Journal of Official Statistics
Vol. 3, No. 2 1987, pp. 133–140

Chi-Squared Tests with Categorical Data from Complex Surveys:

Part II — Independence in a Three-Way Table with Applications to the Canada Health Survey (1978–1979)

M.A. Hidiroglou and J.N.K. Rao[1]

Abstract: This is the second of a two-paper series presenting a user's guide to the field of chi-squared tests under complex survey designs. The basic chi-squared tests for independence in a three-way table that take account of the survey design are presented and their use illustrated on data from the Canada Health Survey (1978–1979).

Key words: Chi-squared tests; categorical data; complex surveys.

1. Introduction

The first paper in this series focused on chi-squared tests for simple goodness-of-fit and homogeneity and independence in a two-way table. Corrections to standard chi-squared tests were presented and their use illustrated on data from the Canada Health Survey (1978–1979). In this second paper the commonly used tests of independence in a three-way table are investigated in accordance with the work presented in the first paper and their use illustrated, again using data from the Canada Health Survey.

We use the same notation as in the first paper with obvious extensions to three-way tables.

[1] M.A. Hidiroglou is Chief, Business Survey Methods Division, Statistics Canada, 11 "P", R.H.C. Building, Tunney's Pasture, Ottawa, Canada, K1A 0T6. J.N.K. Rao is Professor of Statistics, Department of Mathematics and Statistics, Carleton University, Ottawa, Canada, K1S 5B6.
Acknowledgments are given in the first paper in this series.

2. Concepts, Theoretical Results, and Computational Aspects

Suppose that a three-way table has $I+1$ rows (variable A), $J+1$ columns (variable B) and $K+1$ layers (variable C). Let $p_{ijk} = N_{ijk} / N$ be the population proportion in (i,j,k)-th cell and $N=\Sigma\Sigma\Sigma\, N_{ijk}$, $i=1, \ldots, I+1, j=1, \ldots, J+1$; $k=1, \ldots, K+1$. The survey estimate $\hat{p}_{ijk} = \hat{N}_{ijk} / \hat{N}$ of p_{ijk} is obtained from (2.2) of Part I with 'i' replaced by 'ijk'. Finally, the one-way and two-way estimated marginal proportions are denoted by $(\hat{p}_{i++}, \hat{p}_{+j+}, \hat{p}_{++k})$ and $(\hat{p}_{ij+}, \hat{p}_{+jk}, \hat{p}_{i+k})$ respectively.

In a three-way table, four different types of hypotheses of independence can be specified in terms of the saturated loglinear model

$$\ln p_{ijk} = \mu_{ijk} = \tilde{u} + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{23(jk)} + u_{13(ik)} + u_{123(ijk)}, \qquad (2.1)$$

where the u-parameters sum to zero when summed over any subscript i,j,k and $\tilde{u}$ is the normalizing factor to ensure that $\Sigma\Sigma\Sigma p_{ijk} = 1$.

The hypothesis of complete independence of A, B, and C (denoted as A^*B^*C) is given by

$$H_o(1) : u_{12(ij)} = u_{13(ik)} = u_{23(jk)} = u_{123(ijk)} = 0$$

$$<=> p_{ijk} = p_{i++}\, p_{+j+}\, p_{++k} \text{ for all } (i,j,k). \quad (2.2)$$

Under $H_o(1)$, the estimate of p_{ijk} is $\hat{p}_{ijk}(1) = \hat{p}_{i++}\, \hat{p}_{+j+}\, \hat{p}_{++k}$.

The hypothesis of multiple independence, A independent of (B,C), is denoted as A^*BC. It is given by

$$H_o(2) : u_{12(ij)} = u_{13(ik)} = u_{123(ijk)} = 0 <=>$$

$$p_{ijk} = p_{i++}\, p_{+jk} \text{ for all } (i,j,k). \quad (2.3)$$

The hypotheses B^*AC and C^*AB are analogous to (2.3). The estimated proportions under $H_o(2)$ are given by $\hat{p}_{ijk}(2) = \hat{p}_{i++}\, \hat{p}_{+jk}$.

