Analysis of Categorical Data by Linear Models

James E. Grizzle, C. Frank Starmer, Gary G. Koch

Biometrics, Volume 25, Issue 3 (Sep., 1969), 489-504.

Stable URL:
http://links.jstor.org/sici?sici=0006-341X%28196909%2925%3A3%3C489%3AAOCDBL%3E2.0.CO%3B2-N

ANALYSIS OF CATEGORICAL DATA BY LINEAR MODELS

JAMES E. GRIZZLE[1], C. FRANK STARMER[2], AND GARY G. KOCH[1]

[1]*Department of Biostatistics and* [2]*Department of Biomathematics, University of North Carolina, Chapel Hill, North Carolina 27514, U. S. A.*

SUMMARY

Assume there are $n_{i.}$, $i = 1, 2, \cdots, s$, samples from s multinomial distributions each having r categories of response. Then define any u functions of the unknown true cell probabilities $\{\pi_{ij} : i = 1, 2, \cdots, s; j = 1, 2, \cdots, r$, where $\sum_{j=1}^{r} \pi_{ij} = 1\}$ that have derivatives up to the second order with respect to π_{ij}, and for which the matrix of first derivatives is of rank u.

A general noniterative procedure is described for fitting these functions to a linear model, for testing the goodness-of-fit of the model, and for testing hypotheses about the parameters in the linear model.

The special cases of linear functions and logarithmic functions of the π_{ij} are developed in detail, and some examples of how the general approach can be used to analyze various types of categorical data are presented.

INTRODUCTION

Berkson has pointed out in 1968 and on several other occasions that the minimum logit χ^2 gives, for all practical purposes, numerically the same estimates and test statistics as maximum likelihood and Pearson's χ^2 for a variety of problems. There seems to be no widespread realization that the minimum logit χ^2 is only one member of a family whose test statistics have, in large samples, the χ^2-distribution if the null hypothesis is true; and that exploiting this fact leads to a unified approach, both conceptually and computationally, for analyzing categorical data. The computing aspects of the analysis of categorical data have been oriented to the desk calculator which tends to keep powerful general approaches from emerging, even though they have been available for some time. The general approach has the advantages of giving the analyst more latitude in choosing models and testing hypotheses which are precisely tailored to specific data, and makes it possible to have computer programs for the analysis of categorical data comparable in generality to those developed at many places for the analysis of linear models. The purpose of this paper is to present a general approach to the analysis of categorical data and to illustrate its use by examining some special cases. These methods represent an alternative set of procedures to those of Lewis [1968] which are based on maximum likelihood estimation.

The theoretical justification for the method presented can be found in Wald [1943] and Neyman [1949]. The equivalence of these two approaches for the class of problems we shall consider was demonstrated by Bhapkar [1966].

NOTATION

To fix ideas, consider the hypothetical data shown in Table 1 and the expected cell probabilities shown in Table 2.

TABLE 1

FREQUENCY DISTRIBUTION

Populations (factors)	Categories of response 1	2	· · ·	r	Total
1	n_{11}	n_{12}	· · ·	n_{1r}	$n_{1.}$
2	n_{21}	n_{22}	· · ·	n_{2r}	$n_{2.}$
·	·	·	· · ·	·	·
·	·	·	· · ·	·	·
·	·	·	· · ·	·	·
s	n_{s1}	n_{s2}	· · ·	n_{sr}	$n_{s.}$

Define

$$\underset{1\times r}{\boldsymbol{\pi}_i'} = [\pi_{i1}, \pi_{i2}, \cdots, \pi_{ir}]; \quad \underset{1\times rs}{\boldsymbol{\pi}'} = [\boldsymbol{\pi}_1', \boldsymbol{\pi}_2', \cdots, \boldsymbol{\pi}_s'];$$

$$p_{ij} = n_{ij}/n_{i.}; \quad \underset{1\times r}{\mathbf{p}_i'} = [p_{i1}, p_{i2}, \cdots, p_{ir}]; \quad \underset{1\times rs}{\mathbf{p}'} = [\mathbf{p}_1', \mathbf{p}_2', \cdots, \mathbf{p}_s'];$$

$$\text{var}(\mathbf{p}_i) = \underset{r\times r}{\mathbf{V}}(\boldsymbol{\pi}_i) = \frac{1}{n_{i.}} \begin{bmatrix} \pi_{i1}(1-\pi_{i1}) & -\pi_{i1}\pi_{i2} & \cdots & -\pi_{i1}\pi_{ir} \\ -\pi_{i2}\pi_{i1} & \pi_{i2}(1-\pi_{i2}) & \cdots & -\pi_{i2}\pi_{ir} \\ \vdots & \vdots & \cdots & \vdots \\ -\pi_{ir}\pi_{i1} & -\pi_{ir}\pi_{i2} & \cdots & \pi_{ir}(1-\pi_{ir}) \end{bmatrix};$$

$\underset{r\times r}{\mathbf{V}}(\mathbf{p}_i)$ = sample estimate of $\mathbf{V}(\boldsymbol{\pi}_i)$;

$\underset{rs\times rs}{\mathbf{V}}(\mathbf{p})$ = block diagonal matrix having $\mathbf{V}(\mathbf{p}_i)$ on the main diagonal;

$f_m(\boldsymbol{\pi})$ = any function of the elements of $\boldsymbol{\pi}$ that has partial derivatives up to

TABLE 2

EXPECTED CELL PROBABILITIES

Populations (factors)	Categories of response 1	2	· · ·	r	Total
1	π_{11}	π_{12}	· · ·	π_{1r}	1
2	π_{21}	π_{22}	· · ·	π_{2r}	1
·	·	·	· · ·	·	·
·	·	·	· · ·	·	·
·	·	·	· · ·	·	·
s	π_{s1}	π_{s2}		π_{sr}	1

second order with respect to the π_{ij}, $m = 1, 2, \cdots, u \leq (r-1)s$;

