SOME TESTS FOR CATEGORICAL DATA

By V. P. Bhapkar[1]

University of North Carolina and University of Poona

1. Introduction and summary. We shall be concerned with experimental data given in the form of frequencies in cells determined by a multiway cross-classification, with predefined categories along each way of classification. Roy and Bhapkar [10] have posed hypotheses, which might be considered generalizations appropriate to this set up of the usual hypotheses in classical "normal" univariate "fixed effects" analysis of variance, "normal" multivariate "fixed effects" analysis of variance and analysis of various kinds of "normal" independence. Large sample tests for such hypotheses are offered here.

The large sample tests suggested are based on the χ^2-test of Karl Pearson [8]. The general probability model is that of a product of several multinomial distributions. According as the marginal frequencies along any dimension are held fixed or left free, that dimension is said to be associated with a "factor" or a "response" (or variable). The probability model is

$$\prod_j \frac{n_{oj}!}{n_{ij}!} \prod_i p_{ij}^{n_{ij}}, \tag{1}$$

where $\sum_i p_{ij} \equiv p_{oj} = 1$ and $\sum_i n_{ij} \equiv n_{oj}$ is held fixed. Thus i refers to categories of the response while j refers to categories of the factor. n_{oj} denotes the preassigned sample-size for the jth factor-category, out of which n_{ij} happen to lie in the ith response-category. It should be noticed that i may be a multiple subscript, say $i_1, i_2, \cdots, i_k$; j also may be a multiple subscript, say $j_1, j_2 \cdots, j_l$. We then speak of a k-response (or k-variate) and l-factor problem According as a set of real numbers is or is not associated with the categories along any way of classification (factor or response), that way of classification will be said to be structured or unstructured.

It is well-known (for example, Neyman [6]) that if a hypothesis H_o is given in the form of certain constraints on the p_{ij}'s, then a large sample test statistic of H_o under (1) for the model is a χ^2 statistic given by

$$\sum_{ij} (n_{ij} - n_{oj}\hat{p}_{ij})^2/(n_{oj}\hat{p}_{ij}),$$

or a χ_1^2 statistic given by $\sum_{ij} (n_{ij} - n_{oj}\hat{p}_{ij})^2/n_{ij}$, where the $\hat{p}_{ij}$'s form any set of BAN estimates [6]. In the particular case when the constraints are linear in p's, the method of minimum χ_1^2 permits a reduction of the problem to the solution of a system of linear equations and hence is more convenient.

Reiersøl [9] considers binomial experiments and makes use of results of Ney-

Received August 24, 1959; revised April 1, 1960.

[1] This research was supported partly by the Office of Naval Research under Contract No. Nonr-855(06) and partly by the Air Force Office of Scientific Research under Contract No. AF 49(638)-213.

man [6] to determine tests for hypotheses appropriate to factorial experiments. Mitra [5] not only generalizes Reiersøl's theorems to multinomial experiments, but also avoids his restriction that the parameter-sets in the different linear forms occurring in the hypothesis be nonoverlapping. We shall prove theorems to cover the cases that cannot be treated by these theorems.

In Section 2, the χ_1^2 statistic based on the minimum χ_1^2 estimates is obtained to test linear hypotheses. It is further shown that, when H_o specifies linear functions of the p's as known linear functions of some unknown parameters, the χ_1^2 statistic, based on the minimum χ_1^2 estimates, is exactly the same as the minimum sum of squares of residuals obtained by a certain general least squares technique to estimate the unknown parameters. This is then applied to derive test criteria appropriate to various hypotheses proposed in [3] and [10].

2. On testing linear hypotheses. In the notation of (1), let p_{ij} be nonzero for all (i, j). Since the event $\{n_{ij} > 0, \text{all } i, j\}$ has probability approaching one under this hypothesis, we may for asymptotic purposes assume that all the n_{ij}'s are nonzero. Consider a hypothesis H_o defined by m linearly independent constraints on the p_{ij}'s (independent of $\sum_{i=1}^r p_{ij} = 1$), say,

$$H_o: F_t(\mathbf{p}) \equiv \sum_{i=1}^{r} \sum_{j=1}^{s} f_{tij} p_{ij} + h_t = 0, \qquad t = 1, \cdots, m, \tag{2}$$

where f_{tij} and h_t are known constants such that the above equations, together with $\sum_{i=1}^r p_{ij} = 1$, have at least one set of solutions $\{p_{ij}\}$ for which the p_{ij}'s are positive.