Another type of independence hypothesis is the conditional independence of A and B given C (denoted as $A^*B|C$). It is given by

$$H_o(3) : u_{12(ij)} = u_{123(ijk)} = 0 <=> p_{ijk} =$$

$$(p_{i+k}\, p_{+jk}) \,/\, p_{++k} \text{ for all } (i,j,k). \quad (2.4)$$

The hypotheses $A^*C|B$ and $B^*C|A$ are analogous to (2.4). The estimated proportions under $H_o(3)$ are given by $\hat{p}_{ijk}(3) = (\hat{p}_{i+k}\, \hat{p}_{+jk}) \,/\, \hat{p}_{++k}$.

Finally, the hypothesis of no three-factor interaction is given by

$$H_o(4) : u_{123}(ijk) = 0 \text{ for all } (i,j,k), \quad (2.5)$$

meaning that the association in the two-way table corresponding to a level of the third variable is constant for all levels. This hypothesis cannot be expressed explicitly in terms of the marginal proportions, unlike the previous hypothesis. As a result, the estimates of p_{ijk} under $H_o(4)$ are obtained by solving "pseudo-likelihood equations" $p_{ij+} = \hat{p}_{ij+}$, $p_{+jk} = \hat{p}_{+jk}$ and $p_{i+k} = \hat{p}_{i+k}$ iteratively for p_{ijk}, using the well-known iterative proportional fitting procedure (IPFP) or some other iterative procedure. The resulting estimates, $\hat{p}_{ijk}(4)$, are consistent, as are the estimates under the previous hypotheses.

The Pearson statistic for testing $H_o(l)$, $l=1,2,3,4$ is given by

$$X_l^2(P) =$$

$$n \sum_{i=1}^{I+1} \sum_{j=1}^{J+1} \sum_{k=1}^{K+1} [\hat{p}_{ijk} - \hat{p}_{ijk}(l)]^2 \,/\, \hat{p}_{ijk}(l). \quad (2.6)$$

In matrix notation the loglinear model under $H_o(l)$ may be expressed as

$$\underset{\sim}{\mu} = \tilde{u}\, \underset{\sim}{1} + \underset{\sim}{X}_{1l}\, \underset{\sim}{\theta}_{1l}, \quad (2.7)$$

where μ is the vector of μ_{ijk}'s (in lexicographical order), $\underset{\sim}{X}_{1l}$ is the full rank model matrix, and $\underset{\sim}{\theta}_{1l}$ is the corresponding vector of u-terms, as explained in Part I for a two-way table. The saturated model (2.1) may be written as $\underset{\sim}{\mu} = \tilde{u}\, \underset{\sim}{1} + \underset{\sim}{X}_{1l}\, \underset{\sim}{\theta}_{1l} + \underset{\sim}{X}_{2l}\, \underset{\sim}{\theta}_{2l}$, where $\underset{\sim}{X}_{2l}$ is orthogonal to $\underset{\sim}{X}_{1l}$ and $\underset{\sim}{1}$.

The Pearson statistic, $X_l^2(P)$, is asymptotically distributed as a weighted sum,

$$\delta_{1l}\, W_1 + \ldots + \delta_{T\text{-}1\text{-}s_l,l}\, W_{T\text{-}1\text{-}s_l},$$

of independent χ_1^2 variables W_i under $H_o(l)$, where s_l is the number of columns of $\underset{\sim}{X}_{1l}$ (i.e., the rank of $\underset{\sim}{X}_{1l}$), $T=(I+1)(J+1)(K+1)$, and the weights δ_{tl} are estimated by $\hat{\delta}_{tl}$, the eigenvalues of "design effects matrix"

$$\hat{D}_l = n\, (\underset{\sim}{X}'_{2l}\, \underset{\sim}{D}^{-1}_{\hat{p}(l)}\, \underset{\sim}{X}_{2l})^{-1}$$

$$(\underset{\sim}{X}'_{2l}\, \underset{\sim}{D}^{-1}_{\hat{p}(l)}\, \hat{\underset{\sim}{\Sigma}}\, \underset{\sim}{\hat{D}}^{-1}_{\hat{p}(l)}\, \underset{\sim}{X}_{2l}), \quad (2.8)$$

with $D_{\hat{p}(l)}$ = diag ($\hat{p}(l)$) and $\hat{p}(l)$ is the T-vector with elements $\hat{p}_{ijk}(l)$ in lexicographical order. Furthermore, $\hat{\Sigma}$ is the estimated covariance matrix of $\hat{p}$, the T-vector of survey estimates $\hat{p}_{ijk}$ (in lexicographical order). Under multinominal sampling, all the δ_{tl} are equal to 1 ($t=1, \ldots, s_l$) and $\Sigma_t \delta_{tl} W_t$ reduces to $\chi^2_{T\text{-}1\text{-}s_l}$, where $s_1 = I+J+K$, $s_2 = I+(J+1)(K+1)\text{-}1$, $s_3 = (K+1)(I+J+1)\text{-}1$ and $s_4 = (I+1)K+(J+1)I+(K+1)J$.

