The Annals of Statistics
1984, Vol. 12, No. 1, 46–60

ON CHI-SQUARED TESTS FOR MULTIWAY CONTINGENCY TABLES WITH CELL PROPORTIONS ESTIMATED FROM SURVEY DATA

BY J. N. K. RAO AND A. J. SCOTT

Carleton University and University of Auckland

The impact of survey design on standard multinomial-based methods for a multiway contingency table is studied, under nested loglinear models. The asymptotic null distribution of the Pearson chi-squared test statistic, X^2 (or the likelihood ratio test statistic, G^2) is obtained as a weighted sum of independent χ_1^2 random variables, and the weights are then related to the familiar design effects (deffs) used by survey samplers. A simple correction to X^2 (or G^2) is also obtained which requires the knowledge of only the cell deffs and the deffs for collapsed tables (marginals), whenever the model admits a direct solution of likelihood equations under multinomial sampling. Finally, an example on the relative performance of X^2 and some corrected X^2 statistics in a three-way table is given, using some data from the Canada Health Survey, 1978–1979.

1. Introduction. Statistical methods for analysing cross-classified categorical data have been extensively developed under the assumption of multinomial sampling (e.g., Bishop, Fienberg and Holland (abbreviated BFH), 1975), utilizing hierarchical loglinear models. Computer packages are also readily available. Researchers in subject matter areas (e.g., social, behavioural and health sciences) have been using these methods to analyse data from complex sample surveys, but most of the commonly used survey designs involve clustering and stratification and hence the multinomial assumption is violated. Nathan (1975), Koch *et al.* (1975), Shuster and Downing (1976) and others have developed asymptotically valid methods, based on the Wald statistic, which take the survey design into account. However, in secondary analyses from published reports containing multiway tables, the researcher may not have access to the necessary information (e.g. the full estimated covariance matrix of cell estimates) for implementing these methods. At best the reports might contain some information about variance estimates (design effects) for marginal totals or cells. Consequently, it is of importance to assess the impact of survey design on standard multinomial-based methods and suggest simple corrections requiring only minimal information on the design effects (abbreviated "deffs"). Even when the necessary information is available, it is not clear that methods based on the Wald statistic would necessarily perform well in finite samples, especially when the number of cells in the table increases leading to unstable sample estimates of the covariance matrix (Fay, 1979). It would be desirable to obtain improved corrections to standard

Received March 1983; revised July 1983.
AMS 1980 *subject classifications*. Primary 62D05; secondary 62H15.
Key words and phrases. Multiway contingency tables, chi-squared and loglikelihood ratio tests, effect of survey design, Wald statistic.

methods utilizing the detailed information and study their finite sample properties relative to those of the Wald statistic and others.

Cohen (1976), Altham (1976), Fellegi (1980), Brier (1980), Rao and Scott (1981) and others have shown that clustering can have a substantial effect on the distribution of the standard Pearson chi-squared test statistic, X^2 and that some adjustment to X^2 may be necessary, without which one can get misleading results in practice. Rao and Scott (1981) and Scott and Rao (1981) have shown that the X^2 statistics for testing goodness of fit (simple hypothesis), independence in a two-way table and homogeneity of proportions across populations are asymptotically distributed as weighted sums of independent χ_1^2 random variables with weights related to particular deffs. They developed a simple correction to X^2 which requires only the knowledge of deffs (or variance estimates) for individual cells in the goodness of fit problem and the deffs of proportions in each population in the homogeneity problem.

In this article we propose to generalize the previous results of Rao and Scott to multiway tables. In Section 2, the asymptotic distribution of X^2 (or equivalently G^2, the likelihood ratio test statistic) is obtained as a weighted sum of independent χ_1^2 random variables under nested loglinear models M_1 and M_2, where M_2 is nested in M_1. The important special case of a saturated model M_1 is also investigated. A simple correction to X^2 (or G^2) is obtained in Section 3 which requires only the cell deffs and the deffs of collapsed tables (marginals) whenever the likelihood equations under multinomial sampling admit explicit solutions. Finally, an example on the relative performance of X^2 and some corrected X^2 statistics in a three-way table is given in Section 4, utilizing some data from the Canada Health Survey, 1978–1979.

2. Asymptotic distribution of X^2 (or G^2). Let $\boldsymbol{\pi}$ denote the T-vector of population cell properties, π_t, in a multiway table ($\Sigma\pi_t = 1$, $t = 1, \cdots, T$). Suppose a sample, $\tilde{s}$, of n ultimate units is drawn according to a specified survey design, $p(\tilde{s})$, and let $\hat{\mathbf{p}}$ denote a consistent estimator of $\boldsymbol{\pi}$ under $p(\tilde{s})$, $\Sigma\hat{p}_t = 1$. Typically, $\hat{p}_t$ is a ratio estimator depending on survey design weights. We assume that a central limit theorem for the specified design is available which ensures that $\sqrt{n}(\hat{\mathbf{p}} - \boldsymbol{\pi})$ converges in distribution to a $N_T(\mathbf{0}, \mathbf{V})$ random vector, say $\mathbf{Y}$, as $n \to \infty$, i.e. $\hat{\mathbf{p}}$ is approximately T-variate normal with mean vector $\mathbf{0}$ and singular covariance matrix $\mathbf{V}/n$ for sufficiently large n (see Rao and Scott, 1981 for the literature on central limit theorems for various survey designs). In the case of multinomial sampling $\mathbf{V}$ reduces to $\mathbf{P} = \mathbf{D}_\pi - \boldsymbol{\pi}\boldsymbol{\pi}'$ where $\mathbf{D}_\pi = \text{diag}(\boldsymbol{\pi})$.

The following algebraic results are useful in our derivations:

LEMMA 1.

(1) $\mathbf{1}'\boldsymbol{\pi} = \mathbf{1}'\hat{\mathbf{p}} = 1$;

(2) $\mathbf{D}_\pi\mathbf{1} = \boldsymbol{\pi}$, $\mathbf{D}_\pi^{-1}\boldsymbol{\pi} = \mathbf{1}$;

(3) $\mathbf{V1} = \mathbf{VD}_\pi^{-1}\boldsymbol{\pi} = \mathbf{0}$. *In particular,* $\mathbf{P1} = \mathbf{PD}_\pi^{-1}\boldsymbol{\pi} = \mathbf{0}$;

(4) $\mathbf{PD}_\pi^{-1}\mathbf{V} = \mathbf{VD}_\pi^{-1}\mathbf{P} = \mathbf{V}$. *In particular,* $\mathbf{PD}_\pi^{-1}\mathbf{P} = \mathbf{P}$ *so that* $\mathbf{D}_\pi^{-1}$ *is a generalized inverse of* $\mathbf{P}$.

In general, a loglinear model on the π_t's may be written as

$$\mu = \tilde{u}(\theta)\mathbf{1} + \mathbf{X}\theta \tag{1}$$

where μ is the T-vector of log probabilities $\mu_t = \ln \pi_t$, $\mathbf{X}$ is a known $T \times r$ matrix of full rank $r(\leq T-1)$ and $\mathbf{X}'\mathbf{1} = \mathbf{0}$, θ is an r-vector of parameters, $\mathbf{1}$ is the T-vector of 1's and $\tilde{u}(\theta) = \ln \{1/[\mathbf{1}'\exp(\mathbf{X}\theta)]\}$ is the normalizing factor where $\exp(\mathbf{X}\theta)$ is the vector of exponential functions $\exp(\mathbf{x}_t'\theta)$, $t = 1, \cdots, T$ and $\mathbf{X}' = (\mathbf{x}_1, \cdots, \mathbf{x}_T)$. If $r = T-1$ in (1) we get the general (or saturated) loglinear model. For instance, the general loglinear model for a $2 \times 2 \times 2$ table may be written as (see BFH, 1975, page 33):

$$\mu_{ijk} = \tilde{u} + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)} \qquad i, j, k = 1, 2 \tag{2}$$

where the sum of any u-term over each of its subscripts is zero, i.e. $u_{1(2)} = -u_{1(1)}$, $u_{12(21)} = -u_{12(11)}$ etc. Here $\theta = (u_{1(1)}, u_{2(1)}, u_{3(1)}, u_{12(11)}, u_{13(11)}, u_{23(11)}, u_{123(111)})'$ and $\mathbf{X}$ is an 8×7 matrix consisting of 1's and -1's. Under the hypothesis of no three factor interaction $H: u_{123(111)} = 0$, for example, $\mathbf{X}$ reduces to an 8×6 matrix $\mathbf{X}_1$ (say) and the deleted column is given by $\mathbf{X}_2 = (1, -1, -1, 1, -1, 1, 1, -1)'$ provided μ is ordered as $(\mu_{111}, \mu_{112}, \mu_{121}, \mu_{122}, \mu_{211}, \mu_{212}, \mu_{221}, \mu_{222})'$.

