} 540 & 0 & 0 & 702 \\ 0 & 270 & 81 & -607.5 \\ 0 & 81 & 1309.5 & -344.25 \\ 702 & -607.5 & -344.25 & 3040.875 \end{pmatrix},$$

$$\underline{S}_{22.1} = \begin{pmatrix} 1309.5 & -344.25 \\ -344.25 & 3040.875 \end{pmatrix} - \begin{pmatrix} 0 & 81 \\ 702 & -607.5 \end{pmatrix} \begin{pmatrix} 540 & 0 \\ 0 & 270 \end{pmatrix}^{-1} \begin{pmatrix} 0 & 702 \\ 81 & -607.5 \end{pmatrix}$$

$$= \begin{pmatrix} 1285.199951 & -162.0 \\ -162.0 & 761.399902 \end{pmatrix},$$

$$\underline{S}^{-1}_{22.1} = \begin{pmatrix} .000800 & .000170 \\ .000170 & .001350 \end{pmatrix},\ \underline{\Delta} = \begin{pmatrix} 0 \\ 0 \\ -54 \\ 171 \end{pmatrix},\ \underline{d} = \begin{pmatrix} -54 \\ 171 \end{pmatrix}.$$

$$X^2 = (-54,171)\begin{pmatrix} .000800 & .000170 \\ .000170 & .001350 \end{pmatrix}\begin{pmatrix} -54 \\ 171 \end{pmatrix} = 38.652,\ 2\ \text{D.F.}$$

After four iterations there is obtained

$x^*(1) = 18.134$	$\ell n x^*(1)=2.897783$	$-2(18.134)=-36.268, 4(18.134)=72.536$
$x^*(2) = 13.081$	$\ell n x^*(2)=2.571174$	$-1(13.081)=-13.081, 1(13.081)=13.081$
$x^*(3) = 4.000$	$\ell n x^*(3)=1.386189$	$0(4) = 0, 0(4) = 0$
$x^*(4) = 16.586$	$\ell n x^*(4)=2.808560$	$1(16.586)=16.586, 1(16.586)=16.586$
$x^*(5) = 8.199$	$\ell n x^*(5)=2.104045$	$2(8.199)=16.398, 4(8.199)=32.796$
$x^*(6) = 70.910$	$\ell n x^*(6)=4.261405$	$-1.5(70.910)=-106.365, (-1.5)^2(70.910) =159.548$
$x^*(7) = 49.090$	$\ell n x^*(7)=3.893661$	$1.5(49.090)=73.635, (1.5)^2(49.090) =110.453$

$2I(x^*:x)=29.546$, 2 D.F.

$(-36.268-13.081+16.586+16.398)/60=-0.2728, (-106.365+73.635)/120=-0.2728$
$(72.536+13.081+16.586+32.796)/60=2.2500, (159.548+110.453)/120=2.2500$

We may summarize in the analysis of information table.

Analysis of Information

Component due to	Information	D.F.
H_2	$2I(x^*_2:x)=29.546$	2
H_2-H_1 (Effect)	$2I(x^*_2:x^*_1)=27.298$	1
H_1	$2I(x^*_1:x)=2.248$	1

We reject the hypothesis H_2 but accept the hypothesis H_1. The effect of the differences in the variances is significant.

We also used the Darroch-Ratcliff iterative scaling procedure for this example.

Example 3. Marginal homogeneity of an rxr contingency table. This example illustrates the application of the k-sample procedure to a set of data previously estimated using a different algorithm. It also serves as an introduction to the next example. It points out a case in which the Π-distribution is not the uniform distribution and shows the estimate to retain properties of the original observations not involved in the null hypothesis. For applications of the notion of marginal homogeneity to higher order contingency tables see Kullback, 1971a, 1971b. The latter paper includes an example of the quadratic approximation to $2I(x^*:x)$.

Example

Marginal Homogeneity of an r x r Contingency Table

In the paper "Symmetry and marginal homogeneity of an r x r contingency table," by C.T. Ireland, H.H. Ku, S. Kullback Journal of the American Statistical Association, Vol. 64 (1969), 1323-1341 the principle of minimum discrimination information estimation was applied to obtain RBAN estimates of the cell frequencies of an r x r contingency table under hypotheses of either symmetry or marginal homogeneity.

The procedures were illustrated with data from case-records of the eye-testing of employees in Royal Ordnance factories analysed by A. Stuart.

Table

7477 Women Aged 30-39; Unaided Distance Vision x(ij)

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

We shall supplement the discussion in Ireland et al. (1969) by using the single-sample algorithm to derive the m.d.i. estimates as well as the minimum modified χ^2 estimates and relate the results to values given by A. Stuart, "A test for homogeneity of the marginal distributions in a two-way classification," Biometrika, Vol. 42 (1955), 412-416 and V.P. Bhapkar, "A note on the equivalence of two criteria for hypotheses in categorical data," Journal

of the American Statistical Association, Vol. 61 (1966), 228-235.

The reader is referred to Ireland et al. (1969) for further discussion and references. The basic table will also be used to illustrate the k-sample algorithm applied to incomplete data. We remind the reader that the graphic form of the log-linear representation using $C_1(\omega) = L$, $C_2(\omega) = T_1(\omega)$, $C_3(\omega) = T_2(\omega)$, $C_4(\omega) = T_3(\omega)$ presents

$$\ln \frac{x^*(\omega)}{x(\omega)} = L + \tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega)$$

where from the output $L=0.000805$, $\tau_1=-0.159043$, $\tau_2=-0.105379$, $\tau_3= -0.050000$. The $\underline{T}$ design matrix is of course the same as $\underline{C}'$.

Bhapkar's test statistic is the minimum modified χ^2 and he gave $\chi^2_B = 11.976$ with 3 D.F. He did not give the minimum modified χ^2 estimates. The program yields $\chi^2 = 11.975717$. Stuart gave no estimates either and he used as his statistic $\chi^2_S = \underline{d}'\underline{S}_{22}^{-1}\underline{d} = 11.957$. Stuart estimated the covariance matrix of the $\underline{d}$'s under the null hypothesis. From the computer output we see that $\underline{S}_{22}$ and $\underline{S}_{22.1}$ are not very much different in this case.

From the log-linear representation of the m.d.i. estimate we see that associations in the original table are the same as in the estimated table, thus

$$\ln \frac{x^*(ii)x^*(jj)}{x^*(ij)x^*(ji)} = \ln \frac{x(ii)\ x(jj)}{x(ij)\ x(ji)},$$

$$\ln \frac{x^*(ij)x^*(44)}{x^*(i4)x^*(4j)} = \ln \frac{x(ij)\ x(44)}{x(i4)\ x(4j)}.$$

Based on the values $\chi_S^2 = 11.957$, $\chi_B^2 = 11.976$ with 3 D.F. Stuart, and also Bhapkar, rejected the null hypothesis of marginal homogeneity. We find that $2I(x^*:x) = 12.017$, 3 D.F. and reject the null Hypothesis of homogeneity.

We remark that the discussion in Ireland et al. (1969) used a different iterative algorithm.

Log-linear representation

ij	ω	L	τ_1	τ_2	τ_3
11	1	1	0	0	0
12	2	1	1	-1	0
13	3	1	1	0	-1
14	4	1	1	0	0
21	5	1	-1	1	0
22	6	1	0	0	0
23	7	1	0	1	-1
24	8	1	0	1	0
31	9	1	-1	0	1
32	10	1	0	-1	1
33	11	1	0	0	0
34	12	1	0	0	1
41	13	1	-1	0	0
42	14	1	0	-1	0
43	15	1	0	0	-1
44	16	1	0	0	0

STUARTS DATA

DESIGN MATRIX

ω	11	12	13	14	21	22	23	24	31	32	33	34	41	42	43	44
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	0	1	1	1	-1	0	0	0	-1	0	0	0	-1	0	0	0
3	0	-1	0	0	1	0	1	1	0	-1	0	0	0	-1	0	0
4	0	0	-1	0	0	0	-1	0	1	1	0	1	0	0	-1	0

```
X( 1)= 1520.000000
X( 2)=  266.000000
X( 3)=  124.000000
X( 4)=   66.000000
X( 5)=  234.000000
X( 6)= 1512.000000
X( 7)=  432.000000
X( 8)=   78.000000
X( 9)=  117.000000
X(10)=  362.000000
X(11)= 1772.000000
X(12)=  205.000000
X(13)=   36.000000
X(14)=   82.000000
X(15)=  179.000000
X(16)=  492.000000
```

Observed data

```
LN_X( 1)=     7.326466
LN_X( 2)=     5.583496
LN_X( 3)=     4.820282
LN_X( 4)=     4.189654
LN_X( 5)=     5.455321
LN_X( 6)=     7.321189
LN_X( 7)=     6.068426
LN_X( 8)=     4.356709
LN_X( 9)=     4.762174
LN_X(10)=     5.891644
LN_X(11)=     7.479864
LN_X(12)=     5.323010
LN_X(13)=     3.583519
LN_X(14)=     4.406719
LN_X(15)=     5.187386
LN_X(16)=     6.198479
```

```
NTHETA( 1)= 7477.000000
NTHETA( 2)=    0.000000
NTHETA( 3)=    0.000000
NTHETA( 4)=    0.000000
```

THAT(1)= 7477.000000
THAT(2)= 69.000000
THAT(3)= 34.000000
THAT(4)= -51.000000

$_xS = CD_xC'$

	1	2	3	4
1	7477.000000	69.000000	34.000000	-51.000000
2	69.000000	8[illegible].000000	-500.000000	-241.000000
3	34.000000	-500.000000	1454.000000	-794.000000
4	-51.000000	-241.000000	-794.000000	1419.000000

$S_{\not{}22.1}$

	1	2	3
1	842.363357	-500.313721	-240.529343
2	-500.313721	1453.6[illegible]5215	-793.768066
3	-240.529343	-793.768066	1418.652100

$S_{\not{}}$_INV

	1	2	3	4
1	0.000134	-0.000021	-0.000014	-0.000007
2	-0.000021	0.002485	0.001563	0.001296
3	-0.000014	0.001563	0.001973	0.001369
4	-0.000007	0.001296	0.001369	0.001691

$S_{\not{}22.1}$_INV

	1	2	3
1	0.002485	0.001563	0.001296
2	0.001563	0.001973	0.001369
3	0.001296	0.001369	0.001691

DELTA(1)= 0.000000
DELTA(2)= -69.000000
DELTA(3)= -34.000000
DELTA(4)= 51.000000

$$= \begin{pmatrix} 0 \\ d_1 \\ d_2 \\ d_3 \end{pmatrix}$$

```
LAMBDA( 1)=   0.001602
LAMBDA( 2)=  -0.158536
LAMBDA( 3)=  -0.10[illegible]0[illegible]
LAMBDA( 4)=  -0.049753

MU( 1)=   0.001[illegible]02
MU( 2)=  -0.051[illegible]37
MU( 3)=  -0.1071[illegible]
MU( 4)=  -0.156932
MU( 5)=   0.055[illegible]
MU( 6)=   0.001[illegible]02
MU( 7)=  -0.053761
MU( 8)=  -0.1034[illegible]
MU( 9)=   0.1104[illegible]
MU(10)=   0.056[illegible]
MU(11)=   0.001[illegible]02
MU(12)=  -0.048[illegible]51
MU(13)=   0.160136
MU(14)=   0.106[illegible]7
MU(15)=   0.051[illegible]
MU(16)=   0.001602

XHAT( 1)= 15[illegible]2.434326
XHAT( 2)=  252.211517
XHAT( 3)=  110.707291
XHAT( 4)=   55.642471
XHAT( 5)=  247.579288
XHAT( 6)= 1514.421631
XHAT( 7)=  405.775145
XHAT( 8)=   70.927444
XHAT( 9)=  129.917099
XHAT(10)=  362.621094
XHAT(11)= 1774.830135
XHAT(12)=  195.133072
XHAT(13)=   41.764877
XHAT(14)=   70.749151
XHAT(15)=  183.183904
XHAT(16)=  492.787842
```

Min. Mod. χ^2 estimate

```
XSQ=     11.975717
```

Min. Mod. $\chi^2 = \underline{d}'\underline{S}_{22.1}^{-1}\underline{d}$

```
ZI(XHAT:X)=    12.02085[illegible]
```

```
XSTAR( 1)= 1521.222412
XSTAR( 2)=  252.361313
XSTAR( 3)=  111.306107
XSTAR( 4)=   56.369278
XSTAR( 5)=  247.043015
XSTAR( 6)= 1513.216553
XSTAR( 7)=  409.061768
XSTAR( 8)=   70.276963
XSTAR( 9)=  130.552902
XSTAR(10)=  362.914307
XSTAR(11)= 1773.425049
XSTAR(12)=  195.211044
XSTAR(13)=   42.218399
```

M.D.I. Estimate
First iterate

```
XSTAR(14)=     91.160263
XSTAR(15)=    168.276397
XSTAR(16)=    492.395752


LN_XSTAR( 1)=       7.327270
LN_XSTAR( 2)=       5.530862
LN_XSTAR( 3)=       4.712204
LN_XSTAR( 4)=       4.031924
LN_XSTAR( 5)=       5.509562
LN_XSTAR( 6)=       7.321993
LN_XSTAR( 7)=       6.013866
LN_XSTAR( 8)=       4.252416
LN_XSTAR( 9)=       4.871776
LN_XSTAR(10)=       5.947811
LN_XSTAR(11)=       7.480668
LN_XSTAR(12)=       5.274081
LN_XSTAR(13)=       3.742356
LN_XSTAR(14)=       4.512619
LN_XSTAR(15)=       5.237922
LN_XSTAR(16)=       6.199283


2I(XSTAR:X)=    11.954031

T =       0.000805

TAU( 1)=    -0.158534
TAU( 2)=    -0.105036
TAU( 3)=    -0.049733
```

Final M.D.I. Estimate

```
XSTAR( 1)=  1521.222412
XSTAR( 2)=   252.304047
XSTAR( 3)=   111.279144
XSTAR( 4)=    56.340576
XSTAR( 5)=   247.099106
XSTAR( 6)=  1513.216553
XSTAR( 7)=   409.055664
XSTAR( 8)=    70.255127
XSTAR( 9)=   130.534518
XSTAR(10)=   382.920166
XSTAR(11)=  1773.425049
XSTAR(12)=   195.158920
XSTAR(13)=    42.239899
XSTAR(14)=    91.135989
XSTAR(15)=   168.328674
XSTAR(16)=   492.395752


LN_XSTAR( 1)=       7.327270
LN_XSTAR( 2)=       5.530535
LN_XSTAR( 3)=       4.712042
LN_XSTAR( 4)=       4.031415
```

```
LN_XSTAR( 5)=     [illegible].509739
LN_XSTAR( 6)=     7.321993
LN_XSTAR( 7)=     [illegible].013351
LN_XSTAR( 8)=     4.259133
LN_XSTAR( 9)=     4.372321
LN_XSTAR(10)=     5.947326
LN_XSTAR(11)=     7.480368
LN_XSTAR(12)=     5.273314
LN_XSTAR(13)=     3.746335
LN_XSTAR(14)=     4.512901
LN_XSTAR(15)=     5.258139
LN_XSTAR(16)=     6.199233

21(XSTAR:X)=   12.016708

T=        0.000805

TAU( 1)=    -0.159043
TAU( 2)=    -0.105379
TAU( 3)=    -0.050000

NTHAT( 1)= 7477.007813
NTHAT( 2)=    -0.000336
NTHAT( 3)=     0.001053
NTHAT( 4)=    -0.000793
```

S
~~P~~22.1

	1	2	3
1	839.046924	-499.402332	-241.863663
2	-499.402332	1452.819530	-791.975342
3	-241.863663	-791.975342	1417.326660

S
~~P~~22.1_INV

	1	2	3
1	0.002500	0.001570	0.001304
2	0.001570	0.001974	0.001372
3	0.001304	0.001372	0.001695

```
DELTA( 1)=   -0.007813
DELTA( 2)=    0.000336
DELTA( 3)=   -0.001053
DELTA( 4)=    0.000793
```

ITERATIONS= 2

Example 4. Several samples, incomplete data.