A first order correction to $X^2_l(P)$, as in the first paper in this series, is given by

$$X^2_l(\hat{\delta}_{.l}) = X^2_l(P) / \hat{\delta}_{.l}, \tag{2.9}$$

where $(T\text{-}1\text{-}s_l)\,\hat{\delta}_{.l} = \Sigma_t \hat{\delta}_{tl}$. The statistic $X^2_l(\hat{\delta}_{.l})$ is treated as $\chi^2_{T\text{-}1\text{-}s_l}$ under $H_o(l)$. In the case of $H_o(1)$, $H_o(2)$ and $H_o(3)$, $\hat{\delta}_{.l}$ can be expressed in terms of estimated design effects (deffs) of cell estimates $\hat{p}_{ijk}$ and of their two-way and one-way marginals, as follows:

$$(IJK+IJ+JK+IK)\,\hat{\delta}_{.1} = \sum_{i=1}^{I+1}\sum_{j=1}^{J+1}\sum_{k=1}^{K+1} \frac{\hat{p}_{ijk}\ (1-\hat{p}_{ijk})}{\hat{p}_{i++}\ \hat{p}_{+j+}\ p_{++k}}\,\hat{d}_{ijk} - \sum_{i=1}^{I+1} (1\text{-}\hat{p}_{i++})\,\hat{d}_{A(i)} - \sum_{j=1}^{J+1} (1\text{-}\hat{p}_{+j+})\,\hat{d}_{B(j)} - \sum_{k=1}^{K+1} (1\text{-}\hat{p}_{++k})\,d_{C(k)}, \tag{2.10}$$

$$I(JK+J+K)\,\hat{\delta}_{.2} = \sum_{i=1}^{I+1}\sum_{j=1}^{J+1}\sum_{k=1}^{K+1} \frac{\hat{p}_{ijk}\ (1\text{-}\hat{p}_{ijk})}{\hat{p}_{i++}\hat{p}_{+jk}}\,\hat{d}_{ijk} - \sum_{i=1}^{I+1} (1\text{-}\hat{p}_{i++})\,\hat{d}_{A(i)} - \sum_{j=1}^{J+1}\sum_{k=1}^{K+1} (1\text{-}\hat{p}_{+jk})\,\hat{d}_{BC(jk)}, \tag{2.11}$$

and

$$IJ(K+1)\,\hat{\delta}_{.3} = \sum_{i=1}^{I+1}\sum_{j=1}^{J+1}\sum_{k=1}^{K+1} \frac{\hat{p}_{ijk}\ (1\text{-}\hat{p}_{ijk})}{(\hat{p}_{i+k}\ \hat{p}_{+jk} / \hat{p}_{..k})}\,\hat{d}_{ijk} - \sum_{i=1}^{I+1}\sum_{k=1}^{K+1} (1\text{-}\hat{p}_{i+k})\,\hat{d}_{AC(ik)} - \sum_{j=1}^{J+1}\sum_{k=1}^{K+1} (1\text{-}\hat{p}_{+jk})\,\hat{d}_{BC(jk)} + \sum_{k=1}^{K+1} (1\text{-}\hat{p}_{++k})\,\hat{d}_{C(k)}, \tag{2.12}$$

(Rao and Scott (1984)). Here $\hat{d}_{ijk}$ = estvar $(\hat{p}_{ijk}) / [\hat{p}_{ijk}\,(1\text{-}\hat{p}_{ijk})\,n^{-1}]$ is the estimated deff of $\hat{p}_{ijk}$, $\hat{d}_{A(i)}$ = estvar $(\hat{p}_{i++}) / [\hat{p}_{i++}\,(1\text{-}\hat{p}_{i++})\,n^{-1}]$ is the estimated deff of one-way marginal $\hat{p}_{i++}$, and $\hat{d}_{BC(jk)}$ = estvar $(\hat{p}_{+jk}) / [\hat{p}_{+jk}\,(1\text{-}\hat{p}_{+jk})\,n^{-1}]$ is the estimated deff of two-way marginal $\hat{p}_{+jk}$, and so on.