$f_m(\mathbf{p}) = f_m(\pi)$ evaluated at $\pi = \mathbf{p}$;

$[\mathbf{F}(\pi)]' = [f_1(\pi), f_2(\pi), \cdots, f_u(\pi)]$;

$\mathbf{F}' = [\mathbf{F}[(\mathbf{p})]' = [f_1(\mathbf{p}), f_2(\mathbf{p}), \cdots, f_u(\mathbf{p})]$;

$$\underset{u\times rs}{\mathbf{H}} = \left[\frac{\partial f_m(\pi)}{\partial \pi_{ij}} \,\middle|\, \pi_{ij} = p_{ij}\right]; \qquad \underset{u\times u}{\mathbf{S}} = \mathbf{HV}(\mathbf{p})\mathbf{H}'.$$

The matrix $\mathbf{S}$ is the sample estimate of the covariance matrix of $\mathbf{F}$. When $f_m(\mathbf{p})$ is a linear function of the elements of $\mathbf{p}$, $\mathbf{S}$ is the exact covariance matrix of $\mathbf{F}$; when $f_m(\mathbf{p})$ is a non-linear function of the elements of $\mathbf{p}$, $\mathbf{S}$ is the asymptotic covariance matrix of $\mathbf{F}$ which is obtained by the 'delta' method.

We assume that the functions $f_i(\pi)$ are jointly independent of one another and of the constraint $\sum_{j=1}^{r} \pi_{ij} = 1$, $\quad i = 1, 2, \cdots, s$; i.e., both $\mathbf{H}$ and $\mathbf{HV}(\pi)\mathbf{H}'$ are of rank u. When these conditions hold, then $\mathbf{S}$ is of rank u. However, for some types of data, if some of the $n_{ij} = 0$, $\mathbf{S}$ will be of rank less than u. Therefore, if difficulty is created by an occasional $n_{ij} = 0$, we follow Berkson [1955], appendix 3, and suggest that it be replaced by $1/r$. This has the effect of making the estimate of π_{ij} be $1/rn_{i.}$, which is the extension of Berkson's procedure to the multinomial case. However, we have made no extensive investigation of the effect of this rule in the multinomial case such as Berkson did for the binomial case. Incidentally, the method of estimation and testing set forth in this paper will yield Berkson's minimum 'logit χ^2' when specialized to s binomial distributions, $r = 2$, and the logit transformation, $f_m(\mathbf{p}) = \log_e(p_{i1}/p_{i2})$. Thus Berkson's [1955] work throws considerable light on the properties of our method for the special case he investigated so thoroughly.

ESTIMATION AND TESTING

To summarize, thus far we have defined u parametric, possibly non-linear, functions of π, the $\{f_m(\pi)\}$, their estimates $\{f_m(\mathbf{p})\}$, and their asymptotic covariance matrix $\mathbf{HV}(\mathbf{p})\mathbf{H}'$.

Assume that $\underset{u\times 1}{\mathbf{F}(\pi)} = \underset{u\times v}{\mathbf{X}}\ \underset{v\times 1}{\boldsymbol{\beta}}$, where $\mathbf{X}$ is a known design matrix (which is different from the usual design matrix when more than one function is constructed within each population as will be illustrated later) of rank $v \leq u$ and $\boldsymbol{\beta}$ is a vector of unknown parameters.

Several workers have shown that if the hypothesized model fits the data, a best asymptotic normal (BAN) estimate of $\boldsymbol{\beta}$ is given by $\mathbf{b}$, when $\mathbf{b}$ is the vector which minimizes $(\mathbf{F} - \mathbf{Xb})'\mathbf{S}^{-1}(\mathbf{F} - \mathbf{Xb})$. The minimum value of this form may be used to test the fit of the model $\mathbf{F}(\pi) = \mathbf{X}\boldsymbol{\beta}$. Given that the presumed model provides an adequate fit to the data, a test of the hypothesis $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is produced by conventional methods of weighted multiple regression, where $\mathbf{C}$ is a $(d \times v)$ matrix of arbitrary constants of full rank $d \leq v$.

The test statistic for the fit of the model is

$$SS[\mathbf{F}(\pi) = \mathbf{X}\boldsymbol{\beta}] = \mathbf{F}'\mathbf{S}^{-1}\mathbf{F} - \mathbf{b}'(\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})\mathbf{b}$$

which has asymptotically a (central) χ^2-distribution with $u - v$ D.F. if the model fits, where $\mathbf{b} = (\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}^{-1}\mathbf{F}$. Given the model, the test of the hypothesis $H_0 : \mathbf{C}\beta = \mathbf{0}$ is produced by $SS[\mathbf{C}\beta = \mathbf{0}] = \mathbf{b}'\mathbf{C}'[\mathbf{C}(\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\mathbf{b}$ which has asymptotically a χ^2-distribution with d D.F. if H_0 is true.

In many cases there is only one population and the objective of the statistical analysis is to study the relationships among several ways of classification of the sample units. Many tests appropriate to this problem can be formulated as $\underset{u\times 1}{\mathbf{F}(\pi)} = \mathbf{0}$. This fits into the general framework by setting $\mathbf{X} = \mathbf{0}$, the null matrix. Thus the test statistic is $\mathbf{F}'\mathbf{S}^{-1}\mathbf{F}$, which has asymptotically a χ^2-distribution with u D.F. if H_0 is true.