This example uses the complete contingency table of the preceding example and row and column marginals only of additional samples. The example illustrates the application of the procedure to samples which may include fragmentary data.

Example

Several Samples, Incomplete Data

We shall illustrate the k-sample algorithm of testing several samples with incomplete data in terms of a specific sample. In Table 1 the 7477 observations in the 4 x 4 contingency table are Stuart's data, which we have already examined under the null hypothesis of marginal homogeneity.

"The remaining 1100 observations are artificial data for 600 women for whom only left eye vision was reported and 500 women for whom only right eye vision was reported. It will be presumed that the incomplete data for women with vision classified only for one eye arose in a completely random manner which was statistically independent of the true classification of their vision with respect to both eyes. This assumption allows us to say that the marginal probabilities pertaining to left eye vision and right eye vision for women classified on both eyes are the same parameters as the probabilities pertaining to left eye vision for women only for the left eye and to right eye vision for women classified only for the right eye respectively" (Koch et al 1972. p. 665, 666).

The results for the k-sample algorithm computer output are summarized in Table 2, in which we also give the values derived by Koch et al (1972) by their approach.

We also estimated this set of data using the Darroch-Ratcliff algorithm.

In view of the small values of the test statistics with 6 D.F. we accept the null hypothesis of the homogeneity of the data with respect to the underlying population.

Using the m.d.i. estimates of the entries in the cells of the complete contingency table as "improved" values over the original observations we repeat the test for the null hypothesis of marginal homogeneity. The resulting values are summarized in Table 2. There is no change in our inference that the data show no evidence of marginal homogeneity.

Table 3 gives the graphic presentation of the log-linear representation. The relationships may be checked using the appropriate values from the computer output.

Table 4 lists the input for the KULLITR2 computer program.

Table 1

UNAIDED DISTANCE VISION; 8577 WOMEN AGED 30-39

Left eye

Right Eye	Highest Grade (1)	Second Grade (2)	Third Grade (3)	Lowest Grade (4)	Sub-Total	Right Only	Total
Highest Grade (1)	1520	266	124	66	1976	140	2116
Second Grade (2)	234	1512	432	78	2256	150	2406
Third Grade (3)	117	362	1772	205	2456	160	2616
Lowest Grade (4)	36	82	179	492	789	50	839
Sub_Total	1907	2222	2507	841	7477	500	7977
Left Only	160	180	200	60	600	*	*
Total	2067	2402	2707	901	8077	*	8577

See Koch, G.G., Imrey, P.B., and Reinfurt, D.W. (1972), Linear model analysis of categorical data with incomplete response vectors, Biometrics 28, 663-692, in particular p.665.

Table 2

	j	ω	x(ω)	x*(ω)	$\tilde{x}(\omega)$	$\hat{x}(\omega)$ [a]	$\tilde{\tilde{x}}(\omega)$ [b]	x**(ω) [c]
1	1	1	1520	1530.227	1530.155	1529.495	1532.573	1531.372
1	2	2	266	267.148	267.151	266.331	253.107	253.216
1	3	3	124	124.403	124.405	123.670	110.966	111.552
1	4	4	66	65.671	65.643	65.573	55.529	56.202
2	1	5	234	234.664	234.676	235.301	247.726	247.955
2	2	6	1512	1512.657	1512.810	1512.672	1515.085	1513.898
2	3	7	432	431.729	431.773	430.600	408.454	408.742
2	4	8	78	77.311	77.282	77.387	69.555	69.857
3	1	9	117	117.190	117.195	117.838	130.215	130.894
3	2	10	362	361.721	361.751	362.784	382.343	382.648
3	3	11	1772	1768.752	1768.905	1769.357	1771.597	1770.209
3	4	12	205	202.944	202.863	203.748	193.837	193.836
4	1	13	36	36.006	36.007	36.114	41.662	42.121
4	2	14	82	81.818	81.822	81.798	90.284	90.690
4	3	15	179	178.413	178.422	177.878	186.975	187.084
4	4	16	492	486.360	486.142	486.528	487.092	486.710
1	.	17	140	132.904	132.898	132.745		
2	.	18	150	150.887	150.899	150.860		
3	.	19	160	163.876	163.884	164.085		
4	.	20	50	52.333	52.320	52.310		
.	1	21	160	153.919	153.915	153.966		
.	2	22	180	178.415	178.430	178.434		
.	3	23	200	200.880	200.897	200.736		
.	4	24	60	66.787	66.759	66.864		
				2I(x*:x) =1.771 6 D.F.	χ^2=1.764 6 D.F.	χ^2=2.33 6 D.F.	χ^2=11.741 3 D.F.	2I(x**:x*) =11.730 3 D.F.

a) See Koch et al (1972) p.669

b),c) Using "improved" estimate to test marginal homogeneity
b) is min. mod. χ^2 and c) is m.d.i.

Table 3

Log-linear representation

i	j	ω	L_1	L_2	L_3	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6
1	1	1	v_1			v_1			v_1		
1	2	2	v_1			v_1				v_1	
1	3	3	v_1			v_1					v_1
1	4	4	v_1			v_1					
2	1	5	v_1				v_1		v_1		
2	2	6	v_1				v_1			v_1	
2	3	7	v_1				v_1				v_1
2	4	8	v_1				v_1				
3	1	9	v_1					v_1	v_1		
3	2	10	v_1					v_1		v_1	
3	3	11	v_1					v_1			v_1
3	4	12	v_1					v_1			
4	1	13	v_1						v_1		
4	2	14	v_1							v_1	
4	3	15	v_1								v_1
4	4	16	v_1								
1	.	17		v_2		$-v_2$					
2	.	18		v_2			$-v_2$				
3	.	19		v_2				$-v_2$			
4	.	20		v_2							
.	1	21			v_3			$-v_3$			
.	2	22			v_3				$-v_3$		
.	3	23			v_3					$-v_3$	
.	4	24			v_3						

$v_1 = 1/w_1 = 1.147118$

$v_2 = 1/w_2 = 17.153992$

$v_3 = 1/w_3 = 14.294999$

$$\ell n \frac{x^*(1)}{x(1)} = v_1 L_1 + \tau_1 v_1 + \tau_4 v_1$$

etc.

$$\ell n \frac{x^*(17)}{x(17)} = v_2 L_2 - \tau_1 v_2$$

etc.

$$\ell n \frac{x^*(21)}{x(21)} = v_3 L_3 - \tau_4 v_3$$

etc.

$$\ell n \frac{x^*(1)}{x(1)} - \ell n \frac{x^*(4)}{x(4)} = v_1 \tau_4 = \ell n \frac{x^*(5)}{x(5)} - \ell n \frac{x^*(8)}{x(8)}$$

or

$$\ell n \frac{x^*(1)\, x^*(8)}{x^*(4)\, x^*(5)} = \ell n \frac{x(1)\, x(8)}{x(4)\, x(5)}$$

etc.

Certain associations are retained.

Table 4

Input for KULLITR2 Computer Program

```
TITLE = 'SEVERAL SAMPLES'  TOL1 =.001  TOL2 =.001
INTERNAL ='0'B
NUMSET = 3  BMAT = '1'B  CNSTRNT = 9  OBS = 24;
16  4  4

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1
1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0-1 0 0 0 0 0 0 0
0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0-1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0-1 0 0 0 0 0
1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0-1 0 0 0
0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0-1 0 0
0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0-1 0

1520 266 124 66 234 1512 432 78 117 362
1772 205 36 82 179 492 140 150 160 50
160 180 200 60

8577 8577 8577 0 0 0 0 0 0
```

SEVERAL SAMPLES , Incomplete Data

B_MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
4	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0
5	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	-1	0	0	0	0	0
7	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	-1	0	0	0
8	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	-1	0	0
9	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	-1	0

WEIGHT(1)= 0.871750
WEIGHT(2)= 0.058295
WEIGHT(3)= 0.069955

INV_WEIGHT(1)= 1.147118
INV_WEIGHT(2)= 17.153992
INV_WEIGHT(3)= 14.294995

C_DESIGN MATRIX

	1	2	3	4	5	6	7	8	9	10
1	1.147115	1.147118	1.147118	1.147118	1.147118	1.147118	1.147118	1.147118	1.147118	1.147118
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	1.147118	1.147118	1.147118	1.147118	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000	1.147118	1.147118	1.147118	1.147118	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.147118	1.147118
7	1.147118	0.000000	0.000000	0.000000	1.147118	0.000000	0.000000	0.000000	1.147118	0.000000
8	0.000000	1.147118	0.000000	0.000000	0.000000	1.147118	0.000000	0.000000	0.000000	1.147118
9	0.000000	0.000000	1.147118	0.000000	0.000000	0.000000	1.147118	0.000000	0.000000	0.000000

	11	12	13	14	15	16	17	18	19	20
1	1.147118	1.147118	1.147118	1.147118	1.147118	1.147118	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	17.153992	17.153992	17.153992	17.153992
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-17.153992	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-17.153992	0.000000	0.000000
6	1.147118	1.147118	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-17.153992	0.000000
7	0.000000	0.000000	1.147118	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	1.147118	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	1.147118	0.000000	0.000000	0.000000	1.147118	0.000000	0.000000	0.000000	0.000000	0.000000

21	22	23	24

1	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000
3	14.294999	14.294999	14.294999	14.294999
4	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000
7	-14.294999	0.000000	0.000000	0.000000
8	0.000000	-14.294999	0.000000	0.000000
9	0.000000	0.000000	-14.294999	0.000000

OBSERVED VALUES

X(1)= 1520.000000	LN_X(1)= 7.326466
X(2)= 266.000000	LN_X(2)= 5.583496
X(3)= 124.000000	LN_X(3)= 4.820282
X(4)= 66.000000	LN_X(4)= 4.189654
X(5)= 234.000000	LN_X(5)= 5.455321
X(6)= 1512.000000	LN_X(6)= 7.321189
X(7)= 432.000000	LN_X(7)= 6.068426
X(8)= 78.000000	LN_X(8)= 4.356709
X(9)= 117.000000	LN_X(9)= 4.762174
X(10)= 362.000000	LN_X(10)= 5.891644
X(11)= 1772.000000	LN_X(11)= 7.479864
X(12)= 205.000000	LN_X(12)= 5.323010
X(13)= 36.000000	LN_X(13)= 3.583519
X(14)= 82.000000	LN_X(14)= 4.406719
X(15)= 179.000000	LN_X(15)= 5.187386
X(16)= 492.000000	LN_X(16)= 6.198479
X(17)= 140.000000	LN_X(17)= 4.941643
X(18)= 150.000000	LN_X(18)= 5.010635
X(19)= 160.000000	LN_X(19)= 5.075173
X(20)= 50.000000	LN_X(20)= 3.912023
X(21)= 160.000000	LN_X(21)= 5.075173
X(22)= 180.000000	LN_X(22)= 5.192957
X(23)= 200.000000	LN_X(23)= 5.298317
X(24)= 60.000000	LN_X(24)= 4.094344

CONSTRAINTS

NTHETA(1)= 8577.000000
NTHETA(2)= 8577.000000
NTHETA(3)= 8577.000000
NTHETA(4)= 0.000000
NTHETA(5)= 0.000000
NTHETA(6)= 0.000000
NTHETA(7)= 0.000000
NTHETA(8)= 0.000000
NTHETA(9)= 0.000000

ESTIMATE OF NTHETA AT COUNT= 1
NTHAT(1)= 8576.996094
NTHAT(2)= 8576.992188
NTHAT(3)= 8576.996094
NTHAT(4)= -134.854431
NTHAT(5)= 14.798584
NTHAT(6)= 72.682190
NTHAT(7)= -59.046561
NTHAT(8)= -24.204498

NTHAT(9)= 0.624036

S

	1	2	3	4	5	6	7
1	9838.824219	0.000000	0.000000	2600.176514	2968.622559	3231.798340	2509.380859
2	0.000000	147129.687500	0.000000	-41196.320313	-44138.914063	-47061.507813	0.000000
3	0.000000	0.000000	122608.187500	0.000000	0.000000	0.000000	-32695.519531
4	2600.176514	-41196.320313	0.000000	43790.496094	0.000000	0.000000	2000.135742
5	2968.622559	-44138.914063	0.000000	0.000000	47107.535156	0.000000	307.915527
6	3231.798340	-47061.507813	0.000000	0.000000	0.000000	50313.304688	153.957809
7	2509.380859	0.000000	-32695.519531	2000.135742	307.915527	153.957809	35204.85[illegible]
8	2923.882568	0.000000	-36782.457031	350.023682	1989.008643	476.347900	0.000000
9	3298.908203	0.000000	-40869.398438	163.168961	568.459473	2331.737061	0.000000

	8	9
1	2923.882568	3298.908203
2	0.000000	0.000000
3	-36782.457031	-40869.398438
4	350.023682	163.168961
5	1989.008643	568.459473
6	476.347900	2331.737061
7	0.000000	0.000000
8	39706.339844	0.000000
9	0.000000	44168.304688

S22.1

	1	2	3	4	5	6
1	31574.355469	-13143.433594	-14036.910156	1336.963623	-422.691650	-708.656982
2	-13143.433594	32970.148438	-15099.566406	-449.228027	1107.399170	-426.904541
3	-14036.910156	-15099.566406	34185.656250	-670.308594	-484.071533	1248.131348
4	1336.963623	-449.228027	-670.308594	25846.074219	-10554.386719	-11739.686719
5	-422.691650	1107.399170	-484.071533	-10554.386719	27802.687500	-13241.179688
6	-708.656982	-426.904541	1248.131348	-11739.686719	-13241.179688	29439.062500