In the case of $H_o(4)$, the hypothesis of no three-factor interaction, $\hat{\delta}_{.4}$ cannot be expressed in terms of deffs of cell estimates and of their marginals. It requires the knowledge of full estimated covariance matrix $\hat{\Sigma}$, but Rao and Scott (1987) proposed an approximation to $\hat{\delta}_{.4}$ in terms of deffs of cell estimates and of their marginals. Using the formula for $\hat{\delta}_{.3}$ (similar to (2.12)) for the conditional independence hypothesis "closest" to the hypothesis of no three-factor interaction in terms of $X^2(P)$-value, the approximation is given by

$$\delta^*_{.4} = \frac{(T\text{-}1\text{-}s_3)}{(T\text{-}1\text{-}s_4)}\,\hat{\delta}_{.3}, \tag{2.13}$$

where s_3 is the d.f. for the "closest" conditional independence hypothesis.

The resulting correction to $X^2_4(P)$, namely $X^2_4(\delta^*_{.4})$, is "nearly conservative" relative to

$X_4^2\ (\hat{\delta}_{.4})$ when the two hypotheses are very "close".

A second order correction to $X_l^2(P)$, based on the Satterthwaite approximation of the weighted sum of independent χ_1^2 variables, requires the knowledge of the estimated covariance matrix $\hat{\Gamma}_l = [\gamma_{ijk,i'j'k'}\ (l)]$ of the residuals $\hat{h}_{ijk}\ (l) = \hat{p}_{ijk} - \hat{p}_{ijk}\ (l)$. In general, $\hat{\Gamma}_l$ is given by

$$\hat{\Gamma}_l = M_l \hat{\Sigma} M_l', \tag{2.14}$$

where $M_l = I - \hat{P}_l X_{1l} (X_{1l}' \hat{P}_l X_{1l})^{-1}$ and $\hat{P}_l = D_{\hat{p}(l)} - \hat{p}(l)\ \hat{p}(l)'$. The covariance matrices $\hat{\Gamma}_l$ for $l=1,2,3$, however, can be obtained more simply from formula (2.4) of Part I with z_{iht} replaced $z_{ijkht}\ (l)$ given in the Appendix.

The Satterthwaite correction to $X_l^2(P)$, $l=1,2,3,4$, is obtained by treating

$$X_l^2\ (S) = \frac{X_l^2\ (P)}{\hat{\delta}_{.l}\ (1+\hat{C}_{\delta l}^2)} \text{ as } \chi^2_{\upsilon_l}\ ;\ \upsilon_l = \frac{T\text{-}1\text{-}s_l}{1+\hat{C}_{\delta l}^2}, \tag{2.15}$$

where

$$(T\text{-}1\text{-}s_l)\ \hat{\delta}_{.l} = n \sum_{i=1}^{I+1} \sum_{j=1}^{J+1} \sum_{k=1}^{K+1} \frac{\hat{\gamma}_{ijk,\ ijk}(l)}{\hat{p}_{ijk}\ (l)}, \tag{2.16}$$

and

$$\sum_{t=1}^{T\text{-}1\text{-}s_l} \hat{\delta}_{tl}^2 = n^2 \sum_{i,i'=1}^{I+1} \sum_{j,j'=1}^{J+1} \sum_{k,k'=1}^{K+1} \frac{\hat{\gamma}^2_{ijk,\ i'j'k'}\ (l)}{\hat{p}_{ijk}\ (l)\ \hat{p}_{i'j'k'}\ (l)}. \tag{2.17}$$

Note that $\hat{\delta}_{.l}$ for $l=1,2,3$ can also be calculated from (2.10), (2.11), and (2.12) respectively, but (2.16) is readily calculated from the diagonal elements of Γ_l. As noted in Part I, it is convenient to use the critical point $\chi^2_{T\text{-}1\text{-}s_l}\ (\alpha)$, the customary upper α-point of $\chi^2_{T\text{-}1\text{-}s_l}$, rather than $\chi^2_{\upsilon_l}\ (\alpha)$, in which case $X_l^2(S)$ should be modified to

$$X_l^2\ (S, \alpha) = X_l^2\ (S)\ [\chi^2_{T\text{-}1\text{-}s_l}(\alpha)\ /\ \chi^2_{\upsilon_l}(\alpha)]. \tag{2.18}$$

The null hypothesis $H_0(l)$ is rejected at the α-level if $X_l^2(S,\ \alpha)$ exceeds $\chi^2_{T\text{-}1\text{-}s_l}\ (\alpha)$.