SPECIAL CASES OF $f(\pi)$

The form of $\mathbf{S}$ depends on $\mathbf{H}$ and through $\mathbf{H}$ on the function $\mathbf{F}(\pi)$. Therefore for each family of functions $\mathbf{F}(\pi)$, $\mathbf{S}$ will be different. Fortunately two classes of functions cover most applications discussed in the literature thus far. For linear relationships, one can define a family of functions, $\underset{u\times 1}{\mathbf{F}(\pi)} = \underset{u\times rs}{\mathbf{A}}\ \underset{rs\times 1}{\pi}$, where

$$\mathbf{A} = \begin{bmatrix} a_{111} & a_{112} & \cdots & a_{11r} ; & a_{121} & a_{122} & \cdots & a_{12r} ; & \cdots ; & a_{1s1} & a_{1s2} & \cdots & a_{1sr} \\ a_{211} & a_{212} & \cdots & a_{21r} ; & a_{221} & a_{222} & \cdots & a_{22r} ; & \cdots ; & a_{2s1} & a_{2s2} & \cdots & a_{2sr} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ a_{u11} & a_{u12} & \cdots & a_{u1r} ; & a_{u21} & a_{u22} & \cdots & a_{u2r} ; & \cdots ; & a_{us1} & a_{us2} & \cdots & a_{usr} \end{bmatrix}$$

is of rank $u \leq s(r-1)$; the $a_{\gamma ij}$ are arbitrary constants. For logarithmic relationships, one can define the family of functions

$$\underset{t\times 1}{\mathbf{F}(\pi)} = \underset{t\times u}{\mathbf{K}} \log_e \underset{u\times rs}{\mathbf{A}}\ \underset{rs\times 1}{\pi} ;$$

the α-th element of $\mathbf{F}(\pi)$ has the form

$$F_\alpha(\pi) = \sum_{\gamma=1}^{u} k_{\alpha\gamma} \log_e \left(\sum_{i,j} a_{\gamma ij}\pi_{ij}\right),$$

where the $a_{\gamma ij}$ and $k_{\alpha\gamma}$ are the appropriate elements of $\mathbf{A}$ and $\mathbf{K}$, respectively. Here, $\mathbf{K}$ is a matrix of arbitrary constants of rank $t \leq u \leq rs$. Some care must be exercised to make sure that the $\mathbf{H}$ associated with the functions described above is of full rank (i.e., of rank u for the linear case and of rank t for the logarithmic case).

The matrix of partials of the first transformation $\mathbf{F}(\pi) = \mathbf{A}\pi$ is $\mathbf{H} = \partial\mathbf{F}/\partial\pi = \mathbf{A}$, and $\mathbf{S} = \mathbf{A}\hat{\mathbf{V}}(\mathbf{p})\mathbf{A}'$. In the second case $\mathbf{H} = [\partial\mathbf{F}/\partial\pi \mid \pi = \mathbf{p}] = \mathbf{K}\mathbf{D}^{-1}\mathbf{A}$ and $\mathbf{S} = \mathbf{K}\mathbf{D}^{-1}\mathbf{A}\mathbf{V}(\mathbf{p})\mathbf{A}'\mathbf{D}^{-1}\mathbf{K}'$, where $\mathbf{A}$ is as defined previously; and

$$\mathbf{D} = \begin{bmatrix} \mathbf{a}_1'\mathbf{p} & 0 & \cdots & 0 \\ 0 & \mathbf{a}_2'\mathbf{p} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & & \mathbf{a}_u'\mathbf{p} \end{bmatrix},$$

where $\mathbf{a}_\gamma'$ represents the γ-th row of $\mathbf{A}$.

Many hypotheses can be produced as special cases of this formulation when correctly chosen matrices of constants are inserted for **X**, **C**, **A**, and **K**. This will now be illustrated.

ONE POPULATION PROBLEMS

Data collected on a single population are often tabulated in a multi-way table to study the association among the attributes defining the ways of classification; also other aspects of the structure of the data may be of interest. The data in Bhapkar's [1966] numerical example, shown in Table 3, are a good illustration of the latter case.

The hypothesis to be tested is homogeneity of marginal distributions. From a sampling point of view we have 7477 observations from a bivariate distribution, or a single multinomial distribution cross-classified two ways. Thus the appropriate hypothesis written in terms of the multinomial parameters is $H_0 : \pi_{1.} = \pi_{.1}$, $\pi_{2.} = \pi_{.2}$, $\pi_{3.} = \pi_{.3}$, $\pi_{4.} = \pi_{.4}$, where $\pi_{i.} = \sum_j \pi_{ij}$ and $\pi_{.j} = \sum_i \pi_{ij}$. This hypothesis can be written $\mathbf{F}(\boldsymbol{\pi}) = \mathbf{A}\boldsymbol{\pi} = \mathbf{0}$. Notice that $\pi_{1.} = \pi_{.1}$ implies $\pi_{12} + \pi_{13} + \pi_{14} - \pi_{21} - \pi_{31} - \pi_{41} = 0$. Thus H_0 can be written as $\mathbf{A}\boldsymbol{\pi} = \mathbf{0}$ if we note that

$$\boldsymbol{\pi}' = (\pi_{11}\pi_{12}, \cdots, \pi_{14}; \pi_{21}\pi_{22}, \cdots, \pi_{24}; \cdots; \pi_{41}\pi_{42}, \cdots, \pi_{44})$$

and choose

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 1 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

This matrix gives the correct hypothesis but it is singular since the sum of the first two rows is the negative of the sum of the last two rows. Therefore the test can be produced by deleting any row of **A**. These calculations reduce to

TABLE 3

7477 WOMEN AGED 30–39; UNAIDED DISTANCE VISION

Right eye	Left eye: Highest grade	Left eye: Second grade	Left eye: Third grade	Left eye: 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
Total	1907	2222	2507	841	7477

$X^2 = \mathbf{p}'\mathbf{A}^{*\prime}(\mathbf{A}^*\mathbf{V}(\mathbf{p})\mathbf{A}^{*\prime})^{-1}\mathbf{A}^*\mathbf{p}$, where $\mathbf{A}^*$ is the original $\mathbf{A}$ with one row deleted. When $\mathbf{A}$ is singular, the X^2 produced is invariant under the choice of a row basis of $\mathbf{A}$.