S22.1_INV

	1	2	3	4	5	6
1	0.000086	0.000064	0.000064	-0.000004	-0.000003	-0.000003
2	0.000064	0.000085	0.000064	-0.000003	-0.000004	-0.000003
3	0.000064	0.000064	0.000083	-0.000003	-0.000003	-0.000003
4	-0.000004	-0.000003	-0.000003	0.000103	0.000075	0.000075
5	-0.000003	-0.000004	-0.000003	0.000075	0.000100	0.000075
6	-0.000003	-0.000003	-0.000003	0.000075	0.000075	.000098

```
DELTA(1)=    0.003906
DELTA(2)=    0.007613
DELTA(3)=    0.003906
DELTA(4)=  134.654431
DELTA(5)=  -14.796564
DELTA(6)=  -72.682190
DELTA(7)=   99.646561
DELTA(8)=   24.204498
DELTA(9)=  -16.824036


XSW=     1.763553

ESTIMATE OF X AT COUNT=  1
XSTAR( 1)= 1530.171631      LN_XSTAR( 1)=   7.333136
XSTAR( 2)=  267.150879      LN_XSTAR( 2)=   5.587813
XSTAR( 3)=  124.404739      LN_XSTAR( 3)=   4.823541
XSTAR( 4)=   65.643082      LN_XSTAR( 4)=   4.184232
XSTAR( 5)=  234.674194      LN_XSTAR( 5)=   5.458199
XSTAR( 6)= 1512.794678      LN_XSTAR( 6)=   7.321714
XSTAR( 7)=  431.769287      LN_XSTAR( 7)=   6.067891
XSTAR( 8)=   77.284515      LN_XSTAR( 8)=   4.347493
XSTAR( 9)=  117.193665      LN_XSTAR( 9)=   4.763827
XSTAR(10)=  361.747559      LN_XSTAR(10)=   5.890946
XSTAR(11)= 1768.689160      LN_XSTAR(11)=   7.478107
XSTAR(12)=  202.871552      LN_XSTAR(12)=   5.312573
XSTAR(13)=   36.006302      LN_XSTAR(13)=   3.583694
XSTAR(14)=   81.821625      LN_XSTAR(14)=   4.404541
XSTAR(15)=  178.421432      LN_XSTAR(15)=   5.184149
XSTAR(16)=  486.171631      LN_XSTAR(16)=   6.186562
XSTAR(17)=  133.000092      LN_XSTAR(17)=   4.890349
XSTAR(18)=  150.816544      LN_XSTAR(18)=   5.016064
XSTAR(19)=  163.838669      LN_XSTAR(19)=   5.098883
XSTAR(20)=   52.345062      LN_XSTAR(20)=   3.957657
XSTAR(21)=  153.895798      LN_XSTAR(21)=   5.036276
XSTAR(22)=  178.282852      LN_XSTAR(22)=   5.183372
XSTAR(23)=  200.725555      LN_XSTAR(23)=   5.301937
XSTAR(24)=   67.096619      LN_XSTAR(24)=   4.206154



2I(XSTAR:X)=     1.837387

TAU(1)= 0.005662
TAU(2)= 0.002355
TAU(3)= 0.001290
TAU(4)= 0.010541
TAU(5)= 0.008491
TAU(6)= 0.007567
```

```
ESTIMATE OF X AT COUNT= 15
XSTAR( 1)= 1530.227051      LN_XSTAR( 1)=  7.333172
XSTAR( 2)=  267.147705      LN_XSTAR( 2)=  5.587802
XSTAR( 3)=  124.403076      LN_XSTAR( 3)=  4.823527
XSTAR( 4)=   65.670639      LN_XSTAR( 4)=  4.184651
XSTAR( 5)=  234.664124      LN_XSTAR( 5)=  5.458156
XSTAR( 6)= 1512.657471      LN_XSTAR( 6)=  7.321624
XSTAR( 7)=  431.729004      LN_XSTAR( 7)=  6.067799
XSTAR( 8)=   77.310837      LN_XSTAR( 8)=  4.347834
XSTAR( 9)=  117.190430      LN_XSTAR( 9)=  4.763800
XSTAR(10)=  361.720703      LN_XSTAR(10)=  5.890872
XSTAR(11)= 1768.752441      LN_XSTAR(11)=  7.478030
XSTAR(12)=  202.943741      LN_XSTAR(12)=  5.312929
XSTAR(13)=   36.006409      LN_XSTAR(13)=  3.583696
XSTAR(14)=   81.818115      LN_XSTAR(14)=  4.404499
XSTAR(15)=  176.413269      LN_XSTAR(15)=  5.134103
XSTAR(16)=  486.360107      LN_XSTAR(16)=  6.186950
XSTAR(17)=  132.904297      LN_XSTAR(17)=  4.889629
XSTAR(18)=  150.886810      LN_XSTAR(18)=  5.016530
XSTAR(19)=  163.876404      LN_XSTAR(19)=  5.099115
XSTAR(20)=   52.332794      LN_XSTAR(20)=  3.957625
XSTAR(21)=  153.919235      LN_XSTAR(21)=  5.036428
XSTAR(22)=  178.414627      LN_XSTAR(22)=  5.184111
XSTAR(23)=  200.879913      LN_XSTAR(23)=  5.302706
XSTAR(24)=   66.786652      LN_XSTAR(24)=  4.201503


ZI(XSTAR:X)=      1.770519

TAU(1)= 0.005690
TAU(2)= 0.002315
TAU(3)= 0.001263
TAU(4)= 0.010207
TAU(5)= 0.008115
TAU(6)= 0.007189
```

```
ESTIMATE OF NTHETA AT COUNT=  15
NTHAT(1)= 8577.019531
NTHAT(2)= 8577.003906
NTHAT(3)= 8577.003906
NTHAT(4)=     0.000290
NTHAT(5)=    -0.001691
NTHAT(6)=    -0.001867
NTHAT(7)=    -0.001707
NTHAT(8)=    -0.000347
NTHAT(9)=    -0.002171
```

S

	1	2	3	4	5	6	7
1	9838.647656	0.000000	0.000000	2615.244141	2969.098389	3224.702881	2523.973877
2	0.000000	147129.812500	0.000000	-39108.343750	-44399.914063	-48222.226563	0.000000
3	0.000000	0.000000	122608.250000	0.000000	0.000000	0.000000	-31452.949219
4	2615.244141	-39108.343750	0.000000	41723.589844	0.000000	0.000000	2013.595459
5	2969.098389	-44399.914063	0.000000	0.000000	47369.011719	0.000000	308.789551
6	3224.702881	-48222.226563	0.000000	0.000000	0.000000	51446.929688	154.268557
7	2523.973877	0.000000	-31452.949219	2013.595459	308.789551	154.268557	33976.921875
8	2925.652100	0.000000	-36458.511719	351.534424	1990.474121	475.980469	0.000000
9	3294.036865	0.000000	-41049.222656	163.699387	568.103027	2327.464111	0.000000

	8	9
1	2925.652100	3294.036865
2	0.000000	0.000000
3	-36458.511719	-41049.222656
4	351.534424	163.699387
5	1990.474121	568.103027
6	475.980469	2327.464111
7	0.000000	0.000000
8	39384.164063	0.000000
9	0.000000	44343.261719

S22.1

	1	2	3	4	5	6
1	30635.109375	-12591.078125	-13675.023438	1342.703125	-426.126953	-711.881592
2	-12591.078125	33074.289063	-15525.328125	-452.877441	1107.591309	-425.947998
3	-13675.023438	-15525.328125	34585.050781	-673.029053	-482.907959	1247.836670
4	1342.703125	-452.877441	-673.029053	25260.750000	-10103.296875	-11375.464844
5	-426.126953	1107.591309	-482.907959	-10103.296875	27672.980469	-13165.804688
6	-711.881592	-425.947998	1247.836670	-11375.464844	-13165.804688	29497.148438

S22.1_INV

35

	1	2	3	4	5	6
1	0.000085	0.000061	0.000061	-0.000004	-0.000003	-0.000002
2	0.000061	0.000082	0.000061	-0.000003	-0.000005	-0.000002
3	0.000061	0.000061	0.000083	-0.000002	-0.000002	-0.000003
4	-0.000004	-0.000003	-0.000002	0.000097	0.000060	[illegible]
5	-0.000003	-0.000003	-0.000002	0.000058	0.000095	[illegible]
6	-0.000002	-0.000002	-0.000003	0.000068	0.000065	[illegible]

DELTA(1)=-0.015531
DELTA(2)=-0.003906
DELTA(3)=-0.003906
DELTA(4)=-0.000290
DELTA(5)= 0.001691
DELTA(6)= 0.001867
DELTA(7)= 0.001707
DELTA(8)= 0.000347
DELTA(9)= 0.002171

OUTLIER(1)= 0.008579
OUTLIER(2)= 0.004941
OUTLIER(3)= 0.001308
OUTLIER(4)= 0.001648
OUTLIER(5)= 0.001882
OUTLIER(6)= 0.000285
OUTLIER(7)= 0.000170
OUTLIER(8)= 0.006116
OUTLIER(9)= 0.000310
OUTLIER(10)= 0.000216
OUTLIER(11)= 0.005957
OUTLIER(12)= 0.020730
OUTLIER(13)= 0.000001
OUTLIER(14)= 0.000404
OUTLIER(15)= 0.001926
OUTLIER(16)= 0.065025
OUTLIER(17)= 0.369070
OUTLIER(18)= 0.005227
OUTLIER(19)= 0.092795
OUTLIER(20)= 0.106375
OUTLIER(21)= 0.235604
OUTLIER(22)= 0.014025
OUTLIER(23)= 0.003863
OUTLIER(24)= 0.727247

ITERATIONS=15

TOL1=0.1000 TOL2=0.0100

S_INV

	1	2	3	4	5	6	7	8	9
1	0.000219	-0.000054	-0.000064	-0.000060	-0.000060	-0.000061	-0.000071	-0.000071	-0.000071
2	-0.000054	0.000064	-0.000003	0.000064	0.000064	0.000064	-0.000003	-0.000003	-0.000003
3	-0.000064	-0.000003	0.000076	-0.000003	-0.000003	-0.000003	0.000075	0.000075	0.000075
4	-0.000060	0.000064	-0.000003	0.000086	0.000064	0.000064	-0.000004	-0.000003	-0.000003
5	-0.000060	0.000064	-0.000003	0.000064	0.000065	0.000064	-0.000003	-0.000004	-0.000003
6	-0.000061	0.000064	-0.000003	0.000064	0.000064	0.000083	-0.000003	-0.000003	-0.000003
7	-0.000071	-0.000003	0.000075	-0.000004	-0.000003	-0.000003	0.000103	0.000075	0.000075
8	-0.000071	-0.000003	0.000075	-0.000003	-0.000004	-0.000003	0.000075	0.000100	0.000075
9	-0.000071	-0.000003	0.000075	-0.000003	-0.000003	-0.000003	0.000075	0.000075	0.000096

LAMBDA(1)=-0.010380
LAMBDA(2)= 0.002705
LAMBDA(3)= 0.007881
LAMBDA(4)= 0.005662
LAMBDA(5)= 0.002356
LAMBDA(6)= 0.001290
LAMBDA(7)= 0.010541
LAMBDA(8)= 0.008491
LAMBDA(9)= 0.007567

MU(1)= 0.000681
MU(2)= 0.004329
MU(3)= 0.003269
MU(4)=-0.005411
MU(5)= 0.002888
MU(6)= 0.000536
MU(7)=-0.000524
MU(8)=-0.009204
MU(9)= 0.001665
MU(10)=-0.000687
MU(11)=-0.001746
MU(12)=-0.010427
MU(13)= 0.003185
MU(14)=-0.002167
MU(15)=-0.003226
MU(16)=-0.011907
MU(17)=-0.050729
MU(18)= 0.005993
MU(19)= 0.024272
MU(20)= 0.046402
MU(21)=-0.038034
MU(22)=-0.008722
MU(23)= 0.004483
MU(24)= 0.112653

XSQ= 1.763577

MINIMUM MODIFIED CHI SQ ESTIMATE

XHAT(1)=	1530.154765	LN_XHAT(1)=	7.333124
XHAT(2)=	267.151367	LN_XHAT(2)=	5.587815
XHAT(3)=	124.405350	LN_XHAT(3)=	4.823545
XHAT(4)=	65.642853	LN_XHAT(4)=	4.184229
XHAT(5)=	234.675735	LN_XHAT(5)=	5.458204
XHAT(6)=	1512.809814	LN_XHAT(6)=	7.321724
XHAT(7)=	431.773438	LN_XHAT(7)=	6.067901

XHAT(8)=	77.262057	LN_XHAT(8)=	4.347462
XHAT(9)=	117.194837	LN_XHAT(9)=	4.763838
XHAT(10)=	361.751221	LN_XHAT(10)=	5.890957
XHAT(11)=	1768.905273	LN_XHAT(11)=	7.478116
XHAT(12)=	202.862534	LN_XHAT(12)=	5.312529
XHAT(13)=	36.006660	LN_XHAT(13)=	3.583704
XHAT(14)=	81.822327	LN_XHAT(14)=	4.404550
XHAT(15)=	178.422485	LN_XHAT(15)=	5.184155
XHAT(16)=	486.141640	LN_XHAT(16)=	6.186501
XHAT(17)=	132.697083	LN_XHAT(17)=	4.888061
XHAT(18)=	150.898883	LN_XHAT(18)=	5.016610
XHAT(19)=	163.883545	LN_XHAT(19)=	5.099156
XHAT(20)=	52.320099	LN_XHAT(20)=	3.957380
XHAT(21)=	153.914551	LN_XHAT(21)=	5.036398
XHAT(22)=	178.429962	LN_XHAT(22)=	5.184196
XHAT(23)=	200.896591	LN_XHAT(23)=	5.302791
XHAT(24)=	66.759155	LN_XHAT(24)=	4.201092

2I(XHAT:X)= 1.746163

Example 5. Specified log-linear representation.

In this example the problem specifies the form of the log-linear representation and consequently the design matrix. The general linear hypothesis approach is necessary.

Example

Specified Log-linear Representation

D.V. Gokhale, "Analysis of log-linear models" <u>Jour. Royal Statistical Soc.</u> Series B Vol 34 (1972) p. 371-376 formulates a problem for a 2 x 2 x 3 three-way contingency table of fitting a model such that the log-linear representation is of the form

$$\ell n \frac{x^*(ijk)}{n\pi} = L + (i-1)\tau^{i} + (j-1)\tau^{j} + (k-1)\tau^{k}$$
$$+ (i-1)(j-1)\tau^{ij} + (i-1)(k-1)\tau^{ik} + (j-1)(k-1)\tau^{jk}$$
$$+ (i-1)(j-1)(k-1)\tau^{ijk}$$

This implies that the graphic version of the log-linear representation is as given in Fig. 1.

				1	2	3	4	5	6	7	8
ω	i	j	k	L	τ^{i}	τ^{j}	τ^{k}	τ^{ij}	τ^{ik}	τ^{jk}	τ^{ijk}
1	1	1	1	1	0	0	0	0	0	0	0
2	1	1	2	1	0	0	1	0	0	0	0
3	1	1	3	1	0	0	2	0	0	0	0
4	1	2	1	1	0	1	0	0	0	0	0
5	1	2	2	1	0	1	1	0	0	1	0
6	1	2	3	1	0	1	2	0	0	2	0
7	2	1	1	1	1	0	0	0	0	0	0
8	2	1	2	1	1	0	1	0	1	0	0
9	2	1	3	1	1	0	2	0	2	0	0
10	2	2	1	1	1	1	0	1	0	0	0
11	2	2	2	1	1	1	1	1	1	1	1
12	2	2	3	1	1	1	2	1	2	2	2

Figure 1

The observed values (fictitious) are

i	j	k	x(ijk)	ijk	x(ijk)
1	1	1	58	211	75
1	1	2	49	212	58
1	1	3	33	213	45
1	2	1	11	221	19
1	2	2	14	222	17
1	2	3	18	223	22

Gokhale used an iterative procedure that might be described as a "steepest descent" procedure. We shall set this up using the k-sample algorithm (of course here k=1) and using the uniform distribution as the initial distribution. In this case the $\underline{C}$ matrix is the transpose of the $\underline{T}$ matrix in Fig. 1 and is given again for convenience in Fig. 2.

i	1	1	1	1	1	1	2	2	2	2	2	2
j	1	1	1	2	2	2	1	1	1	2	2	2
k	1	2	3	1	2	3	1	2	3	1	2	3
ω	1	2	3	4	5	6	7	8	9	10	11	12
	1	1	1	1	1	1	1	1	1	1	1	1
	0	0	0	0	0	0	1	1	1	1	1	1
	0	0	0	1	1	1	0	0	0	1	1	1
	0	1	2	0	1	2	0	1	2	0	1	2
	0	0	0	0	[illegible]	0	0	0	0	1	1	1
	0	0	0	0	[illegible]	[illegible]	0	1	2	0	1	2
	0	0	0	0	[illegible]	[illegible]	0	0	0	0	1	2
	0	0	0	0	[illegible]	[illegible]	0	0	0	0	1	2

Figure 2

The estimated values as given by Gokhale are

i	j	k	$x_G^*(ijk)$	ijk	$x_G^*(ijk)$
1	1	1	59.73	211	74.97
1	1	2	45.54	212	58.06
1	1	3	34.73	213	44.97
1	2	1	10.98	221	17.85
1	2	2	14.05	222	19.29
1	2	3	17.98	223	20.85

The goodness-of-fit X^2 statistic is 0.8083, 4 D.F.

The input values for the KULLITR2 computer program are given in table 1.