Under multinomial sampling it is well-known that the likelihood equations are given by

$$\mathbf{X}'\hat{\pi} = \mathbf{X}'(\mathbf{n}/n) \tag{3}$$

where $\mathbf{n}$ is the T-vector of observed frequencies, n_t, in the sample $(\Sigma n_t = n)$ and $\hat{\pi} = \pi(\hat{\theta})$ is the maximum likelihood estimator (m.l.e.) of π under the model $(\Sigma \hat{\pi}_t = 1)$. The method of Iterative Proportional Fitting (IPF) is often used to determine $\hat{\pi}$ from (3) directly (without evaluating the m.l.e. $\hat{\theta}$ of θ) whenever (3) does not admit an explicit solution. The m.l.e. $\hat{\pi}$ are easily obtained for hierarchical models.

For general survey designs, we do not have a m.l.e. of π, due to difficulties in obtaining appropriate likelihood functions. Hence we use a "pseudo m.l.e.," $\hat{\pi}$ obtained from (3) by replacing $\mathbf{n}/n$ by $\hat{\mathbf{p}}$ (Imrey *et al.*, 1982). The consistency of $\hat{\mathbf{p}}$ ensures the consistency of $\hat{\pi}$, under standard regularity conditions.

2.1 *Asymptotic covariance matrix of* $\hat{\pi}$. To obtain the asymptotic covariance matrix of $\hat{\pi}$, we need the following result.

LEMMA 2. *Under the regularity conditions given in Section* 14.8.1 *of* BFH (1975),

$$\hat{\theta} - \theta \sim (\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'(\hat{\mathbf{p}} - \pi) \tag{4}$$

and

$$\hat{\pi} - \pi \sim \mathbf{P}\mathbf{X}(\hat{\theta} - \theta) \tag{5}$$

where "~" denotes "asymptotic equivalence".

PROOF. From BFH (1975), page 517, we have

$$\hat{\theta} - \theta \sim (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{D}_{\pi}^{-1/2}(\hat{\mathbf{p}} - \pi)$$

and

$$\hat{\pi} - \pi \sim D_{\pi}^{1/2}\mathbf{A}(\hat{\theta} - \theta)$$

where $\mathbf{A}$ is the $T \times r$ matrix whose (i, j)th element is $\pi_i^{-1/2}(\partial\pi_i/\partial\theta_j)$. Now, under the loglinear model (1) we have $\mathbf{A} = \mathbf{D}_{\pi}^{-1/2}\mathbf{PX}$, $\mathbf{A}'\mathbf{A} = \mathbf{X}'\mathbf{PX}$ and $\mathbf{A}'\mathbf{D}_{\pi}^{-1/2}(\hat{\mathbf{p}} - \pi) = \mathbf{X}'(\hat{\mathbf{p}} - \pi)$ using Lemma 1.

Since the asymptotic covariance matrix of $\hat{p}$ is $\mathbf{D}(\hat{\mathbf{p}}) = \mathbf{V}/n$, we get from (4) the asymptotic covariance matrix of $\hat{\theta}$:

$$\mathbf{D}(\hat{\theta}) = (1/n)(\mathbf{X}'\mathbf{PX})^{-1}(\mathbf{X}'\mathbf{VX})(\mathbf{X}'\mathbf{PX})^{-1}. \tag{6}$$

Hence the asymptotic covariance matrix of $\hat{\pi}$ is

$$\mathbf{D}(\hat{\pi}) = \mathbf{PXD}(\hat{\theta})\mathbf{X}'\mathbf{P}. \tag{7}$$

Imrey *et al.* (1982) derived (6) and (7) by the δ-method, using implicit differentiation. In the special case of multinomial sampling, we have $\mathbf{V} = \mathbf{P}$ and (6) reduces to the well-known result $\mathbf{D}(\hat{\theta}) = (\mathbf{X}'\mathbf{PX})^{-1}/n$ (see e.g., Fienberg, 1980, page 170).

The asymptotic covariance matrix of "residuals" $\hat{\mathbf{p}} - \hat{\pi}$ is obtained by noting that $\hat{\mathbf{p}} - \hat{\pi} \sim [\mathbf{I} - \mathbf{PX}(\mathbf{X}'\mathbf{PX})^{-1}\mathbf{X}'](\hat{\mathbf{p}} - \pi)$ so that

$$\mathbf{D}(\hat{\mathbf{p}} - \hat{\pi}) = n^{-1}[\mathbf{I} - \mathbf{PX}(\mathbf{X}'\mathbf{PX})^{-1}\mathbf{X}']\mathbf{V}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{PX})^{-1}\mathbf{X}'\mathbf{P}]. \tag{8}$$

If $\mathbf{V} = \mathbf{P}$, (8) reduces to $n^{-1}[\mathbf{P} - \mathbf{PX}(\mathbf{X}'\mathbf{PX})^{-1}\mathbf{X}'\mathbf{P}]$. The diagonal elements of (8) provide the asymptotic variances of the residuals $\hat{p}_t - \hat{\pi}_t$ and hence the standardized residuals which are useful in detecting model deviations.

2.2 *Nested models.* Denoting the model (1) by M_1, let $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ and $\theta = \binom{\theta_1}{\theta_2}$, where $\mathbf{X}_1$ is $T \times s$ and $\mathbf{X}_2$ is $T \times u$ and correspondingly θ_1 is $s \times 1$ and θ_2 is $u \times 1$ ($s + u = r$, $\mathbf{X}_1'\mathbf{1} = \mathbf{0}$, $\mathbf{X}_2'\mathbf{1} = \mathbf{0}$). We are interested in testing the null hypothesis H: $\theta_2 = \mathbf{0}$ so that under H we get the reduced model M_2:

$$\mu = \tilde{u}\begin{pmatrix}\theta_1\\ \mathbf{0}\end{pmatrix}\mathbf{1} + \mathbf{X}_1\theta_1 \tag{9}$$

i.e., M_2 is nested in M_1. Let $\hat{\hat{\theta}}_1$ and $\hat{\hat{\pi}} = \pi(\hat{\hat{\theta}}_1)$ denote the "pseudo m.l.e." of θ_1 and π respectively, under M_2 obtained from the likelihood equations $\mathbf{X}_1'\pi(\hat{\hat{\theta}}_1) = \mathbf{X}_1'\hat{\mathbf{p}}$. The consistency of $\hat{\mathbf{p}}$ ensures that of $\hat{\hat{\pi}}$ under M_2 ($\Sigma\hat{\hat{\pi}}_t = 1$).

The Pearson chi-squared test statistic is given by

$$X^2 = n\ \Sigma(\hat{\pi}_t - \hat{\hat{\pi}}_t)^2/\hat{\hat{\pi}}_t = n(\hat{\pi} - \hat{\hat{\pi}})'\mathbf{D}_{\hat{\hat{\pi}}}^{-1}(\hat{\pi} - \hat{\hat{\pi}}). \tag{10}$$

Alternatively, the likelihood ratio statistic

$$G^2 = 2n\ \Sigma\hat{\pi}_t \ln(\hat{\pi}_t/\hat{\hat{\pi}}_t) = 2n\Sigma\hat{p}_t \ln(\hat{\pi}_t/\hat{\hat{\pi}}_t)$$

$$(11) \qquad = 2n\Sigma\hat{p}_t \ln(\hat{p}_t/\hat{\hat{\pi}}_t) - 2n\Sigma\hat{p}_t \ln(\hat{p}_t/\hat{\pi}_t)$$

$$= G_2^2 - G_1^2 \quad \text{(say)}$$

is used to test H. The statistics X^2 and G^2 are asymptotically equivalent – the proof follows along the lines of Lemma 14.9–1, BFH (1975). For multinomial sampling, it is well-known that X^2 (or G^2) $\approx \chi_u^2$ under H, where "$\approx$" denotes "asymptotically distributed as" (see e.g., Theorem 14.9-8, BFH, 1975). The same asymptotic null distribution holds under a product multinomial sampling scheme which arises with stratified simple random sampling when the strata correspond to levels of one dimension of the contingency table. This will not remain true, however, with more complex survey designs involving clustering or stratification based on variables different from those corresponding to the contingency table.

We now derive the asymptotic null distribution of X^2 (or G^2) for any survey design, $p(\tilde{s})$.