Denote the first row of $\mathbf{A}^*$ by $\mathbf{a}_1$, the second by $\mathbf{a}_2$ and the third by $\mathbf{a}_3$. Then $\mathbf{a}_1\mathbf{p} = 0.00923$, $\mathbf{a}_2\mathbf{p} = 0.00455$, $\mathbf{a}_3\mathbf{p} = -0.00682$; $\mathbf{p}'\mathbf{A}^{*\prime} = 10^{-2}(0.923, 0.455, -0.682)$ and

$$\mathbf{A}^*\mathbf{V}(\mathbf{p})\mathbf{A}^{*\prime} = 10^{-4}\begin{bmatrix} 0.1507 & -0.0894 & -0.0430 \\ -0.0894 & 0.2601 & -0.1420 \\ -0.0430 & -0.1420 & 0.2538 \end{bmatrix}.$$

Thus $X^2 = 11.98$ with 3 D.F.

Cochran [1950] discussed a problem for which he developed a test (sometimes called the Q test) by a permutation argument. An alternative test which is due to Bhapkar [1965] can be derived by the techniques presented. The data on 42 subjects, shown in Table 4, are used as an example. The subjects were given drugs A, B, and C. Some had a favorable response to a single drug, some to two, and some to all three. The patterns of response and the number of subjects showing each pattern are shown in Table 4.

If the three drugs are equally effective, $E(T_1) = E(T_2) = E(T_3)$. These data are considered as a single multinomially distributed sample of 46 in which each sample unit exhibits one of eight patterns of response. The hypothesis can be formulated in terms of cell probabilities as

$$\pi_1 + \pi_2 + \pi_3 + \pi_5 = \pi_1 + \pi_2 + \pi_4 + \pi_6 = \pi_1 + \pi_3 + \pi_4 + \pi_7$$

which simplifies to $\pi_2 - \pi_7 = \pi_4 - \pi_5 = \pi_3 - \pi_6$.
Then choose

TABLE 4

TABULATION OF RESPONSE TO DRUGS A, B, AND C
(1 DENOTES FAVORABLE RESPONSE, 0 DENOTES UNFAVORABLE RESPONSE)

Pattern of response				
A	B	C	Number	Expected probability
1	1	1	6	π_1
1	1	0	16	π_2
1	0	1	2	π_3
0	1	1	2	π_4
1	0	0	4	π_5
0	1	0	4	π_6
0	0	1	6	π_7
0	0	0	6	π_8
Number favorable			46	1
28	28	16		
T_1	T_2	T_3		

$$f_1(\pi) = \pi_2 - \pi_7 - \pi_4 + \pi_5 = 0, \qquad f_2(\pi) = \pi_2 - \pi_7 - \pi_3 + \pi_6 = 0.$$

A test of this hypothesis is produced easily by choosing

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & -1 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 \end{bmatrix}.$$

Then

$$(\mathbf{A}\hat{\mathbf{V}}(\mathbf{p})\mathbf{A}')^{-1} = \frac{46}{1241}\begin{bmatrix} 0.5406 & -0.4101 \\ -0.4101 & 0.5406 \end{bmatrix} 10^4,$$

$$\mathbf{a}_1\mathbf{p} = 0.26, \mathbf{a}_2\mathbf{p} = 0.26,$$

and

$$X^2 = [\mathbf{a}_1\mathbf{p}, \mathbf{a}_2\mathbf{p}](\mathbf{A}\hat{\mathbf{V}}(\mathbf{p})\mathbf{A}')^{-1}\begin{bmatrix} \mathbf{a}_1\mathbf{p} \\ \mathbf{a}_2\mathbf{p} \end{bmatrix} = 6.58$$

which has a χ^2-distribution with two D.F. if the null hypothesis is true. Cochran's Q test yields $X^2 = 8.47$. The relative merits of these two tests are not known. We mention the competitor here to show how the general method can be used to produce tests for a variety of problems which might be considered non-standard.

Another test of interest for the data in Table 4 is that of no interaction (as formulated by Bartlett [1935]) between the responses to the three drugs. One set of procedures (considered by Goodman [1963] and Plackett [1962]) is based on

$$f(\pi) = \ln \pi_1 - \ln \pi_2 - \ln \pi_3 - \ln \pi_4 + \ln \pi_5 + \ln \pi_6 + \ln \pi_7 - \ln \pi_8 = 0.$$

This test is directed at the general question of whether the effect of any two of the three drugs is independent of the third. This test fits into the general formulation by choosing

$$\mathbf{F}(\pi) = \mathbf{K} \log_e \mathbf{A}\pi$$

with $\mathbf{A} = \mathbf{I}$ and

$$\mathbf{K} = [1, -1, -1, -1, 1, 1, 1, -1].$$

Hence

$$\mathbf{K}\mathbf{D}^{-1}\mathbf{A} = \left[\frac{1}{p_1}, -\frac{1}{p_2}, -\frac{1}{p_3}, -\frac{1}{p_4}, \frac{1}{p_5}, \frac{1}{p_6}, \frac{1}{p_7}, -\frac{1}{p_8}\right],$$

and

$$\mathbf{K}\mathbf{D}^{-1}\mathbf{A}\mathbf{V}(\mathbf{p})\mathbf{A}'\mathbf{D}^{-1}\mathbf{K}' = 2.0625; \qquad f(\mathbf{p}) = 0.405.$$

Thus $X^2 = 0.08$, which has one D.F. Therefore the hypothesis of no interaction is not rejected.

The contrast tested is easily produced by observing that it is equivalent to the contrast for no three-way interaction among the logarithms of the cell probabilities of a 2^3 factorial experiment.