The type I error rates of customary $X_l^2(P)$ and the first order correction $X_l^2(\hat{\delta}_{.l})$ are estimated as in Part I, assuming that the Satterthwaite approximation is accurate.

A Wald statistic, which is asymptotically a $\chi^2_{T\text{-}1\text{-}s_l}$ under $H_0(l)$, is given by

$$X_l^2\ (W) = \hat{\phi}_l'\ [\hat{\Sigma}_\phi\ (l)]^{-1}\ \hat{\phi}\ (l), \tag{2.19}$$

where $\hat{\phi}_l = X_{2l}'\ \hat{\mu}$ with estimated covariance matrix $\hat{\Sigma}_\phi\ (l) = X_{2l}'\ D_{\hat{p}}^{-1}\ \hat{\Sigma}\ D_{\hat{p}}^{-1}\ X_{2l}$ assuming that all the elements of $\hat{p}$ are nonzero, and $\hat{\mu}$ is the T-vector of log-probabilities ln $\hat{p}_{ijk}$ (Rao and Scott (1984)). If the degrees of freedom, r, for $\hat{\Sigma}$ is not large relative to T-1-s_l, an improvement compared to $X_l^2(W)$ can be obtained, as in Part I, by treating

$$F_l\ (W) = \frac{(r\text{-}T+s_l+2)}{r(T\text{-}1\text{-}s_l)}\ X_l^2\ (W) \tag{2.20}$$

as an F-variable with T-1-s_l and r-T+s_l+2 d.f. respectively, under $H_0(l)$. In the context of Canada Health Survey, r may be taken as the number of sampled clusters minus the number of strata.

Analysis of residuals, $\hat{h}_{ijk}(l) = \hat{p}_{ijk} - \hat{p}_{ijk}(l)$, is useful to detect deviations from $H_0(l)$. The standardized residuals

$$\hat{e}_{ijk}\ (l) = \frac{\hat{h}_{ijk}\ (l)}{[\ \hat{\gamma}_{ijk,\ ijk}\ (l)\ ]^{1/2}} \tag{2.21}$$

are approximately $N(0,1)$ under $H_0(l)$. For

l=1,2,3, we can express (2.21) as

$$\hat{e}_{ijk}(l) = \frac{e_{ijk}(l)}{[\,\hat{d}\,(\hat{h}_{ijk}(l)\,]^{1/2}} \quad , \tag{2.22}$$

where $e_{ijk}(l)$ are the standardized residuals under multinominal sampling given in the Appendix, and $\hat{d}$ [$\hat{h}_{ijk}(l)$] is the estimated deff of $\hat{h}_{ijk}(l)$ under $H_o(l)$.

As in Part I, the results for a particular domain, say D_1, can be obtained by replacing n by $n_{(1)}$, $\hat{p}_{ijk}$ by $\hat{q}_{ijk} = \hat{N}_{(1)ijk} / \hat{N}_{(1)}$, and so on.

2.1. *Example*

Consider the estimated counts, $\hat{N}_{(1)ijk}$, from the Canada Health Survey (1978 – 79) of females aged 15 – 64, D_1, cross-classified by frequency of breast self-examination (variable B, four categories: monthly, quarterly, less often, never), education (variable A, three categories: secondary or less, some post-secondary, post-secondary) and age (variable C, four categories: 15–19, 20–24, 25–44, 45–64). These counts, given in Table 1, are adjusted for post-stratification, as explained in Part I. Here $\hat{N}_{(1)} = \Sigma\Sigma\Sigma\ \hat{N}_{(1)ijk}$, $\hat{q}_{ijk} = \hat{N}_{1(ijk)} / N_{(1)}$, $n_{(1)} = 8713$, $I+1 = 3$, $J+1 = 4$ and $K+1 = 4$.

The results for testing $H_o(1)$, l=1,2,3,4 are summarized in Table 2, where l=2 and l=3 each involve three different hypotheses.