Let

$$q_{ij} = n_{ij}/n_{oj}, \qquad b_{tj} = \textstyle\sum_i f_{tij} q_{ij}, \qquad b_t = \sum_j b_{tj},$$

$$c_t = b_t + h_t, \qquad e_{tt'j} = \textstyle\sum_i (f_{tij} - b_{tj})(f_{t'ij} - b_{t'j}) q_{ij},$$

$$g_{tt'} = \textstyle\sum_j e_{tt'j}/n_{oj}, \qquad \mathbf{c}' = (c_1, c_2, \cdots, c_m)$$

and $\mathbf{G} = (g_{tt'})$.

We notice that b_{tj} is in the nature of a "sample mean" of "F_t" for the jth sample, while $e_{tt'j}$ is in the nature of a sample covariance of "F_t" and "$F_{t'}$" for the jth sample. Since the F_t's are linearly independent, it follows that $\mathbf{G}$ is positive-definite.

THEOREM 1.

$$\operatorname*{Min}_{\text{subject to } H_0} \chi_1^2 = \mathbf{c}'\mathbf{G}^{-1}\mathbf{c}. \tag{3}$$

PROOF: To minimize χ_1^2 subject to the constraints we introduce Lagrangian multipliers, λ_j and μ_t, and write

$$f = \sum_j n_{oj} \sum_i \frac{(p_{ij} - q_{ij})^2}{q_{ij}} - 2 \sum_j \lambda_j \left(\sum_i p_{ij} - 1 \right) - 2 \sum_t \mu_t F_t(\mathbf{p}).$$

Differentiating with respect to p_{ij} and equating this to zero, we get the minimizing equations

$$n_{oj}\frac{(p_{ij}-q_{ij})}{q_{ij}} - \lambda_j - \sum_t \mu_t f_{tij} = 0, \qquad \begin{array}{l} i = 1, \cdots, r, \\ j = 1, \cdots, s. \end{array}$$

Multiplying by q_{ij} and summing over i, we get

$$-\lambda_j - \textstyle\sum_t \mu_t b_{tj} = 0.$$

Eliminating the λ's we get

$$p_{ij} = q_{ij}\left[1 + \frac{\sum_t \mu_t(f_{tij} - b_{tj})}{n_{oj}}\right],$$

where the μ's are to be determined from (2). Hence,

$$\sum_i \sum_j f_{tij}\, q_{ij}\left[1 + \frac{1}{n_{oj}}\sum_{t'} \mu_{t'}(f_{t'ij} - b_{t'j})\right] + h_t = 0, \quad t = 1, \cdots, m.$$

These may be written as $\mathbf{G}\boldsymbol{\mu} + \mathbf{c} = \mathbf{0}$, where $\boldsymbol{\mu}' = (\mu_1, \mu_2, \cdots, \mu_m)$. Hence

$$\boldsymbol{\mu} = -\mathbf{G}^{-1}\mathbf{c}.$$

Then

$$\begin{aligned} \min \chi_1^2 &= \sum_j n_{oj} \sum_i q_{ij}\left\{\frac{1}{n_{oj}}\sum_t \mu_t(f_{tij} - b_{tj})\right\}^2, \\ &= \sum_j \frac{1}{n_{oj}} \sum_t \sum_{t'} \mu_t\, \mu_{t'}\, e_{tt'j}\,; \\ &= \boldsymbol{\mu}'\mathbf{G}\boldsymbol{\mu}, \\ &= \mathbf{c}'\mathbf{G}^{-1}\mathbf{c}. \end{aligned}$$

REMARK: By Neyman's theorem, (Lemma 12, page 268 in [6]) if H_o is true, (3) is distributed in the limit as χ^2 with m d.f.

The form of (3) suggests that it may be the same as the statistic we would obtain if we test the hypothesis (2) by considering the b_t's, the natural unbiased estimates of $\sum_i \sum_j f_{tij} p_{ij}$, and using asymptotic normality. We have $b_t = \sum_i \sum_j f_{tij} q_{ij}$, so that $\mathcal{E}(b_t) = \sum_i \sum_j f_{tij} p_{ij} = -h_t$ if H_o is true, and

$$\begin{aligned} \operatorname{cov}(b_t, b_{t'}) &= \sum_j \sum_i f_{tij}\, f_{t'ij}\frac{p_{ij}(1-p_{ij})}{n_{oj}} - \sum_j \sum_{i \neq i'} f_{tij}\, f_{t'i'j}\frac{p_{ij}\, p_{i'j}}{n_{oj}}, \\ &= \sum_j \sum_i \frac{f_{tij}\, f_{t'ij}\, p_{ij}}{n_{oj}} - \sum_j \frac{1}{n_{oj}}\left(\sum_i f_{tij}\, p_{ij}\right)\left(\sum_i f_{t'ij}\, p_{ij}\right), \\ &= \phi_{tt'}, \quad \text{say.} \end{aligned}$$