The input values for the DARRAT computer program are given in table 2.

```
TITLE='GOKHALE ANALYSIS'
OBS= 12
CNSTRNT= 8
FACTORS= 3
TOL 1= .001   TOL 2= .001   ;
FACNAME(1)= 'I'
FACNAME(2)= 'J'
FACNAME(3)= 'K'   ;
2   2   3

1   1   1   1   1   1   1   1   1   1   1   1
0   0   0   0   0   0   1   1   1   1   1   1
0   0   0   1   1   1   0   0   0   1   1   1
0   1   2   0   1   2   0   1   2   0   1   2
0   0   0   0   0   0   0   0   0   1   1   1
0   0   0   0   0   0   0   1   2   0   1   2
0   0   0   0   1   2   0   0   0   0   1   2
0   0   0   0   0   0   0   0   0   0   1   2

58   49   33   11   14   18   75   58   45
 19   17   22
```

Table 1

Input to KULLITR2 Computer Program

```
TITLE='GOKHALE"S ANALYSIS'

BLOCKS=8

TOL1=.001  TOL2=.001  CNSTRNT=8

OBS=12  FACTORS=3;

FACNAME(1)='I'  FACNAME(2)='J'  FACNAME(3)='K';

2 2 3

1 1 1 1 1 1 1 1

1 1 1 1 1 1 1 1 1 1 1 1

0 0 0 0 0 0 1 1 1 1 1 1

0 0 0 1 1 1 0 0 0 1 1 1

0 1 2 0 1 2 0 1 2 0 1 2

0 0 0 0 0 0 0 0 0 1 1 1

0 0 0 0 0 0 0 1 2 0 1 2

0 0 0 0 1 2 0 0 0 0 1 2


0 0 0 0 0 0 0 0 0 0 1 2


58 49 33 11 14 18 75 58 45 19 17 22
```

Table 2.

Input to DARRAT Computer Program

GOKHALE'S ANALYSIS

3 FACTOR TABLE:I*J*K

C DESIGN MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	1	1	1	1	1	1	1	1	1	1
2	0	0	0	0	0	0	1	1	1	1	1	1
3	0	0	0	1	1	1	0	0	0	1	1	1
4	0	1	2	0	1	2	0	1	2	0	1	2
5	0	0	0	0	0	0	0	0	0	1	1	1
6	0	0	0	0	0	0	0	1	2	0	1	2
7	0	0	0	0	1	2	0	0	0	0	1	2
8	0	0	0	0	0	0	0	0	0	0	1	2

OBSERVED VALUES

1	1	1	X(1)= 58.000000	LN_X(1)= 4.060443
1	1	2	X(2)= 49.000000	LN_X(2)= 3.891820
1	1	3	X(3)= 33.000000	LN_X(3)= 3.496508
1	2	1	X(4)= 11.000000	LN_X(4)= 2.397895
1	2	2	X(5)= 14.000000	LN_X(5)= 2.639057
1	2	3	X(6)= 18.000000	LN_X(6)= 2.890371
2	1	1	X(7)= 75.000000	LN_X(7)= 4.317488
2	1	2	X(8)= 58.000000	LN_X(8)= 4.060443
2	1	3	X(9)= 45.000000	LN_X(9)= 3.806663
2	2	1	X(10)= 19.000000	LN_X(10)= 2.944439
2	2	2	X(11)= 17.000000	LN_X(11)= 2.833213
2	2	3	X(12)= 22.000000	LN_X(12)= 3.091043

CONSTRAINTS
NTHETA(1)= 419.000000
NTHETA(2)= 236.000000
NTHETA(3)= 101.000000
NTHETA(4)= 374.000000
NTHETA(5)= 58.000000
NTHETA(6)= 209.000000
NTHETA(7)= 111.000000
NTHETA(8)= 61.000000

ESTIMATE OF NTHETA AT COUNT= 1
NTHAT(1)= 418.999756
NTHAT(2)= 209.499939
NTHAT(3)= 209.499939
NTHAT(4)= 418.999756
NTHAT(5)= 104.749969
NTHAT(6)= 209.499939
NTHAT(7)= 209.499939
NTHAT(8)= 104.749969

S22.1_INV

	1	2	3	4	5	6	7
1	0.047733	0.023866	0.014320	-0.047733	-0.028640	-0.014320	0.028640
2	0.023866	0.047733	0.014320	-0.047733	-0.014320	-0.028640	0.028640
3	0.014320	0.014320	0.014320	-0.014320	-0.014320	-0.014320	0.014320
4	-0.047733	-0.047733	-0.014320	0.095465	0.028640	0.028640	-0.057279
5	-0.028640	-0.014320	-0.014320	0.028640	0.028640	0.014320	-0.028640
6	-0.014320	-0.028640	-0.014320	0.028640	0.014320	0.028640	-0.028640
7	0.028640	0.028640	0.014320	-0.057279	-0.028640	-0.028640	0.057279

DELTA(1)= 0.000244
DELTA(2)= 26.500061
DELTA(3)=-108.499939
DELTA(4)= -44.999756
DELTA(5)= -46.749969
DELTA(6)= -0.499939
DELTA(7)= -98.499939
DELTA(8)= -43.749969

XS0= 143.018844

ESTIMATE OF X AT COUNT= 1

1	1	1	XSTAR(1)= 58.329590	LN_XSTAR(1)= 4.066110
1	1	2	XSTAR(2)= 40.777039	LN_XSTAR(2)= 3.708119
1	1	3	XSTAR(3)= 28.506424	LN_XSTAR(3)= 3.350129

1	2	1	XSTAR(4)= 14.612355	LN_XSTAR(4)= 2.681868
1	2	2	XSTAR(5)= 16.153030	LN_XSTAR(5)= 2.782105
1	2	3	XSTAR(6)= 17.856094	LN_XSTAR(6)= 2.882344
2	1	1	XSTAR(7)= 90.060471	LN_XSTAR(7)= 4.500482
2	1	2	XSTAR(8)= 58.608948	LN_XSTAR(8)= 4.070883
2	1	3	XSTAR(9)= 38.141159	LN_XSTAR(9)= 3.641294
2	2	1	XSTAR(10)= 17.856079	LN_XSTAR(10)= 2.882344
2	2	2	XSTAR(11)= 18.639893	LN_XSTAR(11)= 2.925305
2	2	3	XSTAR(12)= 19.458115	LN_XSTAR(12)= 2.968264

Z IS OBSERVED TABLE AND X IS INITIAL DIST.

2I(XSTAR:X)= 154.655060

2I(Z:XSTAR)= 7.895874

TAU(1)= 0.434372
TAU(2)=-1.384242
TAU(3)=-0.357990
TAU(4)=-0.233895
TAU(5)=-0.071604
TAU(6)= 0.458228
TAU(7)= 0.014325

ESTIMATE OF X AT COUNT= 5

1	1	1	XSTAR(1)= 59.727356	LN_XSTAR(1)= 4.089790
1	1	2	XSTAR(2)= 45.543640	LN_XSTAR(2)= 3.818671
1	1	3	XSTAR(3)= 34.728195	LN_XSTAR(3)= 3.547552
1	2	1	XSTAR(4)= 10.976489	LN_XSTAR(4)= 2.395756
1	2	2	XSTAR(5)= 14.047031	LN_XSTAR(5)= 2.642411
1	2	3	XSTAR(6)= 17.976517	LN_XSTAR(6)= 2.889066
2	1	1	XSTAR(7)= 74.968704	LN_XSTAR(7)= 4.317071
2	1	2	XSTAR(8)= 58.062469	LN_XSTAR(8)= 4.061520
2	1	3	XSTAR(9)= 44.968765	LN_XSTAR(9)= 3.805968
2	2	1	XSTAR(10)= 17.852737	LN_XSTAR(10)= 2.882156
2	2	2	XSTAR(11)= 19.294540	LN_XSTAR(11)= 2.959822
2	2	3	XSTAR(12)= 20.852768	LN_XSTAR(12)= 3.037466

Z IS OBSERVED TABLE AND X IS INITIAL DIST.

2I(XSTAR:X)= 141.138519

Z11(2;XSTAR)=   0.814035

TAU(1)= 0.227281
TAU(2)=-1.694035
TAU(3)=-0.271119
TAU(4)= 0.259120
TAU(5)= 0.015567
TAU(6)= 0.517774
TAU(7)=-0.184558

ESTIMATE OF NTHETA AT COUNT=   5
NTHAT(1)= 418.999023
NTHAT(2)= 236.000046
NTHAT(3)= 101.000061
NTHAT(4)= 374.000000
NTHAT(5)= 58.000046
NTHAT(6)= 209.000046
NTHAT(7)= 111.000046
NTHAT(8)= 61.000046

S22.1

	1	2	3	4	5	6	7
1	103.073669	1.112041	-1.654413	25.331680	91.281342	-1.520421	26.641937
2	1.112041	76.653900	20.847031	44.019073	10.620414	84.243393	46.295929
3	-1.654413	20.847031	277.218506	9.229003	154.088654	89.579483	48.256099
4	25.331680	44.019073	9.229003	49.971375	32.069153	45.634827	52.556091
5	91.281342	10.620414	154.088654	32.069153	236.391917	47.337645	72.278244
6	-1.520421	84.243393	89.579483	45.634827	47.337845	159.252731	86.545576
7	26.641937	46.295929	48.256699	52.556091	72.278244	86.545578	[illegible]

S22.1 INV

	1	2	3	4	5	6	7
1	0.026277	0.014641	0.009128	-0.026277	-0.016365	-0.009128	0.016365

2	0.014641	0.086509	0.009128	-0.086509	-0.009128	-0.050935	0.050935
3	0.009128	0.009128	0.011112	-0.009128	-0.011112	-0.011112	0.011112
4	-0.026277	-0.086509	-0.009128	0.144079	0.016365	0.050935	-0.085454
5	-0.016365	-0.009128	-0.011112	0.016365	0.019817	0.011112	-0.019817
6	-0.009128	-0.050935	-0.011112	0.050935	0.011112	0.047066	-0.047066
7	0.016365	0.050935	0.011112	-0.085454	-0.019817	-0.047066	0.081741

DELTA(1)= 0.000977
DELTA(2)=-0.000046
DELTA(3)=-0.000061
DELTA(4)= 0.000000
DELTA(5)=-0.000046
DELTA(6)=-0.000046
DELTA(7)=-0.000046
DELTA(8)=-0.000046

1	1	1	OUTLIERZ(1)= 0.050692
1	1	2	OUTLIERZ(2)= 0.252830
1	1	3	OUTLIERZ(3)= 0.088214
1	2	1	OUTLIERZ(4)= 0.000050
1	2	2	OUTLIERZ(5)= 0.000158
1	2	3	OUTLIERZ(6)= 0.000031
2	1	1	OUTLIERZ(7)= 0.000013
2	1	2	OUTLIERZ(8)= 0.000067
2	1	3	OUTLIERZ(9)= 0.000022
2	2	1	OUTLIERZ(10)= 0.071454
2	2	2	OUTLIERZ(11)= 0.290508
2	2	3	OUTLIERZ(12)= 0.061441

```
1  1  1     OUTLIERX( 1)= 13.319021
1  1  2     OUTLIERX( 2)=  2.823663
1  1  3     OUTLIERX( 3)=  0.001020
1  2  1     OUTLIERX( 4)= 27.703735
1  2  2     OUTLIERX( 5)= 19.002884
1  2  3     OUTLIERX( 6)= 11.246520
2  1  1     OUTLIERX( 7)= 30.604034
2  1  2     OUTLIERX( 8)= 11.770920
2  1  3     OUTLIERX( 9)=  2.543224
2  2  1     OUTLIERX(10)= 11.446600
2  2  2     OUTLIERX(11)=  9.266130
2  2  3     OUTLIERX(12)=  7.249615


ITERATIONS= 5

TOL1=0.0010    TOL2=0.0010
```

GOKHALE'S ANALYSIS

3 FACTOR TABLE:I*J*K

DARRAT

C_DESIGN MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	1	1	1	1	1	1	1	1	1	1
2	0	0	0	0	0	0	1	1	1	1	1	1
3	0	0	0	1	1	1	0	0	0	1	1	1
4	0	1	2	0	1	2	0	1	2	0	1	2
5	0	0	0	0	0	0	0	0	0	1	1	1
6	0	0	0	0	0	0	0	1	2	0	1	2
7	0	0	0	0	1	2	0	0	0	0	1	2
8	0	0	0	0	0	0	0	0	0	0	1	2

OBSERVED VALUES

1	1	1	X(1)= 58.000000	LN_X(1)= 4.060443
1	1	2	X(2)= 49.000000	LN_X(2)= 3.891820
1	1	3	X(3)= 33.000000	LN_X(3)= 3.496508
1	2	1	X(4)= 11.000000	LN_X(4)= 2.397895
1	2	2	X(5)= 14.000000	LN_X(5)= 2.639057
1	2	3	X(6)= 18.000000	LN_X(6)= 2.890371
2	1	1	X(7)= 75.000000	LN_X(7)= 4.317488
2	1	2	X(8)= 58.000000	LN_X(8)= 4.060443
2	1	3	X(9)= 45.000000	LN_X(9)= 3.806663
2	2	1	X(10)= 19.000000	LN_X(10)= 2.944439
2	2	2	X(11)= 17.000000	LN_X(11)= 2.833213
2	2	3	X(12)= 22.000000	LN_X(12)= 3.091043

CONSTRAINTS
NTHETA(1)= 419.000000
NTHETA(2)= 256.000000
NTHETA(3)= 101.000000
NTHETA(4)= 374.000000
NTHETA(5)= 58.000000
NTHETA(6)= 209.000000
NTHETA(7)= 111.000000
NTHETA(8)= 61.000000

THE INITIAL DISTRIBUTION IS UNIFORM

ITERATIONS= 63

ESTIMATED DISTRIBUTION

1	1	1	XSTAR(1)= 59.653397	LN_XSTAR(1)= 4.088551
1	1	2	XSTAR(2)= 45.544296	LN_XSTAR(2)= 3.818680
1	1	3	XSTAR(3)= 34.772247	LN_XSTAR(3)= 3.548820
1	2	1	XSTAR(4)= 11.008032	LN_XSTAR(4)= 2.398024
1	2	2	XSTAR(5)= 14.056249	LN_XSTAR(5)= 2.643209
1	2	3	XSTAR(6)= 17.953644	LN_XSTAR(6)= 2.887795
2	1	1	XSTAR(7)= 74.988144	LN_XSTAR(7)= 4.317330

```
  2  1  2      XSTAR( 8)= 58.063187      LN XSTAR( 8)   4.06[illegible]
  2  1  3      XSTAR( 9)= 44.956206      LN_XSTAR( 9)   3.80[illegible]
  2  2  1      XSTAR(10)= 17.850708      LN_XSTAR(10)=  2.88[illegible]43
  2  2  2      XSTAR(11)= 19.294540      LN_XSTAR(11)=  2.959[illegible]22
  2  2  3      XSTAR(12)= 20.855164      LN_XSTAR(12)=  3.037601


ESTIMATED CONSTRAINTS
NTHAT(1)= 418.[illegible]99756
NTHAT(2)= 236.009549
NTHAT(3)= 101.020325
NTHAT(4)= 374.038574
NTHAT(5)=  58.000412
NTHAT(6)= 208.984467
NTHAT(7)= 110.970398
NTHAT(8)=  61.004868


OUTLIER STATISTIC
Z IS OBSERVED TABLE AND X IS INITIAL DISTRIBUTION
  1  1  1      OUTLIERX( 1)= 13.248681
  1  1  2      OUTLIERX( 2)=  2.823992
  1  1  3      OUTLIERX( 3)=  0.000598
  1  2  1      OUTLIERX( 4)= 27.598648
  1  2  2      OUTLIERX( 5)= 18.976028
  1  2  3      OUTLIERX( 6)= 11.283285
  2  1  1      OUTLIERX( 7)= 30.629288
  2  1  2      OUTLIERX( 8)= 11.771581
  2  1  3      OUTLIERX( 9)=  2.558200
  2  2  1      OUTLIERX(10)= 11.449886
  2  2  2      OUTLIERX(11)=  9.266115
  2  2  3      OUTLIERX(12)=  7.246758

  1  1  1      OUTLIERZ( 1)= 0.046474
  1  1  2      OUTLIERZ( 2)= 0.252732
  1  1  3      OUTLIERZ( 3)= 0.092710
  1  2  1      OUTLIERZ( 4)= 0.000006
  1  2  2      OUTLIERZ( 5)= 0.000242
  1  2  3      OUTLIERZ( 6)= 0.000119
  2  1  1      OUTLIERZ( 7)= 0.000002
  2  1  2      OUTLIERZ( 8)= 0.000069
  2  1  3      OUTLIERZ( 9)= 0.000039
  2  2  1      OUTLIERZ(10)= 0.071710
  2  2  2      OUTLIERZ(11)= 0.290512
  2  2  3      OUTLIERZ(12)= 0.061181


2I(XSTAR:X)=   141.023143

2I(Z:XSTAR)=     0.813257

TOL1=        0.010000     TOL2=        0.010000
```

Example 6. Four point bioassay - fit of logistic function. This example illustrates the application of the k-sample procedure to fitting data based on restraints using the observed values. The procedure was also used on the data of examples 1 and 2 of chapter 4, with results the same as there given. It has also been applied in a number of other cases, not given here as additional examples.