Many other examples could be given of published tests which follow from

this approach. The tests in Bhapkar and Koch [1968] can be generated in an obvious way by selection of the appropriate matrices. The difficult part remaining, which is true of all applied statistics, is that of choosing appropriate models and formulating the most informative hypotheses.

SEVERAL POPULATION PROBLEMS

The analysis of data collected on several populations (groups, factors) has many analogies with the analysis of variance. These analogies can be exploited usefully in many situations. We start with the basic model

$$\mathbf{F}(\pi) = \mathbf{X}\beta$$

and choose $\mathbf{F}(\pi)$ and $\mathbf{X}$ to suit our purpose. The data shown in Table 5 will be used to illustrate this method of analysis.

The data in Table 5 are a tabulation of the severity of the 'dumping syndrome,' an undesirable sequela of surgery for duodenal ulcer.

The four operations are:

A = Drainage and vagotomy;
B = 25% resection (antrectomy) and vagotomy;
C = 50% resection (hemigastrectomy) and vagotomy;
D = 75% resection.

Assign the categories of response, none, slight, and moderate, the scores 1, 2, and 3, respectively. The mean score for each treatment within each

TABLE 5
SEVERITY OF THE DUMPING SYNDROME

Clinical evaluation of severity of dumping syndrome	Hospital															
	1				2				3				4			
	Surgical procedure				Surgical procedure				Surgical procedure				Surgical procedure			
	A	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*
None	23	23	20	24	18	18	13	9	8	12	11	7	12	15	14	13
Slight	7	10	13	10	6	6	13	15	6	4	6	7	9	3	8	6
Moderate	2	5	5	6	1	2	2	2	3	4	2	4	1	2	3	4
Total *N*	32	38	38	40	25	26	28	26	17	20	19	18	22	20	25	23
Average score	1.3	1.5	1.6	1.6	1.3	1.4	1.6	1.7	1.7	1.6	1.5	1.8	1.5	1.4	1.6	1.6

hospital, calculated by $1p_{i1} + 2p_{i2} + 3p_{i3}$, is the $f(\mathbf{p})$ used in the analysis. In matrix terms this $f(\mathbf{p})$ is constructed by choosing

$$\underset{16\times 48}{\mathbf{A}} = \begin{bmatrix} 1\,2\,3 & 0\,0\,0 & 0\,0\,0 & \cdots & 0\,0\,0 \\ 0\,0\,0 & 1\,2\,3 & 0\,0\,0 & \cdots & 0\,0\,0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & \cdots & 1\,2\,3 \end{bmatrix}.$$

Then set $\mathbf{Ap} = \mathbf{X}\beta$, where $\mathbf{X}$ is the design matrix of an additive model having hospital and treatment effects. We show $\mathbf{X}$ in reparametrized form of full rank:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & -1 & -1 & -1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & -1 & -1 & -1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & -1 & -1 & 0 & 1 & 0 \\ 1 & -1 & -1 & -1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 \end{bmatrix} \quad \text{and} \quad \beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix},$$

where

μ = effect of the general mean;
α_i = differential effect of the i-th hospital, $i = 1, 2, 3$;
τ_j = differential effect of j-th treatment, $j = 1, 2, 3$.

The estimate of α_4 can be calculated by $\hat{\alpha}_4 = -\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\alpha}_3$ and similarly $\hat{\tau}_4 = -\hat{\tau}_1 - \hat{\tau}_2 - \hat{\tau}_3$.

The estimated parameters are

$$\hat{\mu} = 1.54, \hat{\alpha}_1 = -0.04, \hat{\alpha}_2 = -0.04, \hat{\alpha}_3 = 0.11,$$

$$\hat{\alpha}_4 = -(-0.04 - 0.04 + 0.11) = -0.03,$$

$$\hat{\tau}_1 = -0.11, \hat{\tau}_2 = -0.07, \hat{\tau}_3 = 0.05,$$

$$\hat{\tau}_4 = -(-0.11 - 0.07 + 0.05) = 0.13.$$

The following $\mathbf{C}$ matrices are used to compute the sum of squares for hospitals adjusted for treatment effects and treatments adjusted for hospital effects.

$$\mathbf{C}_{\text{Hospital}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{C}_{\text{Treatment}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The results of these tests can be summarized in an 'Anova' table.

ANOVA

Source of variation	Sums of squares	Degrees of freedom
Hospitals	2.33	3
Treatments	8.90	3
Error	6.32	9

These sums of squares are compared to the tabular values of χ^2 with the appropriate degrees of freedom. The error term does not approach statistical significance. This is interpreted to mean that an additive model having only mean, hospital, and treatment effects fits the data adequately. There are no significant hospital effects. However, treatment effects are significant at less than the 0.05 level.

A Scheffé type of multiple comparison procedure can be used (see Goodman [1964]) to investigate treatment differences further. It consists of comparing the test statistic for a single contrast $\mathbf{C}\beta = \mathbf{0}$ to the percentage point of the χ^2-distribution with v D.F., where v is the rank of the vector space which generates all possible contrasts for which the analyst wishes to control the frequency of Type I errors. Alternatively, a Scheffé type confidence interval for $\mathbf{C}\beta$ can be derived by

$$\mathbf{Cb} \pm \sqrt{[\chi_v^2 \mathbf{C}(\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{C}']},$$

where χ_v^2 is the appropriate percentage point of the χ^2-distribution with v D.F.