THEOREM 1. *Under* $H: \boldsymbol{\theta}_2 = \mathbf{0}$,

$$(12) \qquad X^2 \sim \hat{\boldsymbol{\theta}}_2'(\tilde{\mathbf{X}}_2'\mathbf{P}\tilde{\mathbf{X}}_2)\hat{\boldsymbol{\theta}}_2$$

where $\hat{\boldsymbol{\theta}} = \begin{pmatrix}\hat{\boldsymbol{\theta}}_1\\ \hat{\boldsymbol{\theta}}_2\end{pmatrix}$ *and*

$$(13) \qquad \tilde{\mathbf{X}}_2 = (\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{P}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{P})\mathbf{X}_2$$

is the projection of $\mathbf{X}_2$ *on the orthogonal complement of the space spanned by the columns of* $\mathbf{X}_1$, *where the inner product is defined with respect to* $\mathbf{P}(\tilde{\mathbf{X}}_2'\mathbf{P}\mathbf{X}_1 = \mathbf{0})$. *Moreover, under H*

$$(14) \qquad X^2 \approx \textstyle\sum_{i=1}^{u} \delta_i W_i$$

where the W_i*'s are independent* χ_1^2 *random variables and the* δ_i*'s* (**all** *greater than zero) are the eigenvalues of the matrix* $(\tilde{\mathbf{X}}_2'\mathbf{P}\tilde{\mathbf{X}}_2)^{-1}(\tilde{\mathbf{X}}_2'\mathbf{V}\tilde{\mathbf{X}}_2)$.

PROOF. Using (5) for $\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}$ and the analogous result $\hat{\hat{\boldsymbol{\pi}}} - \boldsymbol{\pi} \sim \mathbf{P}\mathbf{X}_1(\hat{\hat{\boldsymbol{\theta}}}_1 - \boldsymbol{\theta}_1)$ for $\hat{\hat{\boldsymbol{\pi}}} - \boldsymbol{\pi}$ we get

$$\hat{\boldsymbol{\pi}} - \hat{\hat{\boldsymbol{\pi}}} \sim \mathbf{P}[\mathbf{X}_1(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) + \mathbf{X}_2\hat{\boldsymbol{\theta}}_2 - \mathbf{X}_1(\hat{\hat{\boldsymbol{\theta}}}_1 - \boldsymbol{\theta}_1)]$$

where

$$(15) \qquad \hat{\hat{\boldsymbol{\theta}}}_1 - \boldsymbol{\theta}_1 \sim (\mathbf{X}_1'\mathbf{P}\mathbf{X}_1)^{-1}\mathbf{X}_1'(\hat{\mathbf{p}} - \boldsymbol{\pi}).$$

Now, expressing $\mathbf{X}'\mathbf{P}\mathbf{X}$ as the partitioned matrix

$$\mathbf{X}'\mathbf{P}\mathbf{X} = \begin{pmatrix}\mathbf{X}_1'\mathbf{P}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{P}\mathbf{X}_2\\ \mathbf{X}_2'\mathbf{P}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{P}\mathbf{X}_2\end{pmatrix}$$

and using the standard formula for inverse of a partitioned matrix (e.g., Seber, 1977, page 390), we get from (4) and (15)

$$\hat{\hat{\theta}}_1 - \theta_1 \sim (\hat{\theta}_1 - \theta_1) + (\mathbf{X}_1'\mathbf{P}\mathbf{X}_1)^{-1}(\mathbf{X}_1'\mathbf{P}\mathbf{X}_2)\hat{\theta}_2.$$

Hence

$$\hat{\pi} - \hat{\hat{\pi}} \sim \mathbf{P}\tilde{\mathbf{X}}_2\hat{\theta}_2$$

and

$$X^2 \sim n(\hat{\pi} - \hat{\hat{\pi}})'\mathbf{D}_\pi^{-1}(\hat{\pi} - \hat{\hat{\pi}}) \sim n\hat{\theta}_2'(\tilde{\mathbf{X}}_2'\mathbf{P}\tilde{\mathbf{X}}_2)\hat{\theta}_2$$

noting that $\mathbf{P}\mathbf{D}_\pi^{-1}\mathbf{P} = \mathbf{P}$. Also it follows from (6) and the formula for the inverse of a partitioned matrix that

$$\mathbf{D}(\hat{\theta}_2) = (1/n)(\tilde{\mathbf{X}}_2'\mathbf{P}\tilde{\mathbf{X}}_2)^{-1}(\tilde{\mathbf{X}}_2'\mathbf{V}\tilde{\mathbf{X}}_2)(\tilde{\mathbf{X}}_2'\mathbf{P}\tilde{\mathbf{X}}_2)^{-1}. \tag{16}$$

Hence $\hat{\theta}_2 \approx N_u(\mathbf{0}, \mathbf{D}(\hat{\theta}_2))$ and using the standard result on the distribution of quadratic forms in normal variables (e.g., Johnson and Kotz, 1970, page 150), we get $X^2 \approx \sum \delta_i W_i$ where the δ_i's are eigenvalues of $D(\hat{\theta}_2)(\tilde{\mathbf{X}}_2'\mathbf{P}\tilde{\mathbf{X}}_2)n = (\tilde{\mathbf{X}}_2'\mathbf{P}\tilde{\mathbf{X}}_2)^{-1}(\tilde{\mathbf{X}}_2'\mathbf{V}\tilde{\mathbf{X}}_2)$ and W_i's independent χ_1^2 random variables.

For multinomial sampling, $\mathbf{V} = \mathbf{P}$ so that $\delta_i = 1$, $i = 1, \cdots, u$ and hence we get the standard result $X^2 \approx \chi_u^2$ under H, as a special case of Theorem 1.

Fay (1979) has obtained an alternative representation of the asymptotic distribution of X^2 (or G^2) in which the δ_i's are nonzero eigenvalues of a singular matrix, unlike the nonsingular matrix $(\tilde{\mathbf{X}}_2'\mathbf{P}\tilde{\mathbf{X}}_2)^{-1}(\tilde{\mathbf{X}}_2'\mathbf{V}\tilde{\mathbf{X}}_2)$ in our Theorem 1. Foutz and Srivastava (1978) derived a representation similar to (14) in the context of the asymptotic distribution of the general likelihood ratio statistic (for the iid case) when the model is incorrect.

Note that $\tilde{\mathbf{X}}_2'\hat{\mathbf{p}}$ is a vector of contrasts in the $\hat{p}_i$'s since $\tilde{\mathbf{X}}_2'\mathbf{1} = \mathbf{0}$. The covariance matrix of $\tilde{\mathbf{X}}_2'\hat{\mathbf{p}}$ is $(\tilde{\mathbf{X}}_2'\mathbf{V}\tilde{\mathbf{X}}_2)/n$ for the survey design used, while $(\tilde{\mathbf{X}}_2'\mathbf{P}\tilde{\mathbf{X}}_2)/n$ is the corresponding covariance matrix for multinomial sampling. Thus $\delta_1, \cdots, \delta_u$ are the "generalized design effects" (as defined in Rao and Scott, 1981) for the contrast vector $\tilde{\mathbf{X}}_2'\hat{\mathbf{p}}$; the largest eigenvalue, δ_1 say, is the largest possible deff taken over all linear combinations of the elements of the vector $\tilde{\mathbf{X}}_2'\hat{\mathbf{p}}$.

2.3. *A Wald statistic.* If a consistent estimator, $\hat{\mathbf{V}}/n$, of the covariance matrix of $\hat{\mathbf{p}}$, $\mathbf{V}/n$ is available, one can construct a Wald statistic, X_W^2, which is asymptotically χ_u^2 under $H: \theta_2 = \mathbf{0}$. Let $\mathbf{C}$ be any $T \times u$ matrix of rank u with $\mathbf{C}'\mathbf{X}_1 = \mathbf{0}$, $\mathbf{C}'\mathbf{1} = \mathbf{0}$ and $\mathbf{C}'\mathbf{X}_2$ nonsingular, in particular, if $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$ a convenient choice of $\mathbf{C}$ would be $\mathbf{X}_2$. Then $H: \theta_2 = \mathbf{0}$ is equivalent to $H': \phi = \mathbf{C}'\mu = \mathbf{0}$ and hence a Wald statistic for testing H, based on $\hat{\pi}$, is given by

$$X_W^2 = \hat{\phi}'[\hat{\mathbf{D}}(\hat{\phi})]^{-1}\hat{\phi} \tag{17}$$

where $\hat{\phi} = \mathbf{C}'\hat{\mu}$, $\hat{\mu}$ is the T-vector of logprobabilities $\hat{\mu}_t = \ln \hat{\pi}_t$ and $\hat{\mathbf{D}}(\hat{\phi})$ is the estimated asymptotic covariance matrix of $\hat{\phi}$. Noting that $\hat{\mu} - \mu \sim \mathbf{D}_\pi^{-1}(\hat{\pi} - \pi)$ (by the delta method) we get

$$\mathbf{D}(\hat{\phi}) = \mathbf{C}'\mathbf{D}_\pi^{-1}\mathbf{D}(\hat{\pi})\mathbf{D}_\pi^{-1}\mathbf{C} = \mathbf{\Sigma}_\phi, \quad \text{say} \tag{18}$$

where $\mathbf{D}(\hat{\pi})$ is given by (7). The estimator of $\mathbf{D}(\hat{\phi})$ is obtained by replacing π by

$\hat{\pi}$ and $\mathbf{V}$ by $\hat{\mathbf{V}}$. The Wald statistic (17) is independent of the choice of $\mathbf{C}$. Koch *et al.* (1975) give excellent illustrations of the Wald statistics with survey data.