We reformulate the data first as the 4 x 2 contingency table 2, with entries x(ij), i=1,...4, j = 1,2

		j = 1 Deaths	j = 2 Alive	
i	1	1	9	10
	2	6	4	10
	3	3	7	10
	4	8	2	10
		18	22	40

Table 2

The log-linear diagram of the representation of the minimum discrimination information estimate is shown in Fig. 1.

ω	i	j	L_1	L_2	L_3	L_4	τ_1	τ_2
1	1	1	1				1	0
2	1	2	1				0	0
3	2	1		1			1	1
4	2	2		1			0	0
5	3	1			1		1	2
6	3	2			1		0	0
7	4	1				1	1	3
8	4	2				1	0	0

Figure 1

For the procedure fitting observed marginals or other restraints, we note that Fig. 1 implies the following relations.

$$x^*(i1) + x^*(i2) = x(i1) + x(i2), \ i = 1,2,3,4 \ ,$$

$$x^*(11) + x^*(21) + x^*(31) + x^*(41) = x(11) + x(21) + x(31) + x(41),$$

$$x^*(21) + 2x^*(31) + 3x^*(41) = x(21) + 2x(31) + 3x(41),$$

$$\ell n \frac{x^*(11)}{x^*(12)} = \tau_1 \ ,$$

$$\ell n \frac{x^*(21)}{x^*(22)} = \tau_1 + \tau_2 \ ,$$

$$\ell n \frac{x^*(31)}{x^*(32)} = \tau_1 + 2\tau_2 \ ,$$

$$\ell n \frac{x^*(41)}{x^*(42)} = \tau_1 + 3\tau_2 \ .$$

For the k-sample algorithm this is a case of 4 samples, two observations per sample. The basic B matrix is given in Fig. 2.

	11	12	21	22	31	32	41	42
ω	1	2	3	4	5	6	7	8
	1	1	0	0	0	0	0	0
	0	0	1	1	0	0	0	0
	0	0	0	0	1	1	0	0
	0	0	0	0	0	0	1	1
	1	0	1	0	1	0	1	0
	0	0	1	0	2	0	3	0

Figure 2

In view of the relations given above between values of the x^*'s and the x's, in this case the C matrix is derived from the B matrix by the relations

$$\underline{B} = \begin{pmatrix} \underline{B}_1 \\ \hline \underline{B}_2 \end{pmatrix}, \quad \underline{C} = \begin{pmatrix} \underline{C}_1 \\ \hline \underline{C}_2 \end{pmatrix}, \quad \underline{C}_1 = \underline{B}_1 \underline{W}^{-1}, \quad \underline{C}_2 = \underline{B}_2 ,$$

where $\underline{B}$ is 6 x 8, $\underline{B}_1$ is 4 x 8, $\underline{B}_2$ is 2 x 8 with similar dimensions for the $\underline{C}$ matrix and its components.

We remark that instead of starting the iteration from the uniform distribution, the initial distribution used in the computer output attached was $x_1^*(ij) = x(i.)x(.j)/N$ as calculated from table 2. We comment that another run using the uniform distribution $N\pi(ij) = 5$ as the initial distribution for the iteration yielded the same final values. The computer input data is given in table 5.

By computing the maximum likelihood estimates of α and β in his formulation, Berkson derived the estimates given in table 3.

Berksons-Estimate (Max. Likelihood)

Deaths	Alive	
1.901431	8.098569	10.000000
3.445099	6.554901	10.000000
5.405505	4.594495	10.000000
7.247965	2.752035	10.000000
18.000000	22.000000	40.000000

Table 3

Berkson gave a value $2I(x:x^*) = 5.985432$, 2 D.F. (on p. 447 of Berkson (1972) the degrees of freedom are incorrectly given as 1).

The m.d.i. estimates after 4 iterations are given in table 4.

M.D.I. Estimate--4 iterations

Deaths	Alive	
1.901434	8.098566	10.000000
3.445101	6.554895	9.999996
5.405508	4.594491	9.999999
7.247968	2.752036	10.000004
18.000011	21.999988	39.999999

Table 4

$2I(x:x^*) = 5.985401$, 2 D.F.

We also have the analysis of information

Analysis of Information

Component due to	Information	D.F.
$x(i.)$, $x(.j)$	$2I(x:x_1^*) = 12.863$	3
$x(.j)$, $x(21)+2x(31)+3x(41)$, $x(i.)$	$2I(x^*:x_1^*)=6.878$	1
	$2I(x:x^*)=5.985$	2

From the output and Fig. 1 we see that since x_1^* was the initial distribution

$$\ln\frac{x^*(1)}{x^*(2)} = \ln\frac{x_1^*(1)}{x_1^*(2)} + \tau_1 \quad \text{or} \quad -1.449079 = -0.200671 - 1.248407 ,$$

$$\ln \frac{x^*(3)}{x^*(4)} = \ln \frac{x_1^*(3)}{x_1^*(4)} + \tau_1 + \tau_2 \quad \text{or } -0.643259 = -0.200671$$
$$-1.248407 + 0.805820 \ ,$$

$$\ln \frac{x^*(5)}{x^*(6)} = \ln \frac{x_1^*(5)}{x_1^*(6)} + \tau_1 + 2\tau_2 \quad \text{or } 0.162560 = -0.200671$$
$$-1.248407 + 1.611640 \ ,$$

$$\ln \frac{x^*(7)}{x^*(8)} = \ln \frac{x_1^*(7)}{x_1^*(8)} + \tau_1 + 3\tau_2 \quad \text{or } 0.968380 = -0.200671$$
$$-1.248407 + 2.417460 \ ,$$

or

$$\ln \frac{x^*(3)}{x^*(4)} - \ln \frac{x^*(1)}{x^*(2)} = 0.805820 \ ,$$

$$\ln \frac{x^*(5)}{x^*(6)} - \ln \frac{x^*(3)}{x^*(4)} = 0.805819 \ ,$$

$$\ln \frac{x^*(7)}{x^*(8)} - \ln \frac{x^*(5)}{x^*(6)} = 0.805820 \ .$$

Note that $X^2 = 6.545451 = (9)^2(.080808)$, that is, the quadratic approximation to $2I(x^*:x_1^*)$ also obtainable as

$$2I(x^*:x_1^*) = \Sigma \frac{(x^*(ij)-x_1^*(ij))^2}{x_1^*(ij)} = \frac{10}{(4.5)(5.5)} ((2.599)^2+(1.055)^2 + (.906)^2 + (2.748)^2) =$$
$$0.40404(16.2401) = 6.562$$

Computer Input

```
JOB CARD

EX PROGRAM

TITLE = 'LOGISTIC FIT BERKSON"S MDI'

UNIF = 'O'B

NUMSET = 4

BMAT = '1'B

TOL1 = .001   TOL2 = .001

CNSTRNT = 6   OBS = 8 ;

2  2  2  2

1  1  0  0  0  0  0  0
0  0  1  1  0  0  0  0
0  0  0  0  1  1  0  0
0  0  0  0  0  0  1  1
1  0  1  0  1  0  1  0
0  0  1  0  2  0  3  0

1  9  6  4  3  7  8  2

4.5  5.5  4.5  5.5  4.5  5.5  4.5  5.5
```

Table 5

LOGISTIC FIT BERKSON'S PCT

B MATRIX

	1	2	3	4	5	6	7	8
1	1	1	0	0	0	0	0	0
2	0	0	1	1	0	0	0	0
3	0	0	0	0	1	1	0	0
4	0	0	0	0	0	0	1	1
5	1	0	1	0	1	0	1	0
6	0	0	1	0	2	0	3	0

WEIGHT(1)= 0.250000
WEIGHT(2)= 0.250000
WEIGHT(3)= 0.250000
WEIGHT(4)= 0.250000

INV_WEIGHT(1)= 4.000000
INV_WEIGHT(2)= 4.000000
INV_WEIGHT(3)= 4.000000
INV_WEIGHT(4)= 4.000000

C DESIGN MATRIX

	1	2	3	4	5	6	7	8
1	4	4	0	0	0	0	0	0
2	0	0	4	4	0	0	0	0
3	0	0	0	0	4	4	0	0
4	0	0	0	0	0	0	4	4
5	1	0	1	0	1	0	1	0
6	0	0	1	0	2	0	3	0

OBSERVED VALUES

X(1)= 1.000000	LN_X(1)= 0.000000
X(2)= 9.000000	LN_X(2)= 2.197225
X(3)= 6.000000	LN_X(3)= 1.791759
X(4)= 4.000000	LN_X(4)= 1.386294
X(5)= 3.000000	LN_X(5)= 1.098612
X(6)= 7.000000	LN_X(6)= 1.945910
X(7)= 8.000000	LN_X(7)= 2.079441
X(8)= 2.000000	LN_X(8)= 0.693147

CONSTRAINTS
NTHETA(1)= 40.000000
NTHETA(2)= 40.000000
NTHETA(3)= 40.000000
NTHETA(4)= 40.000000
NTHETA(5)= 18.000000

NTHETA(6)= 56.000000

INITIAL DISTRIBUTION

XSTART(1)= 4.500000	LN_XSTART(1)= 1.504077
XSTART(2)= 5.500000	LN_XSTART(2)= 1.704748
XSTART(3)= 4.500000	LN_XSTART(3)= 1.504077
XSTART(4)= 5.500000	LN_XSTART(4)= 1.704748
XSTART(5)= 4.500000	LN_XSTART(5)= 1.504077
XSTART(6)= 5.500000	LN_XSTART(6)= 1.704748
XSTART(7)= 4.500000	LN_XSTART(7)= 1.504077
XSTART(8)= 5.500000	LN_XSTART(8)= 1.704748

ESTIMATE OF NTHETA AT COUNT= 1

NTHAT(1)= 40.000000
NTHAT(2)= 40.000000
NTHAT(3)= 40.000000
NTHAT(4)= 40.000000
NTHAT(5)= 18.000000
NTHAT(6)= 27.000000

S

	1	2	3	4	5	6
1	160	0	0	0	18	0
2	0	160	0	0	18	18
3	0	0	160	0	18	36
4	0	0	0	160	18	54
5	18	18	18	18	18	27
6	0	18	36	54	27	63

S22.1

	1	2
1	9.900002	14.850004
2	14.850004	34.650009

S22.1_INV

	1	2
1	0.282828	-0.121212
2	-0.121212	0.080808

```
DELTA(1)= 0.000000
DELTA(2)= 0.000000
DELTA(3)= 0.000000
DELTA(4)= 0.000000
DELTA(5)= 0.000000
DELTA(6)= 9.000000


XSQ=     6.545451

ESTIMATE OF X AT COUNT=   1
XSTAR(1)= 2.155855        LN_XSTAR(1)= 0.768168
XSTAR(2)= 7.844142        LN_XSTAR(2)= 2.059767
XSTAR(3)= 3.625511        LN_XSTAR(3)= 1.287955
XSTAR(4)= 6.374481        LN_XSTAR(4)= 1.852303
XSTAR(5)= 5.436513        LN_XSTAR(5)= 1.687634
XSTAR(6)= 4.593488        LN_XSTAR(6)= 1.524639
XSTAR(7)= 7.089394        LN_XSTAR(7)= 1.958655
XSTAR(8)= 2.910608        LN_XSTAR(8)= 1.068361


Z IS OBSERVED TABLE AND X IS INITIAL DIST.

ZI(XSTAR:X)=     5.766655

ZI(Z:XSTAR)=     0.051378

TAU(1)=-1.090908
TAU(2)= 0.727272
```

```
ESTIMATE OF X AT COUNT=   4
XSTAR(1)= 1.901434        LN_XSTAR(1)= 0.642608
XSTAR(2)= 6.098566        LN_XSTAR(2)= 2.091667
XSTAR(3)= 3.445161        LN_XSTAR(3)= 1.236953
XSTAR(4)= 6.554895        LN_XSTAR(4)= 1.880212
XSTAR(5)= 5.405508        LN_XSTAR(5)= 1.687416
XSTAR(6)= 4.594491        LN_XSTAR(6)= 1.524858
XSTAR(7)= 7.247968        LN_XSTAR(7)= 1.980721
XSTAR(8)= 2.752036        LN_XSTAR(8)= 1.012341


Z IS OBSERVED TABLE AND X IS INITIAL DIST.

ZI(XSTAR:X)=     0.875422
```

```
ZI(Z:XSTAR)=     5.985401

TAU(1)=-1.246407
TAU(2)= 0.805820


ESTIMATE OF NTHETA AT COUNT=     4
NTHAT(1)= 39.999969
NTHAT(2)= 40.000000
NTHAT(3)= 40.000000
NTHAT(4)= 39.999985
NTHAT(5)= 16.000031
NTHAT(6)= 36.000000
```

S22.1

	1	2
1	8.276361	13.209361
2	13.209361	30.144561

S22.1_INV

	1	2
1	0.461929	-0.176126
2	-0.176126	0.110352

```
DELTA(1)= 0.000031
DELTA(2)= 0.000000
DELTA(3)= 0.000000
DELTA(4)= 0.000015
DELTA(5)=-0.000031
DELTA(6)= 0.000000

OUTLIER(1)= 2.258585
OUTLIER(2)= 1.555486
OUTLIER(3)= 0.281789
```

```
OUTLIER(4)= 0.165096
OUTLIER(5)= 0.166016
OUTLIER(6)= 0.162892
OUTLIER(7)= 1.309799
OUTLIER(8)= 1.902710


ITERATIONS= 4

TOL1=0.0010    TOL2=0.0010
```

9. Bibliography

The bibliography lists publications, reports, etc., primarily dealing with the analysis of contingency tables. Items are listed by year starting with the most recent. Additional references to related topics may be found in the bibliographies contained in the books by D. R. Cox (1970) and H. O. Lancaster (1969). The bibliography depends in large part on compilations prepared by Dr. Marvin A. Kastenbaum and Dr. H. H. Ku. Permission to use their results is gratefully acknowledged. We make no claim that all items that should have been included are contained herein, and we express our regrets to authors of items so omitted.

1975

KEEGEL, JOHN C. (1975). Several numerical procedures in regression and parameter estimation in contingency tables. Ph.D. dissertation, The Graduate School of Arts and Sciences, The George Washington University, February 1975.

1974

FISHER, MARIAN R. (1974). User's guide to CONTABMOD. (now at National Heart and Lung Institute, Bethesda, Md. 20014)

GAIL, MITCHELL. (1974). Value systems for comparing two independent multinomial trials. Biometrika, 61 (1), 91-100.

IRELAND, C. T. and KULLBACK, S. (1974). The information in contingency tables--an application of information-theoretic concepts to the analysis of contingency tables. Submitted for publication.

KASTENBAUM, M. A. (1974). Analysis of categorical data: some well-known analogues and some new concepts. Communications in Statistics, 3 (5), 401-417.

KU, H. H. and KULLBACK, S. (1974). Loglinear models in contingency table analysis. The American Statistician, Nov. 1974.

KULLBACK, S. and FISHER, MARIAN. (1974). Multivariate logit analysis. Biometrische Zeitschrift, to appear.

KULLBACK, S. and REEVES, P. N. (1974). Analysis of interactions between categorical variables. Biometrische Zeitschrift, No. 8, to appear.

PATIL, KASHINATH D. (1974). Interaction test for three-dimensional contingency tables. Journal Am. Statist. Assn. 69, no. 345, 164-168.