Table 1. Estimated Counts (in Thousands) of Females Aged 15 – 64 in a 3×4×4 Table (National Level: $n_{(1)}$ = 8713)

	Monthly	Quarterly	Less often	Never
			Age 15 – 19	
Secondary or less	92	79	108	615
Some post-secondary	11	10	23	59
Post-secondary	2.4	2.4	0.33	4.9
			Age 20 – 24	
Secondary or less	147	144	106	202
Some post-secondary	41	27	54	44
Post-secondary	53	56	70	54
			Age 25 – 44	
Secondary or less	486	488	446	539
Some post-secondary	60	64	56	43
Post-secondary	213	244	197	157
			Age 45 – 64	
Secondary or less	469	408	312	520
Some post-secondary	26	40	26	13
Post-secondary	72	69	71	38

Table 2. Values of $\hat{\delta}_{.l}$, $\hat{C}_{\delta l}$ and Test Statistics $X_l^2(P)$, $X_l^2(\hat{\delta}_{.l})$, $X_l^2(S,0.05)$, $X_l^2(W)$ and $F_l(W)$ for Eight Hypotheses in a 3×4×4 Table (r=69)

	Independence hypotheses							
	$H_o(1)$	$H_o(2)$			$H_o(3)$			$H_o(4)$
	A^*B^*C	A^*BC	B^*AC	C^*AB	$A^*B\|C$	$B^*C\|A$	$A^*C\|B$	$u_{123}=0$
$\hat{\delta}_{.l}$	2.09	2.16	1.87	2.12	1.91	1.90	2.24	1.84
$\hat{C}_{\delta l}$	1.71	1.58	1.41	1.60	1.22	1.36	1.46	1.14
$X_l^2(P)$	1 867	842	988	1 572	148	767	678	45
$X_l^2(\hat{\delta}_{.l})$	892	389	528	742	78	402	303	24
$X_l^2(S,0.05)$	680	298	425	571	64	322	235	20
$X_l^2(W)$	11 042	2 453	3 774	5 876	653	2 349	762	60
$F_l(W)$	127	47	61	96	18	54	21	2.5
T-1-s_l	39	30	33	33	24	27	24	18
r-T+s_l+2	31	40	37	37	46	43	46	52

The effect of sample design on the Pearson statistic $X_l^2(P)$ is very severe for all the eight hypotheses: estimated type I error rate of 0.54 or larger compared to nominal level 0.05. The first order correction $X_l^2(\hat{\delta}_{.l})$ brings the type I error rate down to the 0.12 – 0.17 range, but this is not entirely satisfactory due to a large $\hat{C}_{\delta l}$, which is the coefficient of variation of the $\hat{\delta}_{il}$'s, for all the hypotheses. The simple correction $X_l^2(P)$ / $\hat{d}_{.}$, based on the average cell deff $\hat{d}_{.}$, independent of $H_o(l)$, exhibits a less stable performance than $X_l^2(\hat{\delta}_{.l})$, the type I error rate ranging from 0.11 to 0.25. It may be noted that $X_l^2(P)$ / $\hat{d}_{.}$ is not conservative here, unlike in the two-way table example of Part I.

The first correction $X_4^2(\delta^*_{.4})$ for testing $H_o(4)$, which depends only on the deffs of cell proportions and of their marginals, is somewhat conservative relative to $X_4^2(\hat{\delta}_{.4})$: $X_4^2(\delta^*_{.4})$ = 17.6 compared to $X_4^2(\hat{\delta}_{.4})$ = 24 where $\delta^*_{.4}$ = 2.55 using $A^*B|C$ as the hypotheses "closest" to $H_o(4)$.

The second order correction $X_4^2(S,0.05)$ for the no three-factor interaction hypothesis is not significant at the 5% level (20.00 compared to $\chi_{18}^2(0.05)$ = 28.87, the upper 5% point of χ^2 with 18 d.f.). The loglinear model (2.1) with u_{123} = 0 thus seems to provide an adequate fit to the data of Table 1. The values of X_l^2(S,0.05) for all the simpler models, given in Table 2, are significant at the 5% level.

The Wald statistic, $X_l^2(W)$, is unstable here, leading to values larger than the unadjusted $X_l^2(P)$ in all cases, and 3 to 5 times larger values in six cases out of eight. It may also be noted that $X_l^2(W)$ is much larger than X_l^2(S,0.05) in all cases. The instability of $X_l^2(W)$ is caused by the fact that $\hat{\Sigma}_{\phi}^{-1}(l)$ is ill-conditioned (see Table 3 which gives the largest and smallest eigenvalues, λ_{max} and λ_{min} and the condition number, λ_{max} / λ_{min}, of the matrix $\hat{\Sigma}_{\phi}(l)$). The condition number for all the eight hypotheses is larger than 1 000, and larger than 10 000 for four of the hypoth-