If no estimator of $\mathbf{V}/n$ is available, the effect of survey design is sometimes ignored, as in the case of X^2 or G^2, and $\mathbf{V}/n$ is replaced by the multinomial covariance matrix $\mathbf{P}/n$ in $\mathbf{D}(\hat{\boldsymbol{\phi}})$. In this case $\mathbf{D}(\hat{\boldsymbol{\phi}})$ reduces to $\boldsymbol{\Sigma}_0 = n^{-1}\mathbf{C}'\mathbf{X}(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}$, using the fact $\mathbf{P}\mathbf{D}_{\pi}^{-1}\mathbf{C} = \mathbf{C}$. This leads to a test statistic alternative to X^2 or G^2:

$$\tilde{\mathbf{X}}_W^2 = n\hat{\boldsymbol{\phi}}'[\mathbf{C}'\mathbf{X}(\mathbf{X}'\hat{\mathbf{P}}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}]^{-1}\hat{\boldsymbol{\phi}}. \tag{19}$$

As in the case of X^2 (or G^2), the true asymptotic null distribution of $\tilde{\mathbf{X}}_W^2$ is a weighted sum of independent χ_1^2 random variables, $\Sigma\gamma_i W_i$, where $\gamma_1, \cdots, \gamma_u$ are the eigenvalues of $\Sigma_0^{-1}\boldsymbol{\Sigma}_{\phi}$. Theorem 2 below shows that $\mathbf{X}_W^2$, in fact, is asymptotically equivalent to X^2 under H which implies that $\mathbf{X}_W^2 \approx \Sigma\delta_i W_i$ and $\{\delta_1, \cdots, \delta_u\}$ is identical to $\{\gamma_1, \cdots, \gamma_u\}$.

THEOREM 2. *Under* $H: \boldsymbol{\theta}_2 = \mathbf{0}$, $\tilde{X}_W^2 \sim X^2$.

PROOF. Under H we have $\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\phi}} - \boldsymbol{\phi} = C'(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \sim \mathbf{C}'\mathbf{D}_{\pi}^{-1}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) \sim \mathbf{C}'\mathbf{D}_{\pi}^{-1}\mathbf{P}\mathbf{X}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ using (5). Hence, noting that $\mathbf{C}'\mathbf{D}_{\pi}^{-1}\mathbf{P}\mathbf{X}_1 = \mathbf{C}'\mathbf{X} = (\mathbf{0}, \mathbf{C}'\mathbf{X}_2)$ and using the formula for the inverse of the partitioned matrix $\mathbf{X}'\mathbf{P}\mathbf{X}$, we get

$$\hat{\boldsymbol{\phi}} \sim \mathbf{C}'\mathbf{X}_2\hat{\boldsymbol{\theta}}_2$$

and

$$\mathbf{C}'\mathbf{X}(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C} = \mathbf{C}'\mathbf{X}_2(\tilde{\mathbf{X}}_2'\mathbf{P}\tilde{\mathbf{X}}_2)^{-1}\mathbf{X}_2'\mathbf{C}.$$

Hence

$$\tilde{X}_W^2 \sim n\hat{\boldsymbol{\theta}}_2'(\tilde{\mathbf{X}}_2'\mathbf{P}\tilde{\mathbf{X}}_2)\hat{\boldsymbol{\theta}}_2 \sim X^2$$

noting that $\mathbf{C}'\mathbf{X}_2$ is nonsingular.

In the Appendix, another (multinomial) Wald statistic, $\tilde{\mathbf{X}}_W^2(1)$, based on a weighted least squares estimator (WLS), $\tilde{\boldsymbol{\theta}}$, of $\boldsymbol{\theta}$ is shown to be asymptotically equivalent to X^2, under H.

2.4. *Special case of saturated model* M_1. In the important special case of saturated model M_1 with $s + u = T - 1$, we can obtain the δ_i's directly from $\mathbf{C}$ without having to calculate the projection matrix $\tilde{\mathbf{X}}_2$. Since $\hat{\boldsymbol{\pi}} = \hat{\mathbf{p}}$ in the saturated case, we have $\mathbf{D}(\hat{\boldsymbol{\pi}}) = \mathbf{V}/n$ under the given survey design and $\mathbf{D}(\hat{\boldsymbol{\pi}}) = \mathbf{P}/n =$ under multinomial sampling. Hence, noting that $\mathbf{C}'\mathbf{D}_{\pi}^{-1}\mathbf{P}\mathbf{D}_{\pi}^{-1}\mathbf{C} = \mathbf{C}'\mathbf{D}_{\pi}^{-1}\mathbf{C}$ we get the result that the δ_i's are eigenvalues of $\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{\phi} = (\mathbf{C}'\mathbf{D}_{\pi}^{-1}\mathbf{C})^{-1}(\mathbf{C}'\mathbf{D}_{\pi}^{-1}\mathbf{V}\mathbf{D}_{\pi}^{-1}\mathbf{C})$.

2.5. *Effect of survey design.* The asymptotic null distribution of X^2 may be approximated by a χ^2 variable, following Satterthwaite (1946): $X_S^2 = X^2/[(1 + a^2)\delta.]$ is treated as χ_ν^2 where $\nu = u/(1 + a^2)$, $\delta. = \Sigma\delta_i/u$ and $a^2 = \Sigma(\delta_i - \delta.)^2/[u\delta.^2]$ is the square of coefficient of variation (c.v.) of the δ_i's. It is, however, not necessary to evaluate the individual δ_i's in order to compute X_S^2, since $\Sigma\delta_i = \bar{E}(X^2)$ and $2\ \Sigma\delta_i^2 = \bar{V}(X^2)$ where $\bar{E}$ and $\bar{V}$ respectively denote the asymptotic expectation and asymptotic variance operators. To evaluate $\bar{E}(X^2)$ we first note that $X^2 \approx \Sigma_1^T Y_t^2/\pi_t$ where $\sqrt{n}(\hat{\pi} - \hat{\hat{\pi}}) \approx \mathbf{Y} =_d N_T(\mathbf{0}, \mathbf{B})$, $\mathbf{B} = (b_{tl}) = \mathbf{P}\tilde{\mathbf{X}}_2\mathbf{D}(\hat{\hat{\theta}}_2)\tilde{\mathbf{X}}_2'\mathbf{P}$ and $\mathbf{D}(\hat{\hat{\theta}}_2)$ is given by (16). Hence $\bar{E}(X^2) = \Sigma b_{tt}/\pi_t$. Similarly, noting that $V(Y_t^2) = 2b_{tt}^2$ and $\text{cov}(Y_t^2, Y_l^2) = 2b_{tl}^2$, $t \neq l = 1, \cdots, T$, we get

$$V(X^2) = 2 \sum_{t=1}^T \sum_{l=1}^T b_{tl}^2/(\pi_t \pi_l).$$

If the estimated full covariance matrix $\hat{\mathbf{V}}/n$ of cell estimates $\hat{\mathbf{p}}$ is available, X_S^2 may be employed as an alternative test statistic to X_W^2, using the estimated $\hat{a}$ and $\hat{\delta}.$ obtained from $\hat{\mathbf{B}}$ in place of unknown a and $\delta.$.

The effect of survey design may be studied by computing the asymptotic significance level of X^2, for a desired nominal level α, i.e.,

$$\text{SL}(X^2) = P[X^2 \geq \chi_u^2(\alpha)] \doteq P[\chi_\nu^2 \geq \{(1 + a^2)\delta.\}^{-1}\chi_u^2(\alpha)]$$

is compared to α, where $\chi_u^2(\alpha)$ is the upper α-point of χ_u^2. In practice, $\text{SL}(X^2)$ is estimated by using $\hat{\mathbf{V}}$ for $\mathbf{V}$ and $\hat{\boldsymbol{\pi}}$ (or $\hat{\hat{\boldsymbol{\pi}}}$) for $\boldsymbol{\pi}$. Holt, Scott and Ewing (1980) report estimated $\text{SL}(X^2)$-values for selected items from the General Household Survey of U.K., for testing simple goodness of fit, independence in a two-way table and homogeneity of proportions across populations. Hidiroglou and Rao (1981) provide simplified formulae for $\mathbf{B}$ for stratified multistage sampling with primary clusters sampled with replacement with probability proportional to size, again for the above three hypotheses (note that M_1 is saturated for testing these hypotheses).