SUGIURA, N. and OTAKE, M. (1974). An extension of Mantel-Haenszel procedure to k 2xc contingency tables and the relation to the logit model. Communications in Statistics, to appear.

1973

GOKHALE, D. V. (1973). Approximating discrete distributions with applications. Journal Am. Statist. Assn. 68, no. 344, 1009-1012.

GOODMAN, L. A. (1973). Guided and unguided methods for selecting models for a set of T multidimensional contingency tables. Journal Am. Statist. Assn. 68, 165-175.

KULLBACK, S. (1973). Estimating and testing interaction parameters in the log-linear model. Biometrische Zeitschrift, 15, 371-388.

KULLBACK, S. and FISHER, MARIAN. (1973). Partitioning second-order interaction in three-way contingency tables. Journal Royal Statist. Soc., Series C (Applied Statistics) 22, 172-184.

MANTEL, NATHAN and BROWN, CHARLES. (1973). A logistic reanalysis of Ashford and Sowden's data on respiratory symptoms in British coal miners. Biometrics, 29 (4), 649-665.

SUGIURA, NARIAKI and OTAKE, MASANORI. (1973). Approximate distribution of the maximum of c-1 χ^2-statistics (2x2) derived from 2xc contingency tables. Communications in Statistics, 1 (1), 9-16.

1972

ADAM, J. and ENKE, H. (1972). Analyse mehrdimensionaler kontingenztafeln mit hilfe des informationsmasses von Kullback. Biometrische Zeitschrift 14, 5, pp. 305-323.

BERKSON, J. (1972). Minimum discrimination information, the "no interaction" problem, and the logistic function. Biometrics 28, 2, pp. 443-468.

BRUNDEN, M. N. (1972). The analysis of non-independent 2x2 tables from 2xc tables using rank sums. Biometrics 28, 2, pp. 603-606.

CAUSEY, B. D. (1972). Sensitivity of raked contingency table totals to changes in problem conditions. Ann. Math. Statist. 43, 2, pp. 656-658.

COX, D. R. (1972). The analysis of multivariate binary data. Appl. Statist. 21, 2, pp. 113-120.

DARROCH, J. N. and RATCLIFF, D. (1972). Generalized iterative scaling for log-linear models. Ann. Math. Statist. 43, 5, pp. 1470-1480.

FIENBERG, S. E. (1972). The analysis of incomplete multiway contingency tables. Biometrics 28, 1, pp. 177-202.

FISHER, MARIAN, R. (1972). An application of minimum discrimination information estimation. Ph.D. dissertation, The Graduate School of Arts and Sciences, The George Washington University, September 1972.

GAIL, M. H. (1972). Mixed quasi-independence models for categorical data. Biometrics 28, 3, pp. 703-712.

GART, J. J. (1972). Interaction tests for 2xsxt contingency tables. Biometrika 59, 2, pp. 309-316.

GOKHALE, D. V. (1972). Analysis of log-linear models. J. Roy. Statist. Soc. Ser. B 34, 3, pp. 371-376.

GOODMAN, L. A. and KRUSKAL, W. H. (1972). Measures of association for cross-classifications, IV: simplification of asymptotic variances. J. Amer. Statist. Assoc. 67, pp. 415-421.

GRIZZLE, J. E. and WILLIAMS, O. D. (1972). Log-linear models and tests of independence for contingency tables. Biometrics 28, 1, pp. 137-156.

GRIZZLE, J. E. and WILLIAMS. O. D. (1972). Contingency tables having ordered response categories. J. Amer. Statist. Assoc. 67, pp. 55-63.

HAMMON, GARY L., JACOBS, STANLEY, E. and REEVES, PHILIP, N. (1972). Surveys of hospital/medical computing. Hospital Financial Management, September 1972.

IRELAND, C. T. (1972). Sequential cell deletion in contingency tables. Statistics Department, The George Washington University.

KOCH, G. G., IMREY, P. B., and REINFURT, D. W. (1972). Linear model analysis of categorical data with incomplete response vectors. Biometrics 28, 3, pp. 663-692.

MARTIN, D. C. and BRADLEY, R. A. (1972). Probability models, estimation, and classification for multivariate dichotomous populations. Biometrics, 28, 203-221.

NATHAN, G. (1972). Asymptotic power of tests for independence in contingency tables from stratified samples. J. Amer. Statist. Assoc. 67, pp. 917-920.

VICTOR, N. (1972). Zur klassifizierung mehrdimensionaler kontingenztafeln. Biometrics 28, 2, pp. 427-442.

1971

ALTHAM, P. M. E. (1971). Exact Bayesian analysis of the intraclass 2x2 table. Biometrika 58, 3, pp. 679-680.

ALTHAM, P. M. E. (1971). The analysis of matched proportions. Biometrika 58, 3, pp. 561-576.

BELLE, G. V. and CORNELL, R. G. (1971). Strengthening tests of symmetry in contingency tables. Biometrics 27, pp. 1074-1078.

BISHOP, Y. M. M. (1971). Effects of collapsing multidimensional contingency tables. Biometrics 27, pp. 545-562.

COHEN, J. E. (1971). Estimation and interaction in a censored 2x2x2 contingency table. Biometrics 27, 379-386.

DEMPSTER, A. P. (1971). An overview of multivariate data analysis. Journal Multivariate Analysis 1, 316-347.

FRYER, J. G. (1971). On the homogeneity of the marginal distributions of a multidimensional contingency table. J. Roy. Statist. Soc. Ser. A 134, pp. 368-371.

GART, J. J. (1971). On the ordering of contingency tables for significance tests. Technometrics 13, pp. 910-911.

GART, J. J. (1971). The comparison of proportions: a review of significance tests, confidence intervals, and adjustments for stratification. Rev. Inst. Internat. Statist. 29, pp. 148-169.

GOKHALE, D. V. (1971). An iterative procedure for analysing log-linear models. Biometrics 27, pp. 681-687.

GOODMAN, L. A. (1971). Partitioning of chi-square, analysis of marginal contingency tables, and estimation of expected frequencies in multidimensional contingency tables. J. Amer. Statist. Assoc. 66, pp. 339-344.

GOODMAN, L. A. (1971). Some multiplicative models for the analysis of cross-classified data. Proc. 6th Berkeley Symp., Berkeley and Los Angeles, University of California Press.

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

GRIZZLE, J. E. (1971). Multivariate logit analysis. Biometrics 27, pp. 1057-1062.

IRELAND, C. T. (1971). A computer program for analyzing contingency tables. (Latest version is CONTAB III) Statistics Department, The George Washington University

JOHNSON, W. D., and KOCH, G. G. (1971). A note on the weighted least squares analysis of the Ries-Smith contingency table data. Technometrics 13, pp. 438-447.

KOCH, G. G., IMREY, P. B., and REINFURT, D. W. (1971). Linear model analysis of categorical data with incomplete response vectors. Institute of Statistics Mimeo Series No. 790, University of North Carolina.

KOCH, G. G., JOHNSON, W. D., and TOLLEY, H. D. (1971). An application of linear models to analyze categorical data pertaining to the relationship between survival and extent of disease. Institute of Statistics Mimeo Series No. 770, University of North Carolina.

KOCH, G. G. and REINFURT, D. W. (1971). The analysis of categorical data from mixed models. Biometrics **27**, pp. 157-173.

KU, H. H. (1971). Analysis of information - an alternative approach to the detection of a correlation between the sexes of adjacent sibs in human families. Biometrics **27**, pp. 175-182.

KU, H. H., VARNER, R., and KULLBACK, S. (1971). On the analysis of multidimensional contingency tables. J. Amer. Statist. Assoc. **66**, pp. 55-64.

KULLBACK, S. (1971). Marginal homogeneity of multidimensional contingency tables. Ann. Math. Statist. **42**. pp. 594-606.

KULLBACK, S. (1971). The homogeneity of the sex ratio of adjacent sibs in human families. Biometrics **27**, pp. 452-457.

NAM, J. (1971). On two tests for comparing matched proportions. Biometrics **27**, pp. 945-959.

PEACOCK, P. B. (1971). The non-comparability of relative risks from different studies. Biometrics **27**, pp. 903-907.

PERITZ, E. (1971). Estimating the ratio of two marginal probabilities in a contingency table. Biometrics **27**, pp. 223-225.

RATCLIFF, D. (1971). Topics on independence and correlation for bounded sum variables. Ph.D. thesis, School of Mathematical Sciences, the Flinders University of South Australia, June 1971.

SIMON, G. A. (1971). Information distances and exponential families, with applications to contingency tables. Technical Report No. 32, November 26, 1971, Department of Statistics, Stanford University.

THOMAS, D. G. (1971). Exact confidence limits for the odds ratio in a 2x2 table. Appl. Statist. **20**, pp. 105-110.

YASAIMAIBODI (YASSAEE), HEDAYAT (1971). On comparison of various estimators and their associated statistics in rxc and rxcx2 contingency tables. Ph.D. dissertation, The George Washington University.

ZELEN, M. (1971). The analysis of several 2x2 contingency tables. Biometrika **58**, pp. 129-137.

1970

ALTHAM, PATRICIA M. E. (1970). The measurement of association of rows and columns for an rxs contingency table. J. Roy. Statist. Soc. Ser. B **32**, pp. 63-73.

ASHFORD, J. R., MORGAN, D. C., RAE, S., SOWDEN, R. R. (1970). Respiratory symptoms in British coal miners. American Review of Respiratory Disease, 102, 370-381.

ASHFORD, J. R. and SOWDEN, R. D. (1970). Multivariate probit analysis. Biometrics, 26, 535-546.

BHAPKAR, V. P. (1970). Categorical data analysis of some multivariate tests. Essays in Probability and Statistics (R. C. Bose et al., eds.). The University of North Carolina Press, pp. 85-110.

CAMPBELL, L. L. (1970). Equivalence of Gauss's principle and minimum discrimination information estimation of probabilities. Ann. Math. Statist. **41**, pp. 1011-1015.

COX, D. R. (1970). The Analysis of Binary Data. Methuen & Co., Ltd., London.

CRADDOCK, J. M. and FLOOD, C. R. (1970). The distribution of the chi-square statistic in small contingency tables. Appl. Statist. **19**, pp. 173-181.

FIENBERG, S. E. (1970). Quasi-independence and maximum likelihood estimation in incomplete contingency tables. J. Amer. Statist. Assoc. **65**, 332, pp. 1610-1616.

FIENBERG, S. E. (1970). An iterative procedure for estimation in contingency tables. Ann. Math. Statist. **41**, pp. 907-917.

FIENBERG, S. E. (1970). The analysis of multidimensional contingency tables. Ecology **51**, 2, pp. 419-433.

FIENBERG, S. E. and GILBERT, J. P. (1970). Geometry of a two by two contingency table. J. Amer. Statist. Assoc. **65**, pp. 694-701.

FIENBERG, S. E. and HOLLAND, P. W. (1970). Methods for eliminating zero counts in contingency tables. Random Counts in Scientific Work (G. P. Patil, ed.). The Pennsylvania State University Press.

GOOD, I. J., GOVER, T. N., and MITCHELL, G. J. (1970). Exact distributions for χ-squared and for the likelihood-ratio statistic for the equiprobable multinomial distribution. J. Amer. Statist. Assoc. **65**, pp. 267-283.

GOODMAN, L. A. (1970). The multivariate analysis of qualitative data: interaction among multiple classifications. J. Amer. Statist. Assoc. **65**, pp. 226-256.

KASTENBAUM, M. A. (1970). A review of contingency tables. Essays in Probability and Statistics (R. C. Bose et al., eds.). The University of North Carolina Press, pp. 407-438.

KULLBACK, S. (1970). Various applications of minimum discrimination information estimation, particularly to problems of contingency table analysis. Proceedings of the Meeting on Information Measures, University of Waterloo, Ontario, Canada, April 10-14, 1970, I-33-I-66.

KULLBACK, S. (1970). Minimum discrimination information estimation and application. Proceedings of the Sixteenth Conference on the Design of Experiments in Army Research, Development and Testing, 21 October 1970. ARO-D Report 71-3, 1-38 Proceedings of the Conference.

MANTEL, N. (1970). Incomplete contingency tables. Biometrics **26**, pp. 291-304.

MOLK, YEHUDA (1970). On estimation of probabilities in contingency tables with restrictions on marginals. Ph.D. dissertation, The George Washington University, February 1970.

ODOROFF, C. L. (1970). Minimum logit chi-square estimation and maximum likelihood estimation in contingency tables. J. Amer. Statist. Assoc. **65**, 332, pp. 1617-1631.

WAGNER, S. S. (1970). The maximum-likelihood estimate for contingency tables with zero diagonal. J. Amer. Statist. Assoc. **65**, 331, pp. 1362-1383.

1969

ALTHAM, PATRICIA M. E. (1969). Exact Bayesian analysis of a 2x2 contingency table and Fisher's "exact" significance test. J. Roy. Statist. Soc. Ser. B. **31**, pp. 261-269.

ARGENTIERO, P. D. (1969). χ-squared statistic for goodness of fit test, its derivation and tables. NASA Technical Report, TR-R-313.

BISHOP, Y. M. M. (1969). Full contingency tables, logits, and split contingency tables. Biometrics **25**, pp. 383-400.

BISHOP, Y. M. M. and FIENBERG, S. E. (1969). Incomplete two-dimensional contingency tables. Biometrics **25**, pp. 119-128.

DEMPSTER, A. P. (1969). Some theory related to fitting exponential models. Research Report S-4, Department of Statistics, Harvard University.

FIENBERG, S. E. (1969). Preliminary graphical analysis and quasi-independence for two-way contingency tables. Appl. Statist. **18**, pp. 153-168.

GOODMAN, L. A. (1969). On partition χ-squared and detecting partial association in the three-way contingency tables. J. Roy. Statist. Soc. Ser. B. **31**, pp. 486-498.

GRIZZLE, J. E., STARMER, C. F., and KOCH, G. G. (1969). Analysis of categorical data by linear models. Biometrics 25, pp. 489-504.

HEALY, M. J. R. (1969). Exact tests of significance in contingency tables. Technometrics 11, pp. 393-395.

IRELAND, C. T., KU, H. H., and KULLBACK, S. (1969). Symmetry and marginal homogeneity of an rxr contingency table. J. Amer. Statist. Assoc. 64, pp. 1323-1341.

KOCH, G. G. (1969). The effect of non-sampling errors on measures of association in 2x2 contingency tables. J. Amer. Statist. Assoc. 64, pp. 852-863.

KU, H. H. and KULLBACK, S. (1969). Analysis of multidimensional contingency tables: an information theoretical approach. Contributed papers, 37th Session of the International Statistical Institute, pp. 156-158.

KU, H. H. and KULLBACK, S. (1969). Approximating discrete probability distributions. IEEE Trans. on Information Theory, IT-15, 444-447.

LANCASTER, H. O. (1969). Contingency tables of higher dimensions. Bulletin of the International Statistical Institute. 43, 1, pp. 143-151.

LANCASTER, H. O. (1969). The Chi-Squared Distribution. Wiley, New York.

NAGNUR, B. N. (1969). LAMST and the hypotheses of no three factor interaction in contingency tables. J. Amer. Statist. Assoc. 64, pp. 207-215.

PLACKETT, R. L. (1969). Multidimensional contingency tables. A Survey of Models and Methods, Bulletin of the International Statistical Institute. 43, 1, pp. 133-142.

1968

BENNETT, B. M. (1968). Notes on χ-squared tests for matched samples. J. Roy.Statist. Soc. Ser. B 30, pp. 368-370.

BERKSON, J. (1968). Application of minimum logit chi-squared estimate to a problem of Grizzle with a notation on the problem of no interaction. Biometrics 24, pp. 75-96.

BHAPKAR, V. P. (1968). On the analysis of contingency tables with a quantitative response. Biometrics 24, pp. 329-338.

BHAPKAR, V. P. and KOCH, G. G. (1968). Hypotheses of "no interaction" in multidimensional contingency tables. Technometrics 10, pp. 107-123.