In this example it should be interesting to relate the severity of the dumping syndrome to the amount of stomach removed, and since the amount removed by the operations are approximately 0, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, we ask if the severity is related linearly to the amount removed. This can be tested by $H_0 : -3\tau_1 - \tau_2 + \tau_3 + 3\tau_4 = 0$ due to the equal spacings in the amount of stomach removed. In the reparametrized form this hypothesis becomes $-3\tau_1 - \tau_2 + \tau_3 + 3(-\tau_1 - \tau_2 - \tau_3) = -6\tau_1 - 4\tau_2 - 2\tau_3 = 0$ or $3\tau_1 + 2\tau_2 + \tau_3 = 0$. This hypothesis can be tested easily by choosing $\mathbf{C} = [0, 0, 0, 0, 3, 2, 1]$.

The resulting sum of squares is 8.74 which we compare to the tabular value of χ^2 with 3 D.F. if we wish to control the error level conservatively. Therefore we reject the hypothesis that there is no linear trend. From the amount of the total sum of squares for treatments accounted for by the test of linear trend and by examining the treatment mean scores, it is apparent that the severity of the

syndrome increases at least approximately linearly with the amount of stomach removed.

When the linear model was introduced, a remark was made that the matrix $\mathbf{X}$ was not always the same as in univariate multiple regression. As long as a single function is constructed within each of the s populations, the analogy with multiple regression remains unbroken. However, when two or more functions are constructed within each population, modifications must be made. Fortunately, they are rather simple.

We shall use the data presented by Kastenbaum and Lamphiear [1959], recently reanalyzed by Berkson [1968], to illustrate the method. The approach we shall present supplies a test of interaction which is equivalent to that of Berkson. Both Berkson's and ours are tests of the same hypothesis originally treated by Kastenbaum and Lamphiear, but they differ in the method of estimation used. The data described in Kastenbaum and Lamphiear's paper are shown in Table 6.

In this case $r = 3$, and $s = 10$. Both Berkson, and Kastenbaum and Lamphiear have shown that there is no interaction in the sense defined originally by Roy and Kastenbaum [1956]. An alternative interpretation of their tests, which leads numerically to the same statistic as calculated by Berkson and to other tests, is the following: define π_{i0}, π_{i1}, π_{i2} to be the expected probabilities of observing 0, 1, 2+ depletions respectively, and write

$$l_{i0} = \ln (\pi_{i0}/\pi_{i2}) \quad \text{and} \quad l_{i1} = \ln (\pi_{i1}/\pi_{i2}).$$

If the logarithmic functions l_{i0} and l_{i1} can be considered as additive functions of the mean effect, litter size effect, and treatment effect, there is no interaction (see Bhapkar and Koch [1968a, b]). The equations associated with this hypoth-

TABLE 6

DATA OF KASTENBAUM AND LAMPHIEAR [1959]

Litter size	Treatment	Number of depletions			
		0	1	2+	Total
7	A	58	11	5	74
	B	75	19	7	101
8	A	49	14	10	73
	B	58	17	8	83
9	A	33	18	15	66
	B	45	22	10	77
10	A	15	13	15	43
	B	39	22	18	79
11	A	4	12	17	33
	B	5	15	8	28

esis are the same as those derived by Roy and Kastenbaum. However, they were concerned with the one population problem with three responses, whereas the data in Table 6 should be considered as 'one response, two factors.' Then given the additive model, tests can be made on the treatment and litter size effects. To generate these tests we set

$$\begin{bmatrix} l_{10} \\ l_{20} \\ l_{30} \\ l_{40} \\ l_{50} \\ l_{60} \\ l_{70} \\ l_{80} \\ l_{90} \\ l_{10,0} \\ l_{11} \\ l_{21} \\ l_{31} \\ l_{41} \\ l_{51} \\ l_{61} \\ l_{71} \\ l_{81} \\ l_{91} \\ l_{10,1} \end{bmatrix} = \left[\begin{array}{cccccc:cccccc} 1 & 1 & 1 & 0 & 0 & 0 & & & & & & \\ 1 & -1 & 1 & 0 & 0 & 0 & & & & & & \\ 1 & 1 & 0 & 1 & 0 & 0 & & & & & & \\ 1 & -1 & 0 & 1 & 0 & 0 & & & & & & \\ 1 & 1 & 0 & 0 & 1 & 0 & & & 0 & & & \\ 1 & -1 & 0 & 0 & 1 & 0 & & & 10 \times 6 & & & \\ 1 & 1 & 0 & 0 & 0 & 1 & & & & & & \\ 1 & -1 & 0 & 0 & 0 & 1 & & & & & & \\ 1 & 1 & -1 & -1 & -1 & -1 & & & & & & \\ 1 & -1 & -1 & -1 & -1 & -1 & & & & & & \\ \hdashline & & & & & & 1 & 1 & 1 & 0 & 0 & 0 \\ & & & & & & 1 & -1 & 1 & 0 & 0 & 0 \\ & & & & & & 1 & 1 & 0 & 1 & 0 & 0 \\ & & & & & & 1 & -1 & 0 & 1 & 0 & 0 \\ & & & & & & 1 & 1 & 0 & 0 & 1 & 0 \\ & & 0 & & & & 1 & -1 & 0 & 0 & 1 & 0 \\ & & 10 \times 6 & & & & 1 & 1 & 0 & 0 & 0 & 1 \\ & & & & & & 1 & -1 & 0 & 0 & 0 & 1 \\ & & & & & & 1 & 1 & -1 & -1 & -1 & -1 \\ & & & & & & 1 & -1 & -1 & -1 & -1 & -1 \end{array}\right] \begin{bmatrix} \mu_0 \\ \alpha_{10} \\ \alpha_{20} \\ \alpha_{30} \\ \alpha_{40} \\ \alpha_{50} \\ \hdashline \mu_1 \\ \alpha_{11} \\ \alpha_{21} \\ \alpha_{31} \\ \alpha_{41} \\ \alpha_{51} \end{bmatrix}$$

Each non-zero block of $\mathbf{X}$ is derived from a reparametrization of the usual analysis of variance model so that it will be of full rank.