Hence, in the limit, when H_o is true, $\mathbf{c}$ is asymptotically $N(\mathbf{0}, \boldsymbol{\Phi})$, so that $\mathbf{c}'\boldsymbol{\Phi}^{-1}\mathbf{c}$ is asymptotically distributed as χ^2 with md.f. If we replace p_{ij} in $\boldsymbol{\Phi}$ by q_{ij} we get $\mathbf{G}$. Hence $\mathbf{G}$ may be considered as an estimate of $\boldsymbol{\Phi}$. Thus we have proved

THEOREM 2: *The minimum χ_1^2 method to test the linear hypothesis* (2) *is exactly equivalent to the "large sample test" based on the asymptotic normality of the unbiased estimates of $F_t(\mathbf{p})$, whose variance-covariance matrix is estimated by the "sample variance-covariance matrix".*

Invariance. We then expect the χ_1^2 statistic to be invariant under the choice of linearly independent constraints (on the p_{ij}'s) defining the same hypothesis (2). This can be easily proved.

2.1 *Structured response.* Sometimes a linear hypothesis is defined by linear functions of unknown parameters. Theoretically, of course, this can be reduced to the case, already considered, when the hypothesis is defined by linear constraints on the p's. But, in many cases, this equivalent expression in terms of linearly independent constraints on the p's may be tedious to work out. We prove a theorem, which might be considered as another version of Theorems 1 and 2, which reduces the problem to that of least squares.

THEOREM 3. *Let a linear hypothesis be defined by*

$$H_o: \textstyle\sum_i a_i p_{ij} = d_{j1}\theta_1 + d_{j2}\theta_2 + \cdots + d_{jt}\theta_t , \qquad j = 1, \cdots , s \tag{4}$$

where the d's are known constants and the θ's are unknown parameters. Then the minimum χ_1^2 to test H_o is the same as the minimum sum of squares of residuals obtained by the general least squares technique on $\sum_i a_i q_{ij}$, with the variances estimated by "sample-variances". Moreover, the min χ_1^2 is asymptotically χ^2 with $s - u$ d.f., where $u = \text{Rank } (d_{jk})$.

INDICATION OF THE PROOF. It can be easily shown that H_o is equivalent to

$$\textstyle\sum_i \sum_j l_{vj} a_i p_{ij} = 0, \qquad v = 1, 2, \cdots , s - u$$

where $\mathbf{LD} = \mathbf{0}$ and $\mathbf{L}$ is of rank $(s - u)$; $\mathbf{D} = (d_{jk})$ and $\mathbf{L} = (l_{vj})$.

Let

$$\alpha_j = \textstyle\sum_i a_i q_{ij} , \qquad \beta_j = \sum_i (a_i - \alpha_j)^2 q_{ij} , \qquad \lambda_j = \beta_j / n_{oj} ,$$

$$\mathbf{\Lambda} = \text{diagonal } (\lambda_1 , \cdots , \lambda_s) \quad \text{and} \quad \boldsymbol{\alpha}' = (\alpha_1 , \cdots , \alpha_s).$$

Then by Theorem 1,

$$\text{Min } \chi_1^2 = \boldsymbol{\alpha}'\mathbf{L}'(\mathbf{L\Lambda L}')^{-1}\mathbf{L}\boldsymbol{\alpha}. \tag{5}$$

On the other hand, the α_j's are independent with variances

$$[\textstyle\sum_i a_i^2 p_{ij} - \{ \sum_i a_i p_{ij} \}^2]/n_{oj} ,$$

so that the "sample variances" are λ_j, $j = 1, 2, \cdots , s$. If we use the least squares technique on the α_j's (using the λ_j's for "variance"), then the sum of squares to be minimized with respect to the parameters is

$$S^2 = \textstyle\sum_j (\alpha_j - d_{j1}\theta_1 - \cdots - d_{jt}\theta_t)^2/\lambda_j .$$

Min S^2, then, can be shown to be (5). The last statement in the theorem follows immediately from the remark on Theorem 1.

2.2 *Applications of Theorem* 3 *to univariate linear hypotheses.* In what follows, "i" denotes a structured response. We discuss some simple cases chosen from those considered in [3, 10].