BHAPKAR, V. P. and KOCH, G. G. (1968). On the hypotheses of "no interaction" in contingency tables. Biometrics 24, pp. 567-594.

FIENBERG, S. E. (1968). The geometry of an rxc contingency table. Ann. Math. Statist. 39, pp. 1186-1190.

GOODMAN, L. A. (1968). The analysis of cross-classified data: independence, quasi-independence, and interactions in contingency tables with or without missing entries. J. Amer. Statist. Assoc. 63, pp. 1091-1131.

HAMDAN, M. A. (1968). Optimum choice of classes for contingency tables. J. Amer. Statist. Assoc. 63, pp. 291-297.

IRELAND, C. T. and KULLBACK, S. (1968). Contingency tables with given marginals. Biometrika 55, pp. 179-188.

IRELAND, C. T. and KULLBACK, S. (1968). Minimum discrimination information estimation. Biometrics 24, pp. 707-713.

KU, H. H. and KULLBACK, S. (1968). Interaction in multidimensional contingency tables: an information theoretic approach. J. Res. Nat. Bur. Standards Sect. B 72, pp. 159-199.

KU, H. H., VARNER, R., and KULLBACK, S. (1968). Analysis of multidimensional contingency tables. Proceedings of the Fourteenth Conference on the Design of Experiments in Army Research, Development and Testing. ARO-D Report 69-2.

KULLBACK, S. (1968). Information Theory and Statistics. Dover Pub., Inc., New York.

KULLBACK, S. (1968). Probability densities with given marginals. Ann. Math. Statist. 39, pp. 1236-1243.

LEYTON, M. K. (1968). Rapid calculation of exact probabilities for 2x3 contingency tables. Biometrics 24, pp. 714-717.

MATHIEU, JEAN-RENE and LAMBERT, E. (1968). Un test de l'identite des marges dun tableau de correlation. C. R. Acad. Sci. Paris 267, pp. 832-834.

MOSTELLER, F. (1968). Association and estimation in contingency tables. J. Amer. Statist. Assoc. 63, pp. 1-28.

SLAKTER, M. J. (1968). Accuracy of an approximation to the power of the chi-square goodness of fit test with small but equal expected frequencies. J. Amer. Statist. Assoc. 63, pp. 912-924.

SUGIURA, N. and OTAKE, M. (1968). Numerical comparison of improved methods of testing in contingency tables with small frequencies. Ann. Inst. Statist. Math. 20, pp. 507-517.

1967

BENNETT, B. M. (1967). Tests of hypothesis concerning matched samples. J. Roy. Statist. Soc. Ser. B 29, pp. 468-474.

BISHOP, Y. M. M. (1967). Multidimensional contingency tables: cell estimates. Ph.D. dissertation, Harvard University.

BLOCH, D. A. and WATSON, G. S. (1967). A Bayesian study of the multinomial distribution. Ann. Math. Statist. 38, pp. 1423-1435.

COX, D. R. and LAUH, E. (1967). A note on the graphical analysis of multidimensional contingency tables. Technometrics 9, pp. 481-488.

GOOD, I. J. (1967). A Bayesian significance test for multinomial distributions. J. Roy. Statist. Soc. Ser. B 29, pp. 339-431.

FERGUSON, T. S. (1967). Mathematical Statistics. Academic Press Inc. N. Y.

SNEDECOR, G. W., and COCHRAN, W. G. (1967). Statistical Methods. The Iowa State University Press, Ames, Iowa.

1966

ARMITAGE, P. (1966). The chi-square test for heterogeneity of proportions after adjustment for stratification. J. Roy. Statist. Soc. Ser. B 28, pp. 150-163.

BHAPKAR, V. P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data. J. Amer. Statist. Assoc. 61, pp. 228-235.

BHAPKAR, V. P. (1966). Notes on analysis of categorical data. Institute of Statistics Mimeo Series No. 477, University of North Carolina.

BHAT, B. R. and KULKARNI, S. R. (1966). LAMP test of linear and loglinear hypotheses in multinomial experiments. J. Amer. Statist. Assoc. 61, pp. 236-245.

COX, D. R. (1966). A simple example of a comparison involving quantal data. Biometrika 53, pp. 215-220.

CRADDOCK, J. M. (1966). Testing the significance of a 3x3 contingency table. The Statistician 16, pp. 87-94.

GABRIEL, K. R. (1966). Simultaneous test procedures for multiple comparison on categorical data. J. Amer. Statist. Assoc. 61, pp. 1081-1096.

GART, J. J. (1966). Alternative analyses of contingency tables. J. Roy. Statist. Soc. Ser. B 28, pp. 164-179.

GOOD, I. J. (1966). How to estimate probabilities. J. Inst. Math. Appl. 2, pp. 364-383.

KULLBACK, S. and KHAIRAT, M. A. (1966). A note on minimum discrimination information. Ann. Math. Statist. 37, pp. 279-280.

MANTEL, N. (1966). Models for complex contingency tables and polychotomous dosage response curves. Biometrics 22, pp. 83-95.

1965

ASANO, C. (1965). On estimating multinomial probabilities by pooling incomplete samples. Ann. Inst. Statist. Math. 17, pp. 1-14.

BHAPKAR, V. P. and KOCH, G. G. (1965). On the hypothesis of "no interaction" in three-dimensional contingency tables. Institute of Statistics Mimeo Series No. 440, University of North Carolina.

BHAPKAR, V. P. and KOCH, G. G. (1965). Hypothesis of no interaction in four-dimensional contingency tables. Institute of Statistics Mimeo Series No. 449, University of North Carolina.

BHAT, B. R. and NAGNUR, B. N. (1965). Locally asymptotically most stringent tests and Lagrangian multiplier tests of linear hypotheses. Biometrika 52, 3 and 4, pp. 459-468.

BIRCH, M. W. (1965). The detection of partial association II: the general case. J. Roy. Statist. Soc. Ser. B 27, pp. 111-124.

CAUSSINUS, H. (1965). Contribution a l'analyse statistique des tableaux de correlation. Ann. Fac. Sci. Univ. Toulouse 29, pp. 77-182.

GOOD, I. J. (1965). The Estimation of Probabilities: An Essay on Modern Bayesian Methods. Research Monograph, 30. The MIT Press, Cambridge, Massachusetts.

KASTENBAUM, M. A. (1965). Contingency tables: a review. MRC Technical Summary Report No. 596. Mathematical Research Center, The University of Wisconsin.

KATTI, S. K. and SASTRY, A. N. (1965). Biological examples of small expected frequencies and the chi-square test. Biometrics 21, pp. 49-54.

LANCASTER, H. O. and BROWN, T. A. I. (1965). Size of χ-squared test in the symmetrical multinomials. Austral. J. Statist. 7, p. 40.

LEWONTIN, R. C. and FELSENSTEIN (1965). The robustness of homogeneity tests in 2xn tables. Biometrics 21, pp. 19-33.

MOTE, V. L. and ANDERSON, R. L. (1965). An investigation of the effect of misclassification on the properties of chi-squared tests in the analysis of categorical data. Biometrika 52, pp. 95-109.

RADHAKRISHNA, S. (1965). Combination of results from several 2x2 contingency tables. Biometrics 21, pp. 86-98.

1964

ALLISON, H. E. (1964). Computational forms for chi-square. Amer. Statist. 18, 1, pp. 17-18.

BENNETT, B. M. and NAKAMURA, E. (1964). Tables for testing significance in a 2x3 contingency table. Tecnnometrics 6, 4, pp. 439-458.

BIRCH, M. W. (1964). The detection of partial association I: the 2x2 case. Jour. Roy. Statist. Soc. Series B, 26, 313-324.

BROSS, I. D. J. (1964). Taking a covariable into account. J. Amer. Statist. Assoc. 59, 307, pp. 725-736.

CHEW, V. (1964). Application of the negative binomial distribution with probability of misclassification. Virginia Journal of Science 15, 1, pp. 34-40.

GOODMAN, L. A. (1964). Simultaneous confidence limits for cross-product ratios in contingency tables. J. Roy. Statist. Soc. Ser. B 26, 1, pp. 86-102.

GOODMAN, L. A. (1964). Simple methods for analyzing three-factor interaction in contingency tables. J. Roy. Statist. Soc. 59, pp. 319-352.

GOODMAN, L. A. (1964). Interactions in multidimensional contingency tables. Ann. Math. Statist. 35, 2, pp. 632-646.

GOODMAN, L. A. (1964). Simultaneous confidence intervals for contrasts among multinomial populations. Ann. Math. Statist. 35, 2, pp. 716-725.

HARKNESS, W. L. and KATZ, L. (1964). Comparison of the power functions for the test of independence in 2x2 contingency tables. *Ann. Math. Statist.* **35**, 3, pp. 1115-1127.

KIHLBERG, J. K., NARRAGON, E. A., and CAMPBELL, B. J. (1964). Automobile crash injury in relation to car size. *Cornell Aeronautical Lab. Inc. Report*, VJ-1823R11.

LINDLEY, D. V. (1964). The Bayesian analysis of contingency tables. *Ann. Math. Statist.* **35**, 4, pp. 1622-1643.

PLACKETT, R. L. (1964). The continuity correction in 2x2 tables. *Biometrika* **21**, Parts 3 and 4, pp. 327-338.

PUTTER, J. (1964). The χ^2 goodness-of-fit test for a class of cases of dependent observations. *Biometrika* **51**, pp. 250-252.

SOMERS, R. H. (1964). Simple measures of association for the triple dichotomy. *J. Roy. Statist. Soc. Ser. A* **127**, 3, pp. 409-415.

TALLIS, G. M. (1964). The use of models in the analysis of some classes of contingency tables. *Biometrics* **24**, 4, pp. 832-839.

1963

BENNETT, B. M. and NAKAMURA, E. (1963). Tables for testing significance in a 2x3 contingency table. *Technometrics* **5**, 4, pp. 501-511.

BIRCH, M. W. (1963). Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc. Ser. B* **25**, 1, pp. 220-233.

DARROCH, J. N. and SILVEY, S. D. (1963). On testing more than one hypothesis. *Ann. Math. Statist.* **34**, 2, pp. 555-567.

DIAMOND, E. L. (1963). The limiting power of categorical data chi-square tests analogous to normal analysis of variance. *Ann. Math. Statist.* **34**, 4, pp. 1432-1441.

EDWARDS, A. W. F. (1963). The measure of association in a 2x2 table. *J. Roy. Statist. Soc. Ser. A* **126**, 1, pp. 109-114.

FELDMAN, S. E. and KLINGER, E. (1963). Short cut calculation of the Fisher-Yates exact test. *Psychometrika* **28**, 3, pp. 289-291.

FINNEY, D. J., LATSCHA, R., BENNETT, B. M., HSU, P. and PEARSON, E. S. (1963). Tables for testing significance in a 2x2 contingency table. (Supplement by B. M. Bennett and C. Horst i+28). Camb. Univ. Press. 103 pp.

GOLD, R. A. (1963). Tests auxiliary to χ^2 tests in a Markov chain. *Ann. Math. Statist.* **34**, 1, pp. 56-74.

GOOD, I. J. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *Ann. Math. Statist.* **34**, 3, pp. 911-934.

GOODMAN, L. A. (1963). On methods for comparing contingency tables. *J. Roy. Statist. Soc. Ser. A* **126**, 1, pp. 94-108.

GOODMAN, L. A. (1963). On Plackett's test for contingency table interactions. *J. Roy. Statist. Soc. Ser. B* **25**, 1, pp. 179-188.

GOODMAN, L. A. and KRUSKAL, W. H. (1963). Measures of association for cross classification III: approximate sampling theory. *J. Amer. Statist. Assoc.* **58**, pp. 310-364.

KU, H. H. (1963). A note on contingency tables involving zero frequencies and the 2I test. *Technometrics* **5**, 3, pp. 398-400.

MANTEL, N. (1963). Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. *J. Amer. Statist. Assoc.* **58**, pp. 690-700.

NEWELL, D. J. (1963). Misclassification in 2x2 tables. *Biometrics* **19**, 1, pp. 187-188.

OKAMATO, M. (1963). Chi-square statistic based on the pooled frequencies of several observations. *Biometrika* **50**, pp. 524-528.

RIES, P. N. and SMITH, H. (1963). The use of chi-square for preference testing in multidimensional problems. Chem. Eng. Prog. Symposium Series 59, 42, pp. 39-43.

WALSH, J. E. (1963). Loss in test efficiency due to misclassification for 2x2 tables. Biometrics 19, 1, pp. 158-162.

CORNFIELD, J. (1962). Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: a discriminant function analysis. Federation Proceeding, No. 4, Part II, July-Aug. 1962, Supplement No. 11, 58-61.

1962

DALY, C. (1962). A simple test for trends in a contingency table. Biometrics 18, 1, pp. 114-119.

DARROCH, J. N. (1962). Interactions in multi-factor contingency tables. J. Roy. Statist. Soc. Ser. B 24, 1, pp. 251-263.

FISHER, SIR RONALD A. (1962). Confidence limits for a cross-product ratio. Austral. J. Statist. 4, 1, p. 41.

GART, J. J. (1962). Approximate confidence limits for relative risks. J. Roy. Statist. Soc. Ser. B 24, 2, pp. 454-463.

GART, J. J. (1962). On the combination of relative risks. Biometrics 18, 4, pp. 601-610.

KINCAID, W. M. (1962). The combination of 2xm contingency tables. Biometrics 18, 2, pp. 224-228.

KULLBACK, S. KUPPERMAN, M. and KU, H. H. (1962). An application of information theory to the analysis of contingency tables with a table of 2N ln N, N = 1(1)10,000. J. Res. Nat. Bur. Standards Sect. B 66, pp. 217-243.

KULLBACK, S., KUPPERMAN, M., and KU, H. H. (1962). Tests for contingency tables and Markov chains. Technometrics 4, 4, pp. 573-608.

LEWIS, B. N. (1962). On the analysis of interaction in multi-dimensional contingency tables. *J. Roy. Statist. Soc. Ser. A* **125**, 1, pp. 88-117.

PLACKETT, R. L. (1962). A note on interactions in contingency tables. *J. Roy. Statist. Soc. Ser. B* **24**, 1, pp. 162-166.

TALLIS, G. M. (1962). The maximum likelihood estimation of correlation from contingency tables. *Biometrics* **18**, 3, pp. 342-353.

1961

BERGER, A. (1961). On comparing intensities of association between two binary characteristics in two different populations. *J. Amer. Statist. Assoc.* **56**, pp. 889-908.

BHAPKAR, V. P. (1961). Some tests for categorical data. *Ann. Math. Statist.* **32**, 1, pp. 72-83.

BILLINGSLEY, P. (1961). *Statistical Inference for Markov Processes*. Statistical Research Monographs, **2**, The University of Chicago Press.

CLARINGBOLD, P. J. (1961). The use of orthogonal polynomials in the partition of chi-square. *Austral. J. Statist.* **3**, 2, pp. 48-63.

FRIEDLANDER, D. (1961). A technique for estimating a contingency table, given the marginal totals and some supplementary data. *J. Roy. Statist. Soc. Ser. A* **124**, 3, pp. 412-420.

GARSIDE, R. F. (1961). Tables for ascertaining whether differences between percentages are statistically significant at the 1% level. *British Med. J.* **1**, 874-876.

GREGORY, G. (1961). Contingency tables with a dependent classification. *Austral. J. Statist.* **3**, 2, pp. 42-47.

GRIZZLE, J. E. (1961). A new method of testing hypotheses and esti ting parameters for the logistic model. *Biometrics* **17**, 3, pp. 372-385.

KENDALL, M. G. and STUART, A. (1961). *The Advanced Theory of Statistics*. **2**, Charles Griffin and Company, London.

OKAMATO, M. and ISHII, G. (1961). Test of independence in intraclass 2x2 tables. Biometrika 48, pp. 181-190.

ROGOT, E. (1961). A note on measurement errors and detecting real differences. J. Amer. Statist. Assoc. 56, pp. 314-319.