In terms of the general model originally presented $\mathbf{F}(\pi) = \mathbf{K}\ell$, where ℓ is the vector of logarithms of the elements of $\mathbf{A}\pi$ with $\mathbf{A} = \mathbf{I}$; and

$$\underset{20\times30}{\mathbf{K}} = \begin{bmatrix} 1\,0 & -1 & 0\,0 & 0 & \cdots & 0\,0 & 0 \\ 0\,0 & 0 & 1\,0 & -1 & \cdots & 0\,0 & 0 \\ \vdots & & \vdots & & \vdots\ \vdots\ \vdots & \vdots & \\ 0\,0 & 0 & 0\,0 & 0 & \cdots & 1\,0 & -1 \\ 0\,1 & -1 & 0\,0 & 0 & \cdots & 0\,0 & 0 \\ 0\,0 & 0 & 0\,1 & -1 & \cdots & 0\,0 & 0 \\ \vdots & & \vdots & & \vdots\ \vdots\ \vdots & \vdots & \\ 0\,0 & 0 & 0\,0 & 0 & \cdots & 0\,1 & -1 \end{bmatrix}.$$

The first ten rows of $\mathbf{K}$ pertain to the l_{i0} and the second 10 rows to the l_{i1}. Once $\mathbf{K}$ and $\mathbf{A}$ are specified, it is an easy matter to compute the asymptotic covariance matrix of the sample l_{i0} and l_{i1}, viz. $\hat{l}_{i0}$ and $\hat{l}_{i1}$. Next, the sample estimates are substituted for $\mathbf{K}\ell$ and its covariance matrix, and the remainder of the calculations are identical to weighted multiple regression. The estimated parameters are

$$\hat{\mu}_0 = 0.945, \hat{\alpha}_{10} = -0.278, \hat{\alpha}_{20} = 1.415, \hat{\alpha}_{30} = 0.846, \hat{\alpha}_{40} = 0.195, \hat{\alpha}_{50} = -0.514$$

$$\hat{\mu}_1 = 0.400, \hat{\alpha}_{11} = -0.278, \hat{\alpha}_{21} = 0.474, \hat{\alpha}_{31} = 0.153, \hat{\alpha}_{41} = 0.072, \hat{\alpha}_{51} = -0.401$$

In considering the analysis of $\hat{l}_{i0}$, the estimated parameter $\hat{\alpha}_{10}$ is the estimated A effect; and its negative is the B effect; the effect of the first four litter sizes are estimated by $\hat{\alpha}_{20}$, $\hat{\alpha}_{30}$, $\hat{\alpha}_{40}$, $\hat{\alpha}_{50}$ and the sixth litter size $\hat{\alpha}_{60}$, which is not shown, is the negative of the sum of the $\hat{\alpha}_{i0}$'s; $i = 2, 3, 4, 5$.

The sum of squares of deviation from the model, 3.1269, represents the test statistic for no interaction; it agrees with the result of Berkson [1968] who found 3.128. The two results should be identical except for rounding error. This error sum of squares can be interpreted as a test for no interaction or it can be interpreted as a test of goodness-of-fit of a multivariate additive model in the logarithms, having treatments and litter size as its parameters. If the latter interpretation is adopted, it is reasonable to proceed to test hypotheses about litter sizes and treatment.

The test for no effect of treatment on $\hat{l}_{i0}$ and $\hat{l}_{i1}$ simultaneously is produced by

$$\underset{2\times 12}{\mathbf{C}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and yields $X^2 = 6.41$ with 2 D.F. which is significant at less than the 0.05 level. To test the effect of the litter size on $\hat{l}_{i0}$ and $\hat{l}_{i1}$ simultaneously we choose

$$\underset{8\times 12}{\mathbf{C}} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

which yields $X^2 = 75.32$ with 8 D.F.

More specific hypotheses can be tested. For example, the linear effect of litter size is easily found under the assumption of equal spacing. Then we want to test

$$-2\alpha_2 - \alpha_3 + 0\alpha_4 + \alpha_5 + 2\alpha_6 = 0$$

for $\hat{l}_{i0}$ and $\hat{l}_{i1}$ simultaneously. α_6 was estimated by $-\hat{\alpha}_2 - \hat{\alpha}_3 - \hat{\alpha}_4 - \hat{\alpha}_5$; substituting for α_6 we get

$$-4\alpha_2 - 3\alpha_3 - 2\alpha_4 - \alpha_5 = 0.$$

This test is produced by choosing

$$\mathbf{C} = \begin{bmatrix} 0\ 0\ 4\ 3\ 2\ 1\ 0\ 0\ 0\ 0\ 0\ 0 \\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 4\ 3\ 2\ 1 \end{bmatrix},$$

which yields $X^2 = 67.70$ with 2 D.F. The same hypothesis can be tested for $\hat{l}_{i0}$ and $\hat{l}_{i1}$ separately. In this case we get $X^2 = 59.17$ and $X^2 = 4.674$, respectively, each having one D.F. By similar argument the test for quadratic effect of litter size is produced by

$$\mathbf{C} = \begin{bmatrix} 0\ 0\ 0\ 3\ 4\ 3\ 0\ 0\ 0\ 0\ 0\ 0 \\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 3\ 4\ 3 \end{bmatrix},$$

yielding $X^2 = 5.282$ with two D.F. which does not quite reach the 0.05 level.

From this analysis we can infer the following: a transformation of the observed proportions p_{ij}, $i = 1, \cdots, 10$; $j = 0, 1, 2$ by $\ln(p_{i0}/p_{i2})$ and $\ln(p_{i1}/p_{i2})$ produces a scale on which an additive model containing the effects of litter sizes and treatments produces an adequate fit to data as demonstrated by the lack of significant deviation from the model, i.e., no interaction. On this scale, treatments A and B are significantly different as are the litter sizes; and the number of depletions varies linearly with litter sizes.