3. Modifications to X^2. In practice it is often adequate to make a first order correction to X^2 (or G^2), i.e. treat $X^2/\hat{\delta}.$ or $G^2/\hat{\delta}.$ as χ_u^2 under H, where $\delta.$ may be written as

$$
\begin{aligned}
u\delta. &= \text{tr}[(\tilde{\mathbf{X}}_2'\mathbf{P}\tilde{\mathbf{X}}_2)^{-1}(\tilde{\mathbf{X}}_2'\mathbf{V}\tilde{\mathbf{X}}_2)] \\
&= \text{tr}[(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}\mathbf{X})] - \text{tr}[(\mathbf{X}_1'\mathbf{P}\mathbf{X}_1)^{-1}(\mathbf{X}_1'\mathbf{V}\mathbf{X}_1)] \qquad (20)\\
&= (s + u)\lambda. - s\lambda_1. \quad \text{(say)}.
\end{aligned}
$$

Note that $(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}\mathbf{X})$ is the design effect matrix for the contrast vector $\mathbf{X}'\hat{\mathbf{p}}$ so that $\lambda.$ is the average generalized deff of $\mathbf{X}'\hat{\mathbf{p}}$. Similarly, $(\mathbf{X}_1'\mathbf{P}\mathbf{X}_1)^{-1}(\mathbf{X}_1'\mathbf{V}\mathbf{X}_1)$ is the deff matrix for the contrast vector $\mathbf{X}_1'\hat{\mathbf{p}}$ and $\lambda_1.$ is the average generalized deff of $\mathbf{X}_1'\hat{\mathbf{p}}$. In the special case of saturated model M_1 ($s + u = T - 1$), $u\delta.$ reduces to

$$(T - s - 1)\delta. = (T - 1)\lambda. - s\lambda_1. \qquad (21)$$

where $(T-1)\lambda. = \Sigma(1 - \pi_t)d_t$ and $d_t = v_{tt}/[\pi_t(1 - \pi_t)]$ is the (cell) deff of $\hat{\mathbf{p}}_t$ and $\mathbf{V} = (v_{tl})$. If $T \gg s$, then $\delta. \doteq \lambda.$ and we might expect $X^2/\hat{\lambda}.$ to perform well in large tables if s is fairly small. Note that $\lambda.$ is independent of H, unlike $\delta.$. Another approximation $X^2/\hat{d}.$ (Fellegi, 1980) where $\hat{d}. = \Sigma\, \hat{d}_t/T$ is the average estimated cell deff, is also independent of H. Empirical results reported in Holt, Scott and Ewing (1980) and Hidiroglou and Rao (1981) for testing independence in a two-way table indicated that both $X^2/\hat{\lambda}.$ and $X^2/\hat{d}.$ tend to be conservative, i.e. their asymptotic significance level $< \alpha$, and sometimes very conservative, whereas $X^2/\hat{\delta}.$ works fairly well especially when the c.v. of $\hat{\delta}_i$'s is small; the Pearson statistic X^2 often leads to unacceptably high values of SL(X^2).

When the model admits explicit solutions for $\hat{\pi}$ and $\hat{\hat{\pi}}$, it is possible to simplify $\delta.$ considerably. We now show that $\delta.$ can be computed knowing only the cell deffs, d_t, and the deffs of collapsed tables (or marginals), whenever the model admits explicit solution. For instance, the no three factor interaction hypothesis is the only hierarchical hypothesis that does not permit explicit solution in a three-way table.

BFH (1975, page 75) list the types of direct estimates (i.e., explicit m.l.e.) possible in four dimensions. Using their notation, a hypothesis leading to direct estimates is of the form

$$\pi_\theta = [\Pi_i\ \pi_{\theta_i}]/[\Pi_j\ \pi_{\phi_j}] \tag{22}$$

where θ denotes an arbitrary cell, π_{θ_i} and π_{ϕ_j} are marginal totals of π_θ where θ_i is a subset of θ corresponding to a "sufficient configuration" (under multinomial sampling) and $\phi_j = \theta_i \cap \theta_t$ for some i and t. In this case, the "pseudo m.l.e." $\hat{\pi}_\theta$ (or $\hat{\hat{\pi}}_\theta$) is given by

$$\hat{\pi}_\theta = [\Pi_i\ \hat{p}_{\theta_i}]/[\Pi_j\ \hat{p}_{\phi_j}] \tag{23}$$

where $\hat{p}_{\theta_i}$ and $\hat{p}_{\phi_j}$ are the marginal totals of $\hat{p}_\theta$ corresponding to θ_i and ϕ_j respectively. For instance in a three-way table, $\theta = \{ijk\}$, $\theta_1 = \{ik\}$, $\theta_2 = \{jk\}$ and $\phi_1 = \{k\}$ for the hypothesis that variables 1 and 2 are conditionally independent given the level of variable 3.

It is simpler, in the case of direct estimates, to obtain the formula for $\delta.$ directly by noting that $u\delta. = \bar{E}X^2 = \bar{E}G^2 = \bar{E}G_2^2 - \bar{E}G_1^2$. Note that $G^2 = 2n\ \Sigma_\theta \hat{p}_\theta \ln(\hat{p}_\theta/\hat{\pi}_\theta)$ in the case of saturated model M_1. Using (22), we have

$$\begin{aligned} \frac{G_1^2}{n} &= 2\ \textstyle\sum_\theta\ \hat{p}_\theta \ln(\hat{p}_\theta/\hat{\pi}_\theta) \\ &= 2\ \textstyle\sum_\theta\ \hat{p}_\theta \ln \hat{p}_\theta - 2\ \sum_i\ [\sum_{\theta_i} \hat{p}_{\theta_i} \ln \hat{p}_{\theta_i}] + 2\ \sum_j\ [\sum_{\phi_j} \hat{p}_{\phi_j} \ln \hat{p}_{\phi_j}]. \end{aligned} \tag{24}$$

Now

$$\begin{aligned} 2\ \textstyle\sum_\theta\ \hat{p}_\theta \ln \hat{p}_\theta &= 2\ \textstyle\sum_\theta\ \hat{p}_\theta \left[\ln \pi_\theta + \ln\left\{1 + \frac{\hat{p}_\theta - \pi_\theta}{\pi_\theta}\right\}\right] \\ &\sim 2\ \textstyle\sum_\theta\ \hat{p}_\theta \left[\ln \pi_\theta + \frac{\hat{p}_\theta - \pi_\theta}{\pi_\theta} - \frac{(\hat{p}_\theta - \pi_\theta)^2}{2\pi_\theta^2}\right] \\ &\sim 2\ \textstyle\sum_\theta\ \hat{p}_\theta \ln \pi_\theta + \sum_\theta (\hat{p}_\theta - \pi_\theta)^2/\pi_\theta \end{aligned} \tag{25}$$

noting that $\hat{p}_\theta = \pi_\theta + (\hat{p}_\theta - \pi_\theta)$ and $\hat{p}_\theta - \pi_\theta$ is of the order $n^{-1/2}$ in probability. Similarly,

$$2\ \textstyle\sum_{\theta_i} \hat{p}_{\theta_i} \ln \hat{p}_{\theta_i} \sim 2\ \sum_{\theta_i} \hat{p}_{\theta_i} \ln \pi_{\theta_i} + \sum_{\theta_i} (\hat{p}_{\theta_i} - \pi_{\theta_i})^2/\pi_{\theta_i} \tag{26}$$

and

(27) $$2\sum_{\phi_j}\hat{p}_{\phi_j}\ln\hat{p}_{\phi_j}\sim 2\sum_{\phi_j}\hat{p}_{\phi_j}\ln\pi_{\phi_j}+\sum_{\phi_j}(\hat{p}_{\phi_j}-\pi_{\phi_j})^2/\pi_{\phi_j}.$$

Using (25), (26) and (27) in (24) and noting that

$$2\sum_\theta \hat{p}_\theta[\ln\pi_\theta-\sum_i\ln\pi_{\theta_i}+\sum_j\ln\pi_{\phi_j}]=0$$

due to (22), we get

$$\frac{G_1^2}{n}\sim\sum_\theta\frac{(\hat{p}_\theta-\pi_\theta)^2}{\pi_\theta}-\sum_i\left\{\sum_{\theta_i}\frac{(\hat{p}_{\theta_i}-\pi_{\theta_i})^2}{\pi_{\theta_i}}\right\}+\sum_j\left\{\sum_{\phi_j}\frac{(\hat{p}_{\phi_j}-\pi_{\phi_j})^2}{\pi_{\phi_j}}\right\}$$

or

(28) $$\bar{E}G_1^2=\sum_\theta(1-\pi_\theta)\,d_\theta-\sum_i\{\sum_{\theta_i}(1-\pi_{\theta_i})\,d_{\theta_i}\}+\sum_j\{\sum_{\phi_j}(1-\pi_{\phi_j})\,d_{\phi_j}\}.$$

Here $d_\theta=V(\hat{p}_\theta)/[\pi_\theta(1-\pi_\theta)/n]$ is the deff of cell estimate $\hat{p}_\theta$ under H and π_θ is given by (22), and $d_{\theta_i}=V(\hat{p}_{\theta_i})/[\pi_{\theta_i}(1-\pi_{\theta_i})/n]$ and $d_{\phi_j}=V(\hat{p}_{\phi_j})/[\pi_{\phi_j}(1-\pi_{\phi_j})/n]$ are the deffs of marginals $\hat{p}_{\theta_i}$ and $\hat{p}_{\theta_j}$ respectively. Similarly, $\bar{E}G_2^2$ is obtained when $\hat{\pi}_\theta$ is of the form (23).