SCHULL, W. J. (1961). Some problems of analysis of multi-factor tables. Bull. Inst. Internat. Statist. 28, Part 3, pp. 259-270.

YATES, F. (1961). Marginal percentages in multiway tables of quantal data with disproportionate frequencies. Biometrics 17, 1, pp. 1-9.

1960

BENNETT, B. M. and HSU, P. (1960). On the power function of the exact test for the 2x2 contingency table. Biometrika 47, pp. 393-398.

GRIDGEMAN, N. T. (1960). Card-matching experiments: a conspectus of theory. J. Roy. Statist. Soc. Ser. A 123, 1, pp. 45-49.

ISHII, G. (1960). Intraclass contingency tables. Ann. Inst. Statist. Math. 12, pp. 161-207; corrections, p. 279.

KASTENBAUM, M. A. (1960). A note on the additive partitioning of chi-square in contingency tables. Biometrics 16, 3, pp. 416-422.

KUPPERMAN, M. (1960). On comparing two observed frequency counts. Appl. Statist. 9, 1, pp. 37-42.

LANCASTER, H. O. (1960). On tests on independence in several dimensions. J. Austral. Math. Soc. 1, pp. 241-254.

ROBERTSON, W. H. (1960). Programming Fisher's exact method of comparing two percentages. Technometrics 2, 1, pp. 103-107.

SOLOMON, H. (1960). Classification procedures based on dichotomous response vectors, no. 36 in Contributions to Probability and Statistics, Essays in Honor of Harold Hotelling, Edited by I. Olkin et al. Stanford U. P., Stanford, Cal. 1960, pp. 414-423. (Also in Studies in Item Analysis and Prediction, Edited by H. Solomon, Stanford U. P., Stanford, Cal. 1961, pp. 177-186.)

1959

ANDERSON, R. L. (1959). Use of contingency tables in the analysis of consumer preference studies. Biometrics 15, 4, pp. 582-590.

CHAKRAVARTI, I. M. and RAO, C. R. (1959). Tables for some small sample tests of significance for Poisson distributions and 2x3 contingency tables. Sankhya 21, Parts 3 and 4, pp. 315-326.

GOODMAN, L. A. and KRUSKAL, W. H. (1959). Measures of association for cross classification II: further discussion and references. J. Amer. Statist. Assoc. 54, pp. 123-163.

HALDANE, J. B. S. (1959). The analysis of heterogeneity, I. Sankhya 21, Parts 3 and 4, pp. 209-216.

HOYT, C. J., KRISHNAIAH, P. R., and TORRANCE, E. P. (1959). Analysis of complex contingency data. Journal of Experimental Education 27, pp. 187-194.

KASTENBAUM, M. A. and LAMPHIEAR, D. E. (1959). Calculation of chi-square to test the no three-factor interaction hypothesis. Biometrics 15, 1, pp. 107-115.

KULLBACK, S. (1959). Information Theory and Statistics. John Wiley and Sons, New York.

KUPPERMAN, M. (1959). A rapid significance test for contingency tables. Biometrics 15, 4, pp. 625-628.

NASS, C. A. G. (1959). The χ^2 test for small expectations in contingency tables, with special reference to accidents and absenteeism. Biometrika 46, pp. 365-385.

SILVEY, S. D. (1959). The Lagrangian multiplier test. Ann. Math. Statist. 30, 2, pp. 389-407.

SOMERS, R. H. (1959). The rank analogue of product-moment partial correlation and regression, with application to manifold, ordered contingency tables. Biometrika 46, pp. 241-246.

STEYN, H. S. (1959). On χ^2-tests for contingency tables of negative binomial type. Statistica Neerlandica **13**, pp. 433-444.

WEINER, I. B. (1959). A note of the use of Mood's likelihood ratio test for item analyses involving 2x2 tables with small samples. Psychometrika **24**, 4, pp. 371-372.

1958

BLALOCK. H. M., Jr. (1958). Probabilistic interpretations for the mean square contingency. J. Amer. Statist. Assoc. **53**, pp. 102-105.

GARSIDE, R. F. (1958). Tables for ascertaining whether differences between percentages are statistically significant. British Med. J. 1, 1459-1461.

KASTENBAUM, M. A. (1958). Estimation of relative frequencies of four sperm types in Drosophila melanogaster. Biometrics **14**, 2, pp. 223-228.

MITRA, S. K. (1958). On the limiting power function of the frequency chi-square test. Ann. Math. Statist. **29**, pp. 1221-1233.

SNEDECOR, G. W. (1958). Chi-square of Bartlett, Mood and Lancaster in a 2^3 contingency table. Biometrics **14**, 4, pp. 560-562 (Query).

1957

BROSS, I. D. J. and KASTEN, E. L. (1957). Rapid analysis of 2x2 tables. J. Amer. Statist. Assoc. **52**, pp. 18-28.

CORSTEN, L. C. A. (1957). Partition of experimental vectors connected with multinomial distributions. Biometrics **13**, 4, pp. 451-484.

EDWARDS, J. H. (1957). A note on the practical interpretation of 2x2 tables. Brit. J. Prev. Soc. Med. **11**, pp. 73-78.

LANCASTER, H. O. (1957). Some properties of the bivariate normal distribution considered in the form of a contingency table. Biometrika **44**, pp. 289-292.

MOTE, V. L. (1957). An investigation of the effect of misclassification of the chi-square tests in the analysis of categorical data. Unpublished Ph.D. dissertation, North Carolina State College, Raleigh, North Carolina (also Institute of Statistics Mimeo Series No. 182).

ROY, S. N. (1957). <u>Some Aspects of Multivariate Analysis</u>. John Wiley and Sons, New York.

SAKODA, J. M. and COHEN, B. H. (1957). Exact probabilities for contingency tables using binomial coefficients. <u>Psychometrika</u> **22**, 1, pp. 83-86.

WOOLF, B. (1957). The log likelihood ratio test (the G-test). Methods and tables for tests of heterogeneity in contingency tables. <u>Annals of Human Genetics</u> **21**, pp. 397-409.

<u>1956</u>

FISHMAN, J. A. (1956). A note on Jenkins' "Improved Method for Tetrachoric r." <u>Psychometrika</u> **20**, 3, pp. 305.

GOOD, I. J. (1956). On the estimation of small frequencies in contingency tables. <u>J. Roy. Statist. Soc. Ser. B</u> **18**, 1, pp. 113-124.

GRIDGEMAN, N. T. (1956). A tasting experiment. <u>Appl. Statist.</u> **5**, 2, pp. 106-112.

LEANDER, E. K. and FINNEY, D. J. (1956). An extension of the use of the χ^2 test. <u>Appl. Statist.</u> **5**, 2, pp. 132-136.

MAINLAND, D., HERRERA, L. and SUTCLIFFE, M. I. (1956). Statistical tables for use with binomial samples - contingency tests, confidence limits, and sample size estimates. New York University College of Medicine, New York.

ROY, S. N. and KASTENBAUM, M. A. (1956). On the hypothesis of no "interaction" in a multiway contingency table. <u>Ann. Math. Statist.</u> **27**, 3, pp. 749-757.

ROY, S. N. and MITRA, S. K. (1956). An introduction to some non-parametric generalizations of analysis of variance and multivariate analysis. Biometrika 43, Parts 3 and 4, pp. 361-376.

WATSON, G. S. (1956). Missing and "mixed-up" frequencies in contingency tables. Biometrics 12, 1, pp. 47-50.

1955

ARMITAGE, P. (1955). Tests for linear trends in proportions and frequencies. Biometrics 11, 3, pp. 375-386.

ARMSEN, P. (1955). Tables for significance tests of 2x2 contingency tables. Biometrika 42, pp. 494-505.

COCHRAN, W. G. (1955). A test of a linear function of the deviations between observed and expected numbers. J. Amer. Statist. Assoc. 50, pp. 377-397.

HALDANE, J. B. S. (1955). Substitutes for χ^2. Biometrika 42, pp. 265-266.

HALDANE, J. B. S. (1955). A problem in the significance of small numbers. Biometrika 42, pp. 266-267.

HALDANE, J. B. S. (1955). The rapid calculation of χ^2 as a test of homogeneity from a 2xn table. Biometrika 42, pp. 519-520.

JENKINS, W. L. (1955). An improved method for tetrachoric r. Psychometrika 20, 3, pp. 253-258.

KASTENBAUM, M. A. (1955). Analysis of data in multiway contingency tables. Unpublished doctoral dissertation. North Carolina State College, October 1955.

LESLIE, P. H. (1955). A simple method of calculating the exact probability in 2x2 contingency tables with small marginal totals. Biometrika 42, pp. 522-523.

MITRA, S. K. (1955). Contributions to the statistical analysis of categorical data. North Carolina Institute of Statistics Mimeograph Series No. 142, December 1955.

ROY, S. N. and KASTENBAUM, M. A. (1955). A generalization of analysis of variance and multivariate analysis to data based on frequencies in qualitative categorical or class intervals. North Carolina Institute of Statistics Mimeograph Series No. 131, June 1955.

ROY, S. N. and MITRA, S. K. (1955). An introduction to some non-parametric generalizations of analysis of variance and multivariate analysis. North Carolina Institute of Statistics Mimeograph Series No. 139, November 1955.

SEKAR, C. C., AGARIVALA, S. P. and CHAKRABORTY, P. N. (1955). On the power function of a test of significance for the difference between two proportions. Sankhya **15**, Part 4, pp. 381-390.

STUART, A. (1955). A test of homogeneity of the marginal distributions in a two-way classification. Biometrika **42**, pp. 412-416.

WOOLF, B. (1955). On estimating the relation between blood group and disease. Annals of Human Genetics **19**, pp. 251-253.

YATES, F. (1955). A note on the application of the combination of probabilities test to a set of 2x2 tables. Biometrika **42**, pp. 401-411.

YATES, F. (1955). The use of transformations and maximum likelihood in the analysis of quantal experiments involving two treatments. Biometrika **42**, pp. 382-403.

1954

BROSS, I. D. J. (1954). Misclassification in 2x2 tables. Biometrics **10**, 4, pp. 478-486.

COCHRAN, W. G. (1954). Some methods for strengthening the common chi-square tests. Biometrics **10**, 4, pp. 417-451.

DAWSON, R. B. (1954). A simplified expression for the variance of the χ^2 function on a contingency table. Biometrika **41**, p. 280.

GOODMAN, L. A. and KRUSKAL, W. H. (1954). Measures of association for cross classification. J. Amer. Statist. Assoc. **49**, pp. 732-764.

KIMBALL, A. W. (1954). Short-cut formulas for the exact partition of chi-square in contingency tables. Biometrics 10, 4, pp. 452-458.

MCGILL, W. J. (1954). Multivariate information transmission. Psychometrika 19, 2, pp. 97-116.

1952

COCHRAN, W. G. (1952). The χ^2 test of goodness of fit. Ann. Math. Statist. 23, 3, pp. 315-345.

DYKE, G. V. and PATTERSON, H. D. (1952). Analysis of factorial arrangements when the data are proportions. Biometrics 8, pp. 1-12.

1951

FREEMAN, G. H., and HALTON, J. H. (1951). Note on the exact treatment of contingency, goodness of fit and other problems of significance. Biometrika, 38, 141-149.

LANCASTER, H. O. (1951). Complex contingency tables treated by the partition of chi-square. J. Roy. Statist. Soc. Ser. B 13, pp. 242-249.

SIMPSON, C. H. (1951). The interpretation of interaction in contingency tables. J. Roy. Statist. Soc. Ser. B 13, pp. 238-241.

1950

TOCHER, K. D. (1950). Extension of the Neyman-Pearson theory of tests to discontinuous variates. Biometrika 37, pp. 130-144.

1949

HSU, P. L. (1949). The limiting distributions of functions of sample means and application to testing hypotheses. Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability (1945, 1946), University of California Press, Berkeley and Los Angeles.

IRWIN, J. O. (1949). A note on the subdivision of chi-square into components. Biometrika 36, pp. 130-134.

LANCASTER, H. O. (1949). The derivation and partition of chi-square in certain discrete distributions. Biometrika 36, pp. 117-129.

LANCASTER, H. O. (1949). The combination of probabilities arising from data in discrete distributions. Biometrika, 36, 370-382. Corrig. 37, 452.

1948

FINNEY, D. J. (1948). The Fisher-Yates test of significance in 2x2 contingency tables. Biometrika, 35, 145-156.

SWINEFORD, F. (1948). A table for estimating the significance of the difference between correlated percentages. Psychometrika, 13, 23-25.

YATES, F. (1948). The analysis of contingency tables with groupings based on quantitative characters. Biometrika 35, pp. 176-181.

1947

BARNARD, G. A. (1947). Significance tests for 2x2 tables. Biometrika 34, pp. 123-138.

BARNARD, G. A. (1947). 2x2 tables. A note on E. S. Pearson's paper. Biometrika, 34, 168-169.

PEARSON, E. S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2x2 table. Biometrika 34, pp. 139-167.

1946

CRAMER, H. (1946). Mathematical Methods of Statistics. Princeton University Press, p. 424.

1945

NORTON, H. W. (1945). Calculation of chi-square for complex contingency tables. J. Amer. Statist. Assoc. 40, pp. 251-258.

1943

WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions Amer. Math. Soc. 54, 426-482.

1939

HALDANE, J. B. S. (1939). Note on the preceding analysis of Mendelian segregations. Biometrika, 31, 67-71.

ROBERTS, E., DAWSON, W. M., and MADDEN, M. (1939). Observed and theoretical ratios in Mendelian inheritance. Biometrika, 31, 56-66.

1938

FISHER, R. A., and YATES, F. (1938). Statistical Tables for Biological, Agricultural and Medical Research. Oliver and Boyd, Edinburgh. 6th Ed. (1963) x+146.

SWAROOP, S. (1938). Tables of the exact values of probabilities for testing the significance of differences between proportions based on pairs of small samples. Sankhya, 4, 73-84.

1937

HALDANE, J. B. S. (1937). The exact value of the moments of the distribution of χ^2 used as a test of goodness of fit, when expectations are small. Biometrika 29, pp. 133-143.

1935

BARTLETT, M. S. (1935). Contingency table interactions. J. Roy Statist. Soc. Supplement 2, pp. 248-252.

IRWIN, J. O. (1935). Tests of significance for differences between percentages based on small numbers. Metron, 12 (2), 83-94.

WILKS, S. S. (1935). The likelihood test of independence in contingency tables. Ann. Math. Statist. 6, pp. 190-196.

1934

FISHER, R. A. (1934). Statistical Methods for Research Workers, Edinburgh 5th and subsequent editions. Oliver and Boyd Ltd., Section 21.02.

YATES, F. (1934). Contingency tables involving small numbers and the χ^2 test. J. Roy. Statist. Soc. Supplement 1, pp. 217-235.

1924

FISHER, R. A. (1924). The conditions under which chi-square measures the discrepancy between observation and hypothesis. J. Roy. Statist. Soc. 37, pp. 442-450.

1922

FISHER, R. A. (1922). On the interpretation of chi-square from contingency tables, and the calculation of P. J. Roy. Statist. Soc. 85, pp. 87-94.

1916

PEARSON, K. (1916). On the general theory of multiple contingency with special reference to partial contingency. Biometrika, 11, 145-158.

1915

GREENWOOD, M. and YULE, G. U. (1915). The statistics of anti-typhoid and anti-cholera inoculations and the interpretation of such statistics in general. Proc. Roy. Soc. Medicine, 8, 113-194.

1912

YULE, G. U. (1912). On the methods of measuring association between two attributes. Jour. Roy. Statist. Soc., 75, 579.

1904

PEARSON, K. (1904). Mathematical contributions to the theory of evolution. XIII On the theory of contingency and its relation to association and normal correlation. Draper's Company Research Memoirs, Biometric Series 1, 35 pp.

1900

PEARSON, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philos. Mag., Series 5 50, pp. 157-172.

1898

SHEPPARD, W. F. (1898). On the application of the theory of error to cases of normal distribution and normal correlation. Phil. Trans. Roy. Soc. London, A192, 101-167.