Table 3. Values of the Wald Statistic, X_1^2(w), and the Smallest and Largest Eigenvalues and the Condition Number of $\hat{\Sigma}_\phi(l)$ for Eight Hypotheses

	Independence hypotheses							
	$H_o(1)$	$H_o(2)$			$H_o(3)$			$H_o(4)$
	A^*B^*C	A^*BC	B^*AC	C^*AB	$A^*B\|C$	$B^*C\|A$	$A^*C\|B$	$u_{123}=0$
				3x4x4 Table				
$X_l^2(W)$	11 042	2 453	3 774	5 875	653	2 349	761	60
$\lambda_{\min}(\hat{\Sigma}_\phi(l))$	1.46x10^{-4}	7.52x10^{-4}	3.89x10^{-4}	3.21x10^{-4}	9.62x10^{-4}	8.14x10^{-4}	2.75x10^{-3}	3.71x10_{-3}
$\lambda_{\max}(\hat{\Sigma}_\phi(l))$	12.82	11.03	9.95	9.24	8.23	6.25	7.53	4.79
$\lambda_{\max} / \lambda_{\min}$	87 832	14 676	25 664	28 790	8 551	7 680	2 738	1 294
T-1-s_l	39	30	33	33	24	27	24	18
				2x3x3 Table				
$X_l^2(W)$	752	202	733	384	132	365	78	9.7
$X_l^2(P)$	900	420	655	698	199	467	252	31
$\lambda_{\min}(\hat{\Sigma}_\phi(l))$	2.01x10^{-3}	5.30x10^{-3}	2.65x10^{-3}	2.67x10^{-3}	5.80x10^{-3}	4.53x10^{-3}	6.39x10^{-3}	8.42x10^{-3}
$\lambda_{\max}(\hat{\Sigma}_\phi(l))$	2.87x10^{-1}	2.41x10^{-1}	2.77x10^{-1}	2.84x10^{-1}	1.88x10^{-1}	2.74x10^{-1}	2.37x10^{-1}	1.87x10^{-1}
$\lambda_{\max} / \lambda_{\min}$	143	45	102	106	32	60	37	22
T-1-s_l	12	8	10	10	6	8	6	4

eses, thus clearly indicating that $\hat{\Sigma}_\phi^{-1}(l)$ is ill-conditioned. The instability of $X_l^2(W)$ in the 3×4×4 table can be further demonstrated by comparing its values with the corresponding values from a smaller table (2×3×3) obtained by collapsing the original 3×4×4 table: variable B, three categories: monthly or quarterly, less often, never; variable A, two categories: secondary or less, some post-secondary or post-secondary; variable C, three categories: 15–24, 25–44, 45–64.

The condition number of $\hat{\Sigma}_\phi(l)$ now ranges from 22–143, thus indicating significant improvement in the stability of $X_l^2(W)$ for the smaller table.

The F-version of the Wald statistic, $F_l(W)$, for the 3×4×4 table leads to some improvement over $X_l^2(W)$, but it also suffers from the instability problem (see Table 2). For example, the value of $F_1(W)$ for the hypothesis A^*B^*C is 127 compared to unadjusted $X_1^2(P) / (T\text{-}1\text{-}s_l) = 1\ 867 / 39 \doteq 48$.

3. References

Haberman, S.J. (1973): The Analysis of Residuals in Cross-classified Tables. Biometrics, 29, pp. 205–220.

Rao, J.N.K. and Scott, A.J. (1984): On Chi-Squared Tests for Multi-Way Tables with Cell Proportions Estimated from Survey Data. Annals of Statistics, 12, pp. 46 – 60.

Rao, J.N.K. and Scott, A.J. (1987): On Simple Adjustments of Chi-Squared Tests with Sample Survey Data. Annals of Statistics, 15 (in press).