(i) *One dimensional design* ("j" → "*Treatment*") *Hypothesis of no treatment effects.*

$$H_o: \sum_i a_i p_{ij} \text{ is independent of } j.$$

$$\chi_1^2 = \sum_{j=1}^{s} (n_{oj}\alpha_j^2/\beta_j) - \left[\sum_{j=1}^{s} n_{oj}\alpha_j \Big/ \beta_j\right]^2 \Big/ \left(\sum_{j=1}^{s} n_{oj} \Big/ \beta_j\right),$$

$$\text{d.f.} = s - 1,$$

where

$$\alpha_j = \sum_i a_i q_{ij} \quad \text{and} \quad \beta_j = \sum_i (a_i - \alpha_j)^2 q_{ij}.$$

(ii) *Two-dimensional design* ("j" → "Treatment")

("k" → "Block").

(a) *Hypothesis of no treatment effects on the basic model.*

$$H_o: \sum_i a_i p_{ijk} = b_k, \qquad j = 1, \cdots, s, \quad k = 1, \cdots, t.$$

Note that the design may be incomplete, i.e., all combinations (j, k) may not occur.

$$\chi_1^2 = \sum_j \sum_k \alpha_{jk}^2 h_{jk} - \sum_{k=1}^{t} \left[\sum_j h_{jk}\, \alpha_{jk}\right]^2 \Big/ \sum_j h_{jk}, \quad \text{d.f.} = M - t,$$

where

$$\alpha_{jk} = \sum_i a_i q_{ijk}, \qquad \beta_{jk} = \sum_i (a_i - \alpha_{jk})^2 q_{ijk}$$

and

$$h_{jk} = n_{ojk}/\beta_{jk},$$

the summation is over allowable (j, k) combinations and M is the number of (j, k) combinations. When the design is complete, $M = st$, so that d.f. $= (s - 1)t$.

(b) *Hypothesis of no interaction* (in the additive set up).

$$H_o: \sum_i a_i p_{ijk} = t_j + b_k.$$

$$(6) \quad \chi_1^2 = \sum_j \sum_k \alpha_{jk}^2 h_{jk} - \sum_{j=1}^{s} Q_j t_j - \sum_{k=1}^{t} B_k^2/h_{ok}, \quad \text{d.f.} = M - (s + t - 1),$$

where the t's satisfy

$$(7) \qquad Q_j = \sum_{j=1}^{s} c_{jj'} t_{j'}, \qquad j = 1, \cdots, s,$$

and

$$B_k = \sum_j \alpha_{jk} h_{jk}, \qquad T_j = \sum_k \alpha_{jk} h_{jk},$$

$$h_{ok} = \sum_j h_{jk}, \qquad h_{jo} = \sum_k h_{jk},$$

$$Q_j = T_j - \sum_k B_k h_{jk}/h_{ok}, \qquad c_{jj} = h_{jo} - \sum_k h_{jk}^2/h_{ok}$$

and

$$c_{jj'} = -\sum_k h_{jk} h_{j'k}/h_{ok}.$$

Here M is, as before, the number of (j, k) combinations and the summations are over allowable combinations only.

It may be noted that (6) and (7) are similar to the "error sum of squares" and the "normal equations", respectively, in analysis of variance, T_j and B_k playing the roles of a "treatment total" and a "block total", respectively. The fundamental difference, however, is that the $c_{jj'}$'s here depend not only on the design but also on the observed proportions. In normal ANOVA, the designs can be chosen suitably so that the normal equations have neat closed solutions. This approach fails here for the corresponding equations (7). For example, even for a complete design (which may be called a "randomized block design", there is no essential simplification in the equations (7). (The degrees of freedom for χ_1^2 in that case are $= (s-1)(t-1)$.)

(c) *Hypothesis of no treatment effects on the no interaction model.*

$$\chi_1^2 = \sum_{j=1}^{s} Q_j t_j, \qquad \text{d.f.} = s - 1,$$

where the Q's and the t's are defined as before.

(d) *Hypothesis of linearity of regression on treatment levels (independent of blocks).*

$$H_o: \sum_i a_i p_{ijk} = \lambda + \mu b_j.$$

$$\chi_1^2 = \sum_j \sum_k \alpha_{jk}^2 h_{jk} - [G^2 l - 2G\gamma m + \gamma^2 h]/(hl - m^2), \qquad \text{d.f.} = M - 2,$$

where

$$h = \sum_j \sum_k h_{jk}, \qquad m = \sum_{j=1}^{s} b_j h_{jo},$$

$$l = \sum_{j=1}^{s} b_j^2 h_{jo}, \qquad G = \sum_{j=1}^{s} T_j = \sum_{k=1}^{t} B_k$$

and

$$\gamma = \sum_{j=1}^{s} b_j T_j.$$

(Other quantities are defined as before.)