The estimated deff $\tilde{d}_\theta$ of d_θ is given by $\hat{d}_\theta[\hat{p}_\theta(1-\hat{p}_\theta)/\hat{\pi}_\theta(1-\hat{\pi}_\theta)]$ where $\hat{d}_\theta=\hat{V}(\hat{p}_\theta)/[\hat{p}_\theta(1-\hat{p}_\theta)/n]$ is the estimated deff of $\hat{p}_\theta$ irrespective of H, and $\hat{V}(\hat{p}_\theta)$ is the estimate of variance of $\hat{p}_\theta$. It is a common practice to report $\hat{d}_\theta$ rather than $\tilde{d}_\theta$. The estimated deffs $\hat{d}_{\theta_i}$ and $\hat{d}_{\phi_j}$ are given by $\hat{V}(\hat{p}_{\theta_i})/[\hat{p}_{\theta_i}(1-\hat{p}_{\theta_i})/n]$ and $\hat{V}(\hat{p}_{\phi_i})/[\hat{p}_{\phi_i}(1-\hat{p}_{\phi_i})/n]$ respectively. Hence

(28a) $$u\hat{\delta}.=\sum_\theta\frac{\hat{p}_\theta}{\hat{\pi}_\theta}(1-\hat{p}_\theta)\hat{d}_\theta-\sum_i\{\sum_{\theta_i}(1-\hat{p}_{\theta_i})\hat{d}_{\theta_i}\}+\sum_j\{\sum_{\phi_j}(1-\hat{p}_{\phi_j})\hat{d}_{\phi_j}\}.$$

The result (28a) should have important implications for the publication of tables from survey data.

For ready reference, we now give the results for $I\times J$ and $I\times J\times K$ tables when M_1 is the saturated model:

***I* × *J* table:** Independence $(\bar{1}\otimes\bar{2})$: $\pi_{ij}=\pi_{i+}\pi_{+j}\Leftrightarrow u_{12(ij)}=0$, $i=1,\cdots,I$; $j=1,\cdots,J$.

(29) $$\begin{aligned}&(I-1)(J-1)\delta.\\&=\sum_i\sum_j(1-\pi_{i+}\pi_{+j})\,d_{ij}-\sum_i(1-\pi_{i+})\,d_i(r)-\sum_j(1-\pi_{+j})\,d_j(c)\end{aligned}$$

where π_{i+} and π_{+j} are the row and column marginals, $d_i(r)$ and $d_j(c)$ are the deffs of $\hat{p}_{i+}$ and $\hat{p}_{+j}$ respectively, and d_{ij} is the deff of $\hat{p}_{ij}$. Hence, $X^2/\hat{\delta}.$ can be computed knowing only the cell proportions $\hat{p}_{ij}$ and their estimated deffs $\hat{d}_{ij}$, and the estimated deffs $\hat{d}_i(r)$ and $\hat{d}_j(c)$ of the one-way marginals $\hat{p}_{i+}$ and $\hat{p}_{+j}$.

Bedrick (1983) has independently obtained (29) using an alternative derivation, and stated the general result (28).

***I* × *J* × *K* table:** (a) complete independence $(\bar{1}\otimes\bar{2}\otimes\bar{3})$: $\pi_{ijk}=\pi_{i++}\pi_{+j+}\pi_{++k}$

$\Leftrightarrow u_{123(ijk)} = u_{12(ij)} = u_{13(ik)} = u_{23(jk)} = 0, i = 1, \cdots, I; j = 1, \cdots, J; k = 1, \cdots, K.$

$$(IJK - I - J - K + 2)\delta.$$

$$= \sum_i \sum_j \sum_k (1 - \pi_{i++}\pi_{+j+}\pi_{++k})\, d_{ijk} - \sum_i (1 - \pi_{i++})\, d_i(r) \quad (30)$$

$$- \sum_j (1 - \pi_{+j+})\, d_j(c) - \sum_k (1 - \pi_{++k})\, d_k(l)$$

where π_{i++}, π_{+j+} and π_{++k} are the three-way marginals and $d_i(r)$, $d_j(c)$ and $d_k(l)$ are the corresponding marginal deffs and d_{ijk} is the deff of $\hat{p}_{ijk}$.

(b) Independence of variable 1 from variables (2, 3) jointly ($\overline{1} \otimes \overline{23}$): $\pi_{ijk} = \pi_{i++}\pi_{+jk} \Leftrightarrow u_{123(ijk)} = u_{12(ij)} = u_{13(ik)} = 0$ (similarly $\overline{2} \otimes \overline{13}$ and $\overline{3} \otimes \overline{12}$)

$$(I - 1)(JK - 1)\delta.$$

$$= \sum_i \sum_j \sum_k (1 - \pi_{i++}\pi_{+jk})\, d_{ijk} \quad (31)$$

$$- \sum_i (1 - \pi_{i++})\, d_i(r) - \sum_j \sum_k (1 - \pi_{+jk})\, d_{jk}(c, l)$$

where π_{+jk}'s are the two-way marginals when collapsed over variable 1 with corresponding marginal deffs $d_{jk}(c, l)$.

(c) Conditional independence ($\overline{1} \otimes \overline{2} \mid \overline{3}$): $\pi_{ijk} = (\pi_{i+k}\pi_{+jk})/\pi_{++k} \Leftrightarrow u_{123(ijk)} = u_{12(ij)} = 0$. (Similarly $\overline{2} \otimes \overline{3} \mid \overline{1}$ and $\overline{1} \otimes \overline{3} \mid \overline{2}$)

$$K(I - 1)(J - 1)\delta. = \sum_i \sum_j \sum_k \left(1 - \frac{\pi_{i+k}\pi_{+jk}}{\pi_{++k}}\right) d_{ijk}$$

$$- \sum_i \sum_k (1 - \pi_{i+k})\, d_{ik}(r, l) \quad (32)$$

$$- \sum_j \sum_k (1 - \pi_{+jk})\, d_{jk}(c, l) + \sum_k (1 - \pi_{++k})\, d_k(l).$$

Hence, for the case of a three-way table one can compute the corrected statistic $X^2/\hat{\delta}$ for the hypotheses (a) – (c) knowing only the three-way table of proportions $\hat{p}_{ijk}$ and their estimated deffs $\hat{d}_{ijk}$, estimated deffs $\hat{d}_{ij}(r, c)$, $\hat{d}_{jk}(c, l)$ and $\hat{d}_{ik}(r, l)$ of the two-way marginals $\hat{p}_{ij+}$, $\hat{p}_{+jk}$ and $\hat{p}_{i+k}$, and the estimated deffs $\hat{d}_i(r)$, $\hat{d}_j(c)$ and $\hat{d}_k(l)$ of the one-way marginals $\hat{p}_{i++}$, $\hat{p}_{+j+}$ and $\hat{p}_{++k}$.

When the model does not permit explicit solution for $\hat{\pi}$ or $\hat{\hat{\pi}}$, $\delta.$ cannot be expressed in terms of only the cell deffs and marginal deffs. For instance, in the $2 \times 2 \times 2$ case and the no three-factor interaction hypothesis $H: u_{123(ijk)} = 0$ for all (i, j, k) using $\mathbf{C} = \mathbf{X}_2 = (1, -1, -1, 1, -1, 1, 1, -1)'$, in the equivalent hypothesis $H': \mathbf{C}'\mu = (\mu_{111} - \mu_{121} + \mu_{221} - \mu_{211}) - (\mu_{112} - \mu_{122} + \mu_{222} - \mu_{212}) = \psi_1 - \psi_2 = 0$ where ψ_k is the log cross-product ratio in the kth 2×2 layer ($k = 1, 2$), we get from Section 2.3

$$\delta. = \delta_1 = (\mathbf{C}'\mathbf{D}_\pi^{-1}\mathbf{V}\mathbf{D}_\pi^{-1}\mathbf{C})/(\mathbf{C}'\mathbf{D}_\pi^{-1}\mathbf{C})$$

$$= \left(\sum_i \sum_j \sum_k \frac{1}{\pi_{ijk}}\right)^{-1} \{\textstyle\sum_{a=1}^{8} \sum_{b=1}^{8} v_{(a)(b)}\pi_{(a)}^{-1}\pi_{(b)}^{-1}(-1)^{a+b}\}$$

where we labeled the cells (ijk) as follows: (1) = (111), (2) = (121), (3) = (221), (4) = (211), (5) = (122), (6) = (112), (7) = (212) and (8) = (222). Since δ_1 involves all the covariances $v_{(a)(b)}$ it cannot be expressed in terms of only the cell deffs and marginal deffs. When the hypothesis does not permit explicit solution for $\hat{\pi}$ or $\hat{\hat{\pi}}$ and only the cell deffs and marginal deffs are known, we can take the $\delta.$ corresponding to a hypothesis "closest" to H (in the sense of having approximately the same u-terms in the model) and admitting explicit solution. For instance, to test $H: u_{123(ijk)} = 0$, we can take $\delta.$ corresponding to a conditional independence hypothesis: $u_{123(ijk)} = u_{12(ij)} = 0$.