Received January 1986
Revised April 1987

Appendix

Estimated Variances and Covariances of $\hat{h}_{ijk}(l)$ for l=1,2,3

The estimated variances and covariances of $\hat{h}_{ijl}(l)$ for l=1,2,3 are given by (2.4) of Part I with z_{iht} replaced by $z_{ijkht}(l)$, where

$$z_{ijkht}(1) = \{B_{ijk(ht)} - \hat{p}_{i++}\,\hat{p}_{+j+}\,B_{++k(ht)}$$
$$-\hat{p}_{++j}\,\hat{p}_{++k}\,B_{i++(ht)}$$
$$-\hat{p}_{i++}\,\hat{p}_{++k}\,B_{+j+(ht)}$$
$$+\,2\,\hat{p}_{i++}\,\hat{p}_{+j+}\,\hat{p}_{++k}\,B_{(ht)}\,\}$$
$$-\Sigma_a\,({}_aB_{ht}\,/\,{}_a\hat{N})\,\{{}_a\hat{N}_{ijk} - \hat{p}_{i++}\,\hat{p}_{++k}\,{}_a\hat{N}_{+j+}$$
$$-\hat{p}_{i++}\,\hat{p}_{+j+}\,{}_a\hat{N}_{++k} - \hat{p}_{+j+}\,\hat{p}_{++k}\,{}_a\hat{N}_{i++}$$
$$+\,2\,\hat{p}_{i++}\,\hat{p}_{+j+}\,\hat{p}_{++k}\,{}_a\hat{N}\}, \qquad (A.1)$$

$$z_{ijkht}(2) = \{B_{ijk(ht)} - \hat{p}_{i++}\,B_{+jk(ht)}$$
$$-\hat{p}_{+jk}\,B_{i++(ht)} + \hat{p}_{i++}\,\hat{p}_{+jk}\,B_{(ht)}\,\}$$
$$-\,\Sigma_a\,({}_aB_{ht}\,/\,{}_a\hat{N})\,\{{}_a\hat{N}_{ijk} - \hat{p}_{i++}\,{}_a\hat{N}_{+jk}$$
$$-\hat{p}_{+jk}\,{}_a\hat{N}_{i++} + \hat{p}_{i++}\,\hat{p}_{+jk}\,{}_a\hat{N}\}, \qquad (A.2)$$

and

$$z_{ijkht}(3) = \{B_{ijk(ht)} - \frac{\hat{p}_{+jk}}{\hat{p}_{++k}}\,B_{i+k(ht)}$$
$$-\,\frac{\hat{p}_{i+k}}{\hat{p}_{++k}}\,B_{+jk(ht)} + \frac{\hat{p}_{i+k}\,\hat{p}_{+jk}}{\hat{p}^2_{++k}}\,B_{(ht)}\}$$
$$-\Sigma_a\,({}_aB_{ht}\,/\,{}_a\hat{N})\,\{{}_a\hat{N}_{ijk} - \frac{\hat{p}_{+jk}}{\hat{p}_{++k}}\,{}_a\hat{N}_{i+k}$$
$$-\,\frac{\hat{p}_{i+k}}{\hat{p}_{++k}}\,{}_a\hat{N}_{+jk} + \frac{\hat{p}_{i+k}\,\hat{p}_{+jk}}{\hat{p}^2_{++k}}\,{}_a\hat{N}\}, \qquad (A.3)$$

with obvious extention of notation used in Part I.

Standardized Residuals Under Multinominal Sampling

The standardized residuals $e_{ijk}(l)$ for l=1,2,3 under multinominal sampling are given by (Haberman (1973)).

$$e_{ijk}(1) = n^{1/2}\,[\hat{p}_{ijk} - \hat{p}_{ijk}(1)]$$
$$[\hat{p}_{ijk}(1)\,\{1\text{-}\hat{p}_{i++}\,\hat{p}_{+j+} - \hat{p}_{i++}\,\hat{p}_{++k}$$
$$-\hat{p}_{+j+}\,\hat{p}_{++k} + 2\,\hat{p}_{ijk}(1)\,\}\,]^{-1/2}, \qquad (A.4)$$

$$e_{ijk}(2) = n^{1/2}\,[\hat{p}_{ijk} - \hat{p}_{ijk}(2)\,]$$
$$[\hat{p}_{ijk}(2)\;(1\text{-}\hat{p}_{i++})\;(1\text{-}\hat{p}_{+jk})\,]^{-1/2}, \qquad (A.5)$$

and

$$e_{ijk}(3) = n^{1/2}\,[\hat{p}_{ijk} - \hat{p}_{ijk}(3)\,]$$
$$[\hat{p}_{ijk}(3)\;(1\text{-}\frac{\hat{p}_{i+k}}{\hat{p}_{++k}})\;(1\text{-}\frac{\hat{p}_{+jk}}{\hat{p}_{++k}})\,]^{-1/2} \qquad (A.6)$$

respectively.