(e) *Hypothesis of linearity of regression on treatment levels* (*the regression coefficient being independent of blocks*).

$H_0 : \sum_i a_i\, p_{ijk} = \lambda_k + \mu b_j\,.$

$$\chi_1^2 = \sum_j \sum_k \alpha_{jk}^2\, h_{jk} - \sum_{k=1}^{t} \left(\frac{B_k^2}{h_{ok}}\right) - \left[\gamma - \sum_{k=1}^{t} \frac{B_k\, m_k}{h_{ok}}\right]^2 \Big/ \left[l - \sum_{k=1}^{t} \left(\frac{m_k^2}{h_{ok}}\right)\right],$$

$$\text{d.f.} = M - t - 1,$$

where $m_k = \sum_j b_j h_{jk}$. (Other quantities are defined as before.)

(f) *Hypothesis of linearity of regression on treatment levels.*

$H_0 : \sum_i a_i\, p_{ijk} = \lambda_k + \mu_k\, b_j\,.$

$$\chi_1^2 = \sum_j \sum_k \alpha_{jk}^2\, h_{jk} - \sum_{k=1}^{t} \left(\frac{B_k^2}{h_{ok}}\right) - \sum_{k=1}^{k} \left[\left(\gamma_k - \frac{B_k\, m_k}{h_{ok}}\right)^2 \Big/ \left(l_k - \frac{m_k^2}{h_{ok}}\right)\right],$$

$$\text{d.f.} = M - 2t,$$

where

$$\gamma_k = \sum_j \alpha_{jk} h_{jk} b_j \quad \text{and} \quad l_k = \sum_j b_j^2 h_{jk}\,.$$

(Other quantities are defined as before.)

(iii) *Two-dimensional design* ("j" $\rightarrow$ factor)

("k" $\rightarrow$ another factor).

Hypothesis of linearity of regression on the factor-levels.

$$H_0 : \sum_i a_i p_{ijk} = \lambda + \mu b_j + \nu c_k.$$

$$\chi_1^2 = \sum_j \sum_k \alpha_{jk}^2 h_{jk} - \hat{\lambda} G - \hat{\mu}\gamma - \hat{\nu}\delta, \qquad \text{d.f.} = M - 3,$$

where $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\nu}$ satisfy the equations

$$G = \hat{\lambda} h + \hat{\mu} m + \hat{\nu} w$$
$$\gamma = \hat{\lambda} m + \hat{\mu} l + \hat{\nu} x$$
$$\delta = \hat{\lambda} w + \hat{\mu} x + \hat{\nu} y,$$

where

$$\delta = \sum_{k=1}^{t} c_k B_k\,, \qquad w = \sum_{k=1}^{t} c_k h_{ok}\,,$$

$$x = \sum_j \sum_k b_j c_k h_{jk} \quad \text{and} \quad y = \sum_{k=1}^{t} c_k^2 h_{ok}\,.$$

(Other quantities are defined as before.)

2.3. *Application of* 2.1 *to multivariate linear hypotheses.* Let us consider, as a further illustration, a bivariate randomized block experiment. If the two responses are structured, a_{i_1} and b_{i_2} being the weights associated with the respective categories, the hypothesis of no treatment effects takes the form

$$H_o: \quad \sum_{i_1=1}^{r_1} a_i p_{i_1 o jk} \text{ is independent of } k$$
$$\sum_{i_2=1}^{r_2} b_{i_2} p_{o i_2 jk} \text{ is independent of } k,$$

where j and k denote the block and treatment respectively; o in the place of a subscript denotes a summation over that subscript.

It can be easily seen that this hypothesis, or, more generally such hypotheses for p structured variables can be expressed in the form

$$(8) \qquad H_o: \quad \begin{aligned} &\sum_{i_1=1}^{r_1} \sum_j f_{t i_1 ** \cdots * j} p_{i_1 o \cdots o j} + h_t^{(1)} = 0 \\ &\cdots \\ &\sum_{i_p=1}^{r_p} \sum_j f_{t ** \cdots * i_p j} p_{o \cdots o i_p j} + h_t^{(p)} = 0, \quad t = 1, 2, \cdots, m, \end{aligned}$$

where the linear functions are linearly independent. We can write these as

$$\textstyle\sum_i \sum_j f_{t ** \cdots * i_k ** \cdots * j} p_{ij} + h_t^{(k)} = 0, \begin{cases} k = 1, 2, \cdots, p \\ t = 1, 2, \cdots, m, \end{cases}$$