The general result (28) also covers the case of one set of fixed marginal totals of the observed frequencies, n_t, in the sample, provided the u-terms corresponding to fixed margins are included in the model (1). However, the published tables usually report estimated cell proportions and their marginal totals within each population corresponding to fixed margins, and the associated deffs. Therefore, we need to express $\delta.$ in terms of the within-population cell proportions and associated deffs. We now derive such a formula for the $I \times J$ table when M_1 is a saturated model. The corresponding results for three-way and higher dimensional tables will be reported in a subsequent paper.

***I* × *J* table:** Let n_{ij} be the observed sample frequency in (i, j)th cell and let the row margins $n_{i+} = \sum_j n_{ij}$ be fixed. We take $\hat{p}_{ij} = (n_{i+}/n)\hat{p}_{j(i)}$ and $\pi_{ij} = (n_{i+}/n)P_{j(i)}$ where $P_{j(i)}$ are the cell proportions within the ith row population and $\hat{p}_{j(i)}$ are the corresponding survey estimates ($\sum_j P_{j(i)} = \sum_j \hat{p}_{j(i)} = 1$). Hence $H: \pi_{ij} = \pi_{i+}\pi_{+j} \Leftrightarrow P_{j(i)} = \sum_i (n_{i+}/n)P_{j(i)} = P_j$ (say) which is the test of homogeneity of proportions across populations.

Since $\hat{p}_{i+} = n_{i+}/n$ we get $V(\hat{p}_{i+}) = 0$ and hence

$$(1 - \pi_{i+})\, d_i(r) = nV(\hat{p}_{i+})/\pi_{i+} = 0. \tag{33}$$

Also

$$(1 - \pi_{i+}\pi_{+j})\, d_{ij} = nV(\hat{p}_{ij})/(\pi_{i+}\pi_{+j}) = (1 - P_j)\, d_{j(i)} \tag{34}$$

where

$$d_{j(i)} = n_{i+} V[\hat{p}_{j(i)}]/P_j(1 - P_j)$$

is the jth cell deff in the i-th row population, under H. Similarly,

$$(1 - \pi_{+j})\, d_j(c) = nV(\hat{p}_{+j})/\pi_{+j} = (\textstyle\sum_i (n_{i+}/n)\, d_{j(i)})(1 - P_j) \tag{35}$$

noting that $\hat{p}_{+j} = \sum_i (n_{i+}/n)\hat{p}_{j(i)}$ and $nV(\hat{p}_{+j}) = \sum_i (n_{i+}/n)\, d_{j(i)}P_j(1 - P_j)$, assuming that sampling is done independently within each row population. Hence, using (33), (34) and (35) in (28) we get

$$(I - 1)(J - 1)\delta. = \sum\sum(1 - P_j)\, d_{j(i)}(1 - n_{i+}/n) \tag{36}$$

which agrees with the result given by Scott and Rao (1981). Thus $\delta.$ depends only on the estimated cell deffs within each row population.

The formula for $\delta.$ in the case of fixed column margins n_{+j} is obtained from (36) by interchanging the subscripts i and j.

4. Example. We now provide an example on the relative performance of X^2, $X^2/\hat{\delta}.$, $X^2/\hat{\lambda}.$ and $X^2/\hat{d}.$ in a three-way table, utilizing some data from the

TABLE 1.

Estimated Asymptotic Significance Levels (SL) *of* X^2 *and the Corrected Statistics* $X^2/\hat{\delta}.$, $X^2/\hat{\lambda}.$, $X^2/\hat{d}.$: $2 \times 5 \times 4$ *table and nominal level* $\alpha = 0.05$.

Hypothesis

	(a) $\bar{1}\otimes\bar{2}\otimes\bar{3}$	$\bar{1}\otimes\overline{23}$	(b) $\bar{2}\otimes\overline{13}$	$\bar{3}\otimes\overline{12}$	$\bar{1}\otimes\bar{2}\mid\bar{3}$	(c) $\bar{1}\otimes\bar{3}\mid\bar{2}$	$\bar{2}\otimes\bar{3}\mid\bar{1}$
SL(X^2):	0.72	0.33	0.76	0.72	0.43	0.30	0.78
SL($X^2/\hat{\delta}.$)	0.16	0.11	0.14	0.13	0.095	0.11	0.12
SL($X^2/\hat{\lambda}.$)	0.34	0.056	0.39	0.32	0.098	0.06	0.39
SL($X^2/\hat{d}.$)	0.34	0.054	0.39	0.32	0.097	0.06	0.39
$\hat{\delta}$	2.09	1.40	2.25	2.09	1.63	1.39	2.31
C.V.($\hat{\delta}_i$)	1.54	1.02	1.37	1.27	0.86	1.05	1.11

Canada Health Survey (CHS), 1978–1979. The CHS was designed to provide reliable information on the health status of Canadians. A complex multistage design involving stratification and cluster sampling was employed and the estimates of totals and proportions were subjected to post-stratification adjustment on age-sex, to improve their efficiency. We consider the three hypotheses (a), (b) and (c) leading to explicit solution for m.l.e. Hidiroglou and Rao (1983) provide simplified formulae for the estimated covariance matrix **B** under the above three hypotheses. Table 1, taken from Hidiroglou and Rao (1983), gives the estimated asymptotical significance levels ($\alpha = .05$) of X^2 and the three corrected statistics for a $2 \times 5 \times 4$ table with the following variables: (1) sex (male, female), (2) drug use (0, 1, 2, 3, 4+ drugs in a 2-day period), (3) age group (0–14, 15–44, 45–64, 65+). Here $n = 31{,}668$ and $\hat{\lambda}. = 1.614$, $\hat{d}. = 1.615$. The Satterthwaite approximation was used in computing the SL-values reported in Table 1.

It is clear from Table 1 that SL(X^2) is unacceptably high, ranging from 0.30 to 0.78 whereas $\alpha = 0.05$. Hence the effect of survey design on X^2 is severe. The corrected statistics $X^2/\hat{\lambda}.$ and $X^2/\hat{d}.$, where $\hat{\lambda}.$ and $\hat{d}.$ do not depend on the hypothesis, have essentially the same performance. Their SL is reduced to approximately 0.06 in two cases and 0.10 in one case but in the remaining four cases it is higher than 0.30. The corrected statistic $X^2/\hat{\delta}.$ has a more stable performance than $X^2/\hat{\lambda}.$ or $X^2/\hat{d}.$ (SL ranging from 0.095 to 0.16), but not entirely satisfactory due to the large c.v. of the $\hat{\delta}_i$'s. It may also be noted that, unlike the empirical results for two-way tables previously reported, $X^2/\hat{\lambda}.$ or $X^2/\hat{d}.$ may not be conservative for three-way tables; in fact, as shown in Table 1, their SL could be quite high.

This work was supported by a research grant from the Natural Sciences and Engineering Research Council of Canada. This paper was presented at the International Meeting on Analysis of Sample Survey Data and Sequential Analysis, Jerusalem, June 14–18, 1982. Thanks are due to the referees for constructive suggestions.

APPENDIX

Asymptotic equivalence of X^2 *and* $\tilde{X}^2_W(1)$ *under H.* Let **F** be a $(T-1) \times T$ matrix of rank $T-1$ such that $\mathbf{F1} = \mathbf{0}$, and let $\tilde{\mathbf{f}} = F\tilde{\mu}$ where $\tilde{\mu}$ is the T-vector of logprobabilities $\tilde{\mu}_t = \ln \hat{p}_t$. We have $\sqrt{n}(\tilde{\mathbf{f}} - \mathbf{f}) \approx N(\mathbf{0}, \mathbf{V_f})$ where $\mathbf{f} = \mathbf{F}\boldsymbol{\mu} = \mathbf{FX}\boldsymbol{\theta}$ and $\mathbf{V_f} = \mathbf{FD}_\pi^{-1}\mathbf{VD}_\pi^{-1}\mathbf{F}'$. In the case of multinomial sampling, $\mathbf{V_f} = \mathbf{FD}_\pi^{-1}\mathbf{F}'$.