It is important to note that the results of tests involving both l_{i0} and l_{i1} are invariant in the sense that the same test statistics would have been produced if l_{i0} and l_{i1} had been defined as $\ln(\pi_{i1}/\pi_{i0})$ and $\ln(\pi_{i2}/\pi_{i0})$; or in terms of some other linear transformation of $\mathbf{K}\ell$ in which the same vector space is spanned.

DISCUSSION

The approach presented here relegates the analysis of categorical data to a subclass of problems that can be handled by weighted regression. The authors feel that this unification is worthwhile because of the simplicity with which models and hypotheses can be formulated and tested. For many already well acquainted with linear models, this should be a welcome simplification. Also, the details of computing are greatly simplified. The two classes of functions, linear and logarithmic functions of the cell probabilities, do not exhaust the possibilities. On the other hand, many others could be handled by the same approach.

This approach is not without its problems. The behavior of the tests in 'small' samples is unknown. However, the same can be said for tests based on the maximum likelihood estimates. The occasional empty cell may require adjustment of the data by collapsing into a smaller array or by modification of $f(\pi)$ so that $\mathbf{S}$ is not singular. Recent work by Rao [1963] shows that the maximum likelihood estimate has a smaller variance than the BAN estimate used

here. To counter these arguments, the numerous examples presented by Berkson [1968] suggest that the analyst is not paying a high price for the simplicities that result from adopting this simple non-iterative procedure.

This procedure has been programmed for both the IBM 1130 (8k, 16 bit machine) and the IBM 360/75 (240k byte partition). For the 1130, the program will handle 16 × 16 matrices and for the 360/75, 65 × 65 matrices can be handled. The program deck and documentation for the 360/75 version are available from: The Program Librarian, Department of Biostatistics, University of North Carolina, Chapel Hill, N. C. 27514.

ACKNOWLEDGMENT

This work was supported by National Institutes of Health, Institute of General Medical Science Grant No. GM-12868-04 and GM 13625.

ANALYSE DE DONNES QUALITATIVES PAR DES MODELES LINEAIRES

RESUME

Supposons qu'il y ait n_i $i = 1, 2, \cdots, s$, échantillons de s distributions multinomiales chacune ayant r classes de réponses. Définissons alors u fonctions des probabilités vraies par cellule (inconnues) $(\pi_{ij}: i = 1, 2, \cdots, s; j = 1, 2, \cdots, r,$ où $\sum_{j=1}^{r} \pi_{ij} = 1)$ qui aient des dérivées jusqu'au second ordre par rapport à π_{ij} et pour lesquelles la matrice des dérivées premières soit de rang u.

Un procédé général non itératif est décrit pour ajuster ces fonctions à un modèle linéaire, pour tester la qualité d'ajustement de ce modèle et pour tester des hypothèses concernant les paramètres du modèle linéaire.

Les cas particuliers des fonctions linéaires et des fonctions logarithmiques des π_{ij} sont développées de façon détaillée et quelques exemples de la façon dont cette approche générale peut être utilisée pour analyser différents types de données qualitatives sont présentées.

REFERENCES

Bartlett, M. S. [1935]. Contingency table interactions. *J. R. Statist. Soc. Suppl. 2*, 248–52.

Berkson, J. [1955]. Maximum likelihood and minimum χ^2 estimates of the logistic function. *J. Amer. Statist. Ass. 50*, 130–62.

Berkson, J. [1968]. Application of minimum logit χ^2 estimate to a problem of Grizzle with a notation on the problem of no interaction. *Biometrics 24*, 75–95.

Bhapkar, V. P. [1965]. Categorical data analogs of some multivariate tests. To appear in *S. N. Roy Memorial Volume*, University of North Carolina Press.

Bhapkar, V. P. [1966]. A note on the equivalence of two test criteria for hypotheses in categorical data. *J. Amer. Statist. Ass. 61*, 228–35.

Bhapkar, V. P. and Koch, G. G. [1968a]. Hypotheses of 'no interaction' in multi-dimensional contingency tables. *Technometrics 10*, 107–23.

Bhapkar, V. P. and Koch, G. G. [1968b]. On the hypothesis of 'no interaction' in contingency tables. *Biometrics 24*, 567–94.

Cochran, W. G. [1950]. The comparison of percentages in matched samples. *Biometrika 37*, 256–66.

Goodman, L. A. [1963]. On Plackett's test for contingency table interactions. *J. R. Statist. Soc. B 25*, 179–88.

Goodman, L. A. [1964]. Simultaneous confidence intervals for contrasts among multinomial populations. *Ann. Math. Statist. 35*, 716–25.

Kastenbaum, M. A. and Lamphiear, D. E. [1959]. Calculation of chi-square to test the no three-factor interaction hypothesis. *Biometrics 15*, 107–15.

Lewis, J. A. [1968]. A program to fit constants to multiway tables of quantitative and quantal data. *Appl. Statist. 17*, 33–42.

Neyman, J. [1949]. Contribution to the theory of the χ^2 test. Pp. 239–73 in: *Proc. Berkeley Symp. Math Statist. Prob.* University of California Press, Berkeley and Los Angeles.

Plackett, R. L. [1962]. A note on interactions in contingency tables. *J. R. Statist. Soc. B 24*, 162–6.

Rao, C. R. [1963]. Criteria of estimation in large samples. Pp. 345–62 in: *Contributions to Statistics*. Pergamon Press, New York.

Roy, S. N. and Kastenbaum, M. A. [1965]. On the hypothesis of 'no interaction' in a multiway contingency table. *Ann. Math. Statist. 27*, 749–57.

Wald, A. [1943]. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc. 54*, 426–82.

Received July 1968, *Revised December* 1968