so that, (8) is a particular case of (2). Hence

$$\chi_1^2 = \mathbf{c}'\mathbf{G}^{-1}\mathbf{c}, \qquad \text{d.f.} = pm,$$

where

$$\mathbf{c}'_{1 \times mp} = (c_1^{(1)}, \cdots, c_m^{(1)}; c_1^{(2)}, \cdots, c_m^{(2)}; \cdots c_m^{(p)}),$$

$$\mathbf{G}_{pm \times pm} = \begin{bmatrix} \mathbf{G}^{11}\mathbf{G}^{12} \cdots \mathbf{G}^{1p} \\ \mathbf{G}^{p1}\mathbf{G}^{p2} \cdots \mathbf{G}^{pp} \end{bmatrix}, \qquad \mathbf{G}^{kk'} = (g_{ij}^{(kk')})_{m \times m},$$

$$g_{tt'}^{(kk')} = \textstyle\sum_j e_{tt'j}^{(kk')} / n_{oj},$$

$$e_{tt'j}^{(kk')} = \sum_{i_k=1}^{r_k} \sum_{i_{k'}=1}^{r_{k'}} f_{t* \cdots * i_k * \cdots * j} f_{t' * \cdots * i_{k'} * \cdots * j} q_{o \cdots o i_k o \cdots o i_{k'} o \cdots o j} - b_{tj}^{(k)} b_{t'j}^{(k')},$$

$$b_{tj}^{(k)} = \sum_{i_k=1}^{r_k} f_{t* \cdots * i_k * \cdots * j} q_{o \cdots o i_k o \cdots o j}$$

and

$$c_t^{(k)} = h_t^{(k)} + \textstyle\sum_j b_{tj}^{(k)}.$$

2.4 *Unstructured Response.* In Theorem 1, we considered the test criterion appropriate to a linear hypothesis. Its equivalence to a certain least squares technique for linear hypotheses in structured cases was established in Theorem 3. We shall prove a similar equivalence for linear hypotheses in unstructured cases in Theorem 4.

THEOREM 4. Let a linear hypothesis be defined by

$$H_o: p_{ij} = d_{j1}\theta_{i1} + d_{j2}\theta_{i2} + \cdots + d_{jt}\theta_{it}, \quad i = 1, 2, \cdots, r \quad j = 1, 2, \cdots, s, \tag{9}$$

where the d's are known constants and the θ's are unknown parameters. Then the minimum χ_1^2 to test H_o is the same as the minimum "generalized sum of squares" of residuals obtained by a "generalized least squares technique" on q_{ij}, with the covariance matrix estimated by the "sample covariance matrix". Moreover, the min χ_1^2 is asymptotically χ^2 with $(r - 1)(s - u)$ d.f., where $u = \text{Rank}(d_{jk})$.

INDICATION OF THE PROOF. It can be easily shown that H_o is equivalent to

$$\sum_{j=1}^{s} l_{vj} p_{ij} = 0, \quad i = 1, 2, \cdots, r - 1 = r' \text{ (say)} \quad v = 1, 2, \cdots, s - u,$$

where $\mathbf{LD} = \mathbf{0}$ and $\mathbf{L}$ is of rank $(s - u)$; $\mathbf{D} = (d_{jk})$ and $\mathbf{L} = (l_{vj})$.

Let

$$\mathbf{q}' = (q_{11}, q_{21}, \cdots, q_{r'1}; \cdots; q_{1s}, \cdots, q_{r's}),$$

$$\mathbf{L}^* = \mathbf{L} \times \mathbf{I} = \begin{pmatrix} l_{11}\mathbf{I}_{r'} & \cdots & l_{1s}\mathbf{I}_{r'} \\ l_{s-u,1}\mathbf{I}_{r'} & \cdots & l_{s-u,s}\mathbf{I}_{r'} \end{pmatrix},$$

$$y_{ii'}^{j} = \delta_{ii'} q_{ij} - q_{ij} q_{i'j}, \qquad \delta_{ij} = 0 \text{ if } i \neq j, \quad = 1 \text{ if } i = j,$$

$$\mathbf{Y}_j = (y_{ii'}^{j})$$

and

$$\mathbf{Y} = \begin{bmatrix} n_{o1}^{-1}\mathbf{Y}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & n_{o2}^{-1}\mathbf{Y}_2 & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & n_{os}^{-1}\mathbf{Y}_s \end{bmatrix}.$$

Then, from (3),

$$\chi_1^2 = \mathbf{q}'\mathbf{L}^{*'}[\mathbf{L}^*\mathbf{Y}\mathbf{L}^{*'}]^{-1}\mathbf{L}^*\mathbf{q}. \tag{10}$$

This has already been obtained by Mitra [5] in a slightly different form.