The WLS estimator of $\boldsymbol{\theta}$, under the model $\mathbf{f} = \mathbf{FX}\boldsymbol{\theta}$ is given by

$$\tilde{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{F}'\tilde{\mathbf{V}}_f^{-1}\mathbf{FX})^{-1}\mathbf{X}'\mathbf{F}'\tilde{\mathbf{V}}_f^{-1}\mathbf{F}\tilde{\boldsymbol{\mu}} \tag{A.1}$$

where $\tilde{\mathbf{V}}_f = \mathbf{FD}_{\hat{\mathbf{p}}}^{-1}\hat{\mathbf{V}}\mathbf{D}_{\hat{\mathbf{p}}}^{-1}\mathbf{F}'$. The asymptotic covariance matrix of $\tilde{\boldsymbol{\theta}}$ is given by

$$\mathbf{D}(\tilde{\boldsymbol{\theta}}) = (1/n)(\mathbf{X}'\mathbf{F}'\mathbf{V_f}^{-1}\mathbf{FX})^{-1} = (1/n)\ \mathbf{V}_\theta \quad \text{(say).}$$

Hence, a Wald statistic, which is asymptotically χ_u^2 under $H: \boldsymbol{\theta}_2 = \mathbf{0}$ is given by

$$X_W^2(1) = n\tilde{\boldsymbol{\theta}}_2'\tilde{\mathbf{V}}_{\tilde{\theta}22}^{-1}\tilde{\boldsymbol{\theta}}_2 \tag{A.2}$$

provided $\hat{\mathbf{V}}$ is available. Here $\tilde{\boldsymbol{\theta}}' = (\tilde{\boldsymbol{\theta}}_1', \tilde{\boldsymbol{\theta}}_2')$ and the corresponding partition of $\mathbf{V}_\theta$ is given by

$$\mathbf{V}_\theta = \begin{pmatrix} \mathbf{V}_{\theta 11} & \mathbf{V}_{\theta 12} \\ \mathbf{V}_{\theta 21} & \mathbf{V}_{\theta 22} \end{pmatrix}.$$

Note that $\tilde{\boldsymbol{\theta}}$ and $X_W^2(1)$ are independent of the choice of $\mathbf{F}$, since $\mathbf{F}^{*\prime}\mathbf{V_f}^{-1}\mathbf{F}^* = \mathbf{F}'\mathbf{V}_f^{-1}\mathbf{F}$ for any $\mathbf{F}^* = \mathbf{GF}$, where $\mathbf{G}$ is nonsingular and $\mathbf{f}^* = \mathbf{F}^*\boldsymbol{\mu}$. Consider the particular choice

$$\mathbf{F} = \begin{pmatrix} \mathbf{X}'\mathbf{P} \\ \mathbf{F}_1 \end{pmatrix}$$

where $\mathbf{F}_1$ is $(T - 1 - r) \times T$ matrix of rank $T - 1 - r$ with $\mathbf{F}_1\mathbf{X} = \mathbf{0}$ and $\mathbf{F}_1\mathbf{1} = \mathbf{0}$. With this choice

$$\mathbf{FX} = \begin{pmatrix} \mathbf{X}'\mathbf{PX} \\ \mathbf{0} \end{pmatrix}$$

and, for multinomial sampling with $\mathbf{V} = \mathbf{P}$,

$$\mathbf{V_f} = \begin{pmatrix} \mathbf{X}'\mathbf{PX} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_1'\mathbf{D}_\pi^{-1}\mathbf{F}_1 \end{pmatrix}$$

after some simplification. Thus, replacing $\hat{\mathbf{V}}$ by $\tilde{\mathbf{P}} = \mathbf{D}_{\hat{\mathbf{p}}}^{-1} - \hat{\mathbf{p}}\hat{\mathbf{p}}'$ in (A.1) and (A.2) we get the multinomial-based Wald statistic $\tilde{X}_W^2(1)$ when $\hat{\mathbf{V}}$ is not available.

Now replacing $\hat{\mathbf{V}}$ by $\hat{\mathbf{P}}$ in (A.1) we get

$$\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta} = (\mathbf{X}'\tilde{\mathbf{P}}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{P}}(\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}) \sim (\mathbf{X}'\mathbf{PX})^{-1}\mathbf{X}'\mathbf{PD}_\pi^{-1}(\hat{\mathbf{p}} - \boldsymbol{\pi})$$

$$= (\mathbf{X}'\mathbf{PX})^{-1}\mathbf{X}'(\hat{\mathbf{p}} - \boldsymbol{\pi}) \sim \hat{\boldsymbol{\theta}} - \boldsymbol{\theta},$$

using (4). Moreover, $\mathbf{V}_\theta$ reduces to $(\mathbf{X}'\mathbf{PX})^{-1}$ so that $\mathbf{V}_{\theta 22} = (\tilde{\mathbf{X}}_2'\mathbf{P}\tilde{\mathbf{X}}_2)^{-1}$. Thus, under $H: \boldsymbol{\theta}_2 = \mathbf{0}$,

$$\tilde{X}_W^2(1) = n\tilde{\boldsymbol{\theta}}_2'(\tilde{\mathbf{X}}_2'\tilde{\mathbf{P}}\tilde{\mathbf{X}}_2)^{-1}\tilde{\boldsymbol{\theta}}_2 \sim n\hat{\boldsymbol{\theta}}_2'(\tilde{\mathbf{X}}_2'\mathbf{P}\tilde{\mathbf{X}}_2)^{-1}\hat{\boldsymbol{\theta}}_2 \sim X^2,$$

using (12).

REFERENCES

ALTHAM, P. A. E. (1976). Discrete variable analysis for individuals grouped into families. *Biometrika* **63** 263–269.

BEDRICK, E. J. (1983). Adjusted goodness-of-fit tests for survey data. *Biometrika* **70** (to appear).

BISHOP, Y. M. M., FIENBERG, S. E. and HOLLAND, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* MIT Press, Cambridge, Mass.

BRIER, S. E. (1980). Analysis of contingency tables under cluster sampling. *Biometrika* **67** 591–596.

COHEN, J. E. (1976). The distribution of the chi-squared statistic under cluster sampling from contingency tables. *J. Amer. Statist. Assoc.* **71** 665–670.

FAY, R. E. (1979). On adjusting the Pearson chi-square statistic for clustered sampling. *Proc. of the Social Statistics Section*, Amer. Statist. Assoc. 402–405.

FAY, R. E. (1979). On jackknifing chi-square test statistics—Part II: asymptotic results. Unpublished manuscript.

FELLEGI, I. P. (1980). Approximate tests of independence and goodness of fit based on stratified multistage samples. *J. Amer. Statist. Assoc.* **71** 665–670.

FIENBERG, S. E. (1980). *The Analysis of Cross Classified Data* (2nd ed.). MIT Press, Cambridge, Mass.

FOUTZ, R. V. and SRIVASTAVA, R. C. (1978). The asymptotic distribution of the likelihood ratio when the model is incorrect. *Canadian J. Statist.* **6** 273–279.

HIDIROGLOU, M. A. and RAO, J. N. K. (1981). Chisquare tests for the analysis of categorical data from the Canada Health Survey. Paper presented at the Intl. Statist. Inst. Meetings, Buenos Aires, 1981.

HIDIROGLOU, M., A. and RAO, J. N. K. (1983). Chi-square tests for the analysis of three-way contingency tables from the Canada Health Survey. Technical Report, Statistics Canada.

HOLT, D., SCOTT, A. J., and EWINGS, P. O. (1980). Chi-squared tests with survey data. *J. Roy. Statist. Soc. A* **143** 302–320.

IMREY, P. B., KOCH, G. G. and STOKES, M. E. (1982). Categorical data analysis: some reflections on the log linear model and logistic regression. Part II: data analysis. *Intl. Statist. Rev.* **50** 35–64 (in collaboration with J. N. Darroch, D. H. Freeman, Jr. and H. D. Tolley).

JOHNSON, N. L. and KOTZ, S. (1970). *Continuous Univariate Distributions.* Houghton Mifflin, Boston.

KOCH, G. G., FREEMAN, D. H. JR., and FREEMAN, J. L. (1975). Strategies in the multivariate analysis of data from complex surveys. *Intl. Statist. Rev.* **43** 59–78.

NATHAN, G. (1975). Tests of independence in contingency tables from stratified proportional samples. *Sankhyā C* **37** 77–87.

RAO, J. N. K. and SCOTT, A. J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. *J. Amer. Statist. Assoc.* **76** 221–230.

SATTERTHWAITE, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics* **2** 110–114.

SCOTT, A. J. and RAO, J. N. K. (1981). Chi-squared tests for contingency tables with proportions estimated from survey data. In *Current Topics in Survey Sampling* (D. Krewski, R. Platek and J. N. K. Rao, eds.) 247–266.

SEBER, G. A. F. (1977). *Linear Regression Analysis.* Wiley, New York.

SHUSTER, J. J. and DOWNING, D. J. (1976). Two-way contingency tables for complex sampling schemes. *Biometrika* **63** 271–276.

DEPARTMENT OF MATHEMATICS
AND STATISTICS
ARTS TOWER
CARLETON UNIVERSITY
OTTAWA K1S 5B6
CANADA

DEPARTMENT OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF AUCKLAND
AUCKLAND, NEW ZEALAND