On the other hand, if we consider the asymptotically normal variables q_{ij}'s, then

$$\text{cov}(\mathbf{q}_j) = \mathbf{Y}_j / n_{oj},$$

where

$$\mathbf{q}_j' = (q_{1j}, \cdots, q_{r'j}).$$

Let

$$S^2 = \sum_j n_{oj}(\mathbf{q}_j - d_{j1}\boldsymbol{\theta}_1 - \cdots - d_{jt}\boldsymbol{\theta}_t)'\mathbf{Y}_j^{-1}(\mathbf{q}_j - d_{j1}\boldsymbol{\theta}_1 - \cdots - d_{jt}\boldsymbol{\theta}_t),$$

where

$$\boldsymbol{\theta}_k' = (\theta_{1k}, \theta_{2k}, \cdots, \theta_{r'k}).$$

S^2 may be called the "generalized sum of squares" of residuals and it is to be minimized with respect to $\boldsymbol{\theta}$'s. Min S^2, then can be shown to be (10).

The last statement in the theorem follows immediately from the remark on Theorem 1.

A possible application of Theorem 4, is, for example, to test the hypothesis of no interaction in the additive set up, given by

$$H_o: p_{ijk} = b_{ik} + t_{ij}, \qquad \begin{aligned} j &= 1, 2, \cdots, s \\ k &= 1, 2, \cdots, t, \end{aligned}$$

where j and k refer to a "treatment" and a "block" respectively. We shall not consider the details of the computation. It is easy to verify that the d.f. will be $(r - 1)(M - s - t + 1)$, where M is the possible number of combinations (j, k), and which reduces to $(r - 1)(s - 1)(t - 1)$ for a complete design. Moreover, to test the hypothesis of no treatment effects, given by

$$H_o: p_{ijk} = b_{ik},$$

on the above model of no interaction, the test-statistic would be

$$\sum_{i,j,k} \frac{\left(n_{ijk} - \dfrac{n_{ojk}\, n_{iok}}{n_{ook}}\right)^2}{\dfrac{n_{ojk}\, n_{iok}}{n_{ook}}} - \chi_1^2, \qquad \text{d.f.} = (r-1)(s-1),$$

where χ_1^2 is the statistic appropriate for the hypothesis of no interaction. The first part of the expression has been given already by Roy and Mitra [12] as the appropriate statistic to test the hypothesis of no treatment effects on the basic model.

3. On the test of nonlinear hypotheses. In such cases Neyman's technique of linearization [6] may be adopted, so that the problem is reduced to one of the previous cases. On the other hand, it may happen in some cases that the maximum likelihood equations are fairly simple so that the χ^2 statistic, based on the maximum likelihood estimates, may be used.

3.1 *Minimum* χ_1^2 *by "linearization"*. If the hypothesis is defined by $F_t(\mathbf{p}) = 0$, $t = 1, 2, \cdots, m$, where the F's satisfy the regularity conditions (see page 254 in [6]), the linearization gives

$$F_t^*(\mathbf{q}, \mathbf{p}) \equiv F_t(\mathbf{q}) + \sum_i \sum_j \left[\frac{\partial F_t(\mathbf{p})}{\partial p_{ij}}\right]_{\mathbf{p}=\mathbf{q}} (p_{ij} - q_{ij}) = 0, \qquad t = 1, 2, \cdots, m.$$

Let

$$\left(\frac{\partial F_t(\mathbf{p})}{\partial p_{ij}}\right)_{\mathbf{p}=\mathbf{q}} = f_{tij}$$

and

$$h_t = F_t(\mathbf{q}) - \textstyle\sum_i \sum_j f_{tij} q_{ij} .$$

Then, from (3),

$$\chi_1^2 = \mathbf{f}'\mathbf{G}^{-1}\mathbf{f}, \qquad \text{d.f.} = m,$$

where

$$\mathbf{f}' = [F_1(\mathbf{q}), \cdots, F_m(\mathbf{q})].$$

3.2 *The hypothesis of no interaction (multiplicative set-up) in the two-dimensional design.*

$$H_o: p_{ijk} = b_{ij} t_{ik}, \qquad i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, s;\ k = 1, 2, \cdots, t.$$

This may be tested by the linearization technique mentioned above. On the other hand, in the case of a *complete design*, the maximum likelihood equations appear to be fairly simple and may admit an iterative solution. It can be easily shown that H_o is equivalent to

$$p_{ijk} p_{ist} = p_{isk} p_{ijt}, \qquad i = 1, \cdots, r;\ j = 1, \cdots, s-1;\ k = 1, \cdots, t-1. \tag{11}$$

The maximum likelihood equations, subject to (11) and $\sum_i p_{ijk} = 1$, can be obtained by differentiating

$$f = \sum_{i,j,k} n_{ijk} \log p_{ijk} - \sum_{j=1}^{s} \sum_{k=1}^{t} \lambda_{jk} [\sum_i p_{ijk} - 1]$$
$$- \sum_{i=1}^{r} \sum_{j=1}^{s-1} \sum_{k=1}^{t-1} \mu_{ijk} [\log p_{ijk} + \log p_{ist} - \log p_{isk} - \log p_{tji}] \tag{12}$$

with respect to the p's, where the λ's and μ's are Lagrangian multipliers. The final equations are

$$\frac{(n_{ijk} - \mu_{ijk})(n_{ist} - \mu_{ioo})}{(n_{isk} + \mu_{iok})(n_{ijt} + \mu_{ijo})} = \frac{(n_{ojk} - \mu_{ojk})(n_{ost} - \mu_{ooo})}{(n_{osk} + \mu_{ook})(n_{ojt} + \mu_{ojo})},$$

$$i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, s-1 \text{ and } k = 1, \cdots, t-1,$$

where

$$\mu_{iok} = \sum_{j=1}^{s-1} \mu_{ijk} , \qquad \mu_{ijo} = \sum_{k=1}^{t-1} \mu_{ijk} , \qquad \mu_{ojk} = \sum_{i=1}^{r} \mu_{ijk} , \qquad \text{etc.}$$

In particular, when $r = s = t = 2$, we have just two equations (linear) and these can be explicitly solved. In this special case, Bartlett [2] has posed another hypothesis of no interaction, but the solution of the maximum likelihood equation comes out as a root of a certain cubic equation. Mitra [5] has shown that it is the numerically smallest real root that gives the consistent solution. The equations, in the present case, thus seem to be simpler. Roy and Kastenbaum [11] have extended Bartlett's hypothesis to more general cases where "i", "j" and "k" are variables, and they get equations similar to (12).

4. Acknowledgment. I am indebted to Professor S. N. Roy for suggesting the problem and for his constant encouragement and criticism. Thanks are also due to the referee for suggestions that have improved the form of this paper.

REFERENCES

[1] G. A. Barnard, "Significance tests for 2 × 2 tables," *Biometrika*, Vol. 34 (1947), pp. 123–138.

[2] M. S. Bartlett, "Contingency table interactions," *Suppl., J. Roy. Stat. Soc.*, Vol. 2 (1935), pp. 248–252.

[3] V. P. Bhapkar, "Contributions to the statistical analysis of experiments with one or more responses (not necessarily normal)," North Carolina Institute of Statistics Mimeograph Series No. 229, (1959), pp. 1–123.

[4] R. A. Fisher, "On the interpretation of chi-square from contingency tables and the calculation of p," *J. Roy. Stat. Soc.*, Vol. 85 (1922), pp. 87–94.

[5] S. K. Mitra, "Contributions to the Statistical Analysis of Categorical Data," North Carolina Institute of Statistics Mimeograph Series No. 142, (1955), pp. 1–146.

[6] J. Neyman, "Contributions to the theory of χ^2 test," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1949, pp. 239–273.

[7] E. S. Pearson, "The choice of statistical tests illustrated on the interpretation of data classed in a 2 × 2 table," *Biometrika*, Vol. 34 (1947), pp. 139–163.

[8] Karl Pearson, "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.*, Ser. 5, Vol. 50 (1900), pp. 157–175.

[9] Olav Reiersøl, "Tests of linear hypotheses concerning binomial experiments," *Skandinavisk Aktuarietidskrift*, Vol. 37 (1954), pp. 38–59.

[10] S. N. Roy and V. P. Bhapkar, "Some nonparametric analogs of 'normal' ANOVA, MANOVA, and of studies in 'normal' association," *Contributions to Probability and Statistics, Essays in Honor of Harold Hotelling*, Stanford University Press, Stanford, 1960, pp. 371–387.

[11] S. N. Roy and Marvin A. Kastenbaum, "On the hypothesis of no interaction in a multiway contingency table," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 749–757.

[12] S. N. Roy and S. K. Mitra, "An introduction to some nonparametric generalizations of analysis of variance and multivariate analysis," *Biometrika*, Vol. 43 (1956), pp. 361–376.