

# Chi-Squared Tests with Categorical Data from Complex Surveys:

## Part I – Simple Goodness-of-Fit, Homogeneity and Independence in a Two-Way Table with Applications to the Canada Health Survey (1978–1979)

*M.A. Hidirolou and J.N.K. Rao<sup>1</sup>*

**Abstract:** This is the first 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 that take account of the survey design are presented and their use illustrated on data from the Canada Health Survey (1978–1979). The

commonly used simple goodness-of-fit test and homogeneity and independence tests in two-way tables are studied in Part I.

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

### 1. Introduction

Sample surveys are generally designed to produce reliable estimates of simple descriptive parameters such as population totals, means and proportions. Because of cost and operational constraints in designing and implementing a survey, the sample design that

best meets desired objectives is often complex and highly clustered.

In recent years, the growing trend has been to use survey data for statistical analysis beyond the estimation of simple descriptive parameters. Such analyses often ignore the sample design and apply standard statistical methods appropriate for data collected through simple random sampling. The availability of computer packages (for standard analyses) and wealth of survey data published in tabular form have contributed to this trend. It is unfortunate that software offering traditional statistical methods are so readily used on published data. For this reason, it is important to investigate the effect of sample design on standard statistical methods and suggest, if needed, adjustments that make the methods valid.

<sup>1</sup> M.A. Hidirolou 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. The authors would like to thank Phillip Giles and David Paton for assistance in preparing summary tapes and for suggestions on computational aspects pertaining to the Canada Health Survey. Constructive suggestions from David Binder, Nanjamma Chinnappa, Geoffrey Hole, Carl Särndal, Alastair Scott and Roland Thomas are also appreciated. The referees and the Associate Editor provided many helpful suggestions which enhanced the final paper. The authors would also like to thank Judy Clarke and Lucie Gagné for their excellent typing.

In this two-paper series, we focus on methods for the analysis of categorical data summarized in the form of cross-classified tables of estimated counts, commonly referred to as contingency tables. Our main aim is to give a user's guide to the field of chi-squared tests for categorical data arising from complex sample surveys. We present the basic chi-squared tests that take account of the sample design and illustrate their use on data from the Canada Health Survey (1978–1979), a typical complex survey based on a multi-stage design involving stratification and cluster sampling. Computational methods and hints are also provided. Essentially no mathematical derivations are given, but relevant references (for derivations) are provided.

Standard test statistics for categorical data include the Pearson chi-squared test,  $X^2$ . These statistics, however, are not asymptotically distributed as  $\chi^2$  random variables under stratification and clustering. Rao and Scott (1979, 1981, 1984) have, in fact, shown that  $X^2$  is asymptotically distributed as a weighted sum,  $\sum \delta_i W_i$ , of independent  $\chi^2_1$  random variables  $W_i$ , where the weights  $\delta_i$  are related to the familiar design effects (deffs) used by survey samplers. They also developed an adjustment to  $X^2$ , based on the Satterthwaite (1946) approximation to a weighted sum of independent  $\chi^2$  variables, that requires knowledge of the full estimated covariance matrix of cell estimates. Fay (1979, 1985) developed a jackknife  $X^2$  also taking the design into account, but requiring cell estimates at the psu (primary sampling unit) level. The jackknife  $X^2$ , however, is not investigated in this paper. Koch et al. (1975), Nathan (1975) and others proposed asymptotically valid tests, based on the Wald statistic, that require access to the full estimated covariance matrix of cell estimates. If the degrees of freedom (d.f.) for the estimated covariance matrix is not large relative

to the d.f. for  $X^2$ , the Wald statistic is often unreliable due to instability in the estimated inverse covariance matrix (Fay (1985)). Monte Carlo results (Thomas and Rao (1984)) also indicate that the Wald statistic, although asymptotically valid, does not control the type I error rate satisfactorily in the above situation, unlike the Satterthwaite-adjusted  $X^2$  or the jackknife  $X^2$ .

Rao and Scott (1979, 1981, 1982, 1984), Scott and Rao (1981), Gross (1984) and Bedrick (1983) have provided first order corrections to  $X^2$  requiring only knowledge of cell deffs (or cell variance estimates) and the deffs of marginal totals in the contingency table. These corrections are particularly useful for performing secondary analyses from published tables, for which the researcher may not have access to the detailed information needed for implementing the Satterthwaite-adjusted  $X^2$  or the jackknife  $X^2$  or the Wald statistic.

The plan for the first of this two-paper series is as follows. In Section 2 we briefly describe the Canada Health Survey (1978–1979) and the procedures used for estimating cell counts, proportions, variances, and covariances. The procedures should have wide applicability since the design of the Canada Health Survey is similar to the design of many large-scale surveys. The commonly used simple goodness-of-fit, homogeneity, and independent tests in two-way tables are investigated in Sections 3, 4, and 5 respectively. Each of these sections gives a summary of relevant theoretical results and computational aspects followed by an example from the Canada Health Survey. Convenient formulae are provided to implement the tests and to carry out residual analysis to detect model deviations. In the 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. It may be pointed out that the pub-

lished tables from the Canada Health Survey (see *The Health of Canadians* (1981)) are mostly three-way tables either at the region level (five regions in Canada) or at the national level.

Experimental software for the above methods has been developed at Statistics Canada by modifying the MINI CARP program (Hidiroglou et al. (1980)). These programs are available but they have not been fully documented for general use.

## **2. Canada Health Survey (1978–1979)**

A brief description of the Canada Health Survey (1978 – 1979) and the procedures used for estimating cell counts, proportions, their estimated variances, and covariances are given in this section. The reader is referred to the *Health of Canadians* (1981) for further details and for the various cross-classified tables of estimated counts.

### *2.1. Description of the Survey*

The broad objectives of the Canada Health Survey (1978 – 1979) were to provide reliable statistics on the current health status of the Canadian population and the antecedents (risk factors) to and consequences of the health status, including changes over time at the national and provincial levels. The data were collected from monthly samples, but the survey was terminated after the first year because of budget cuts. The tables published in *The Health of Canadians* (1981) are based on the data pooled over the nine-month period July 1978 to March 1979.

The information collected was made up of two main components. The first, referred to as the interview component, used two types of questionnaires. The first questionnaire covered items which usually require interviewer probing, but could also be obtained for the entire household from a suitable

member. These items included accidents and injuries, chronic conditions, disability days, and utilization of health services. The second questionnaire covered items which could be sensitive and could only be reliably answered by the person sampled. Due to its content and the need for respondent completion, it was limited to persons 15 years and over. The items covered included alcohol use, tobacco use, health related activities, and emotional health. Part-time interviewers collected the data for the interview component.

The second component, referred to as the physical measures component, was divided into two parts. The first part included physical measurements of blood pressure, cardio-respiratory fitness, weight and skinfold on persons age two years and over. The second part involved the taking of blood samples from persons three years and over to determine immune status as well as biochemical and trace metal levels. Part-time nurses collected the physical measures data.

The Canada Health Survey used a multi-stage, stratified cluster sampling design. An aim of the design was to achieve an annual sample of 12 000 households in the interview component from 100 sample geographical clusters in monthly samples of 10 households per cluster. A subsample of 4 200 households was selected for the physical measures component in 50 of the 100 interview sample clusters at the rate of seven out of the ten interview households per cluster per month. Initially the 100 sample clusters were allocated to the provinces proportional to the square root of their 1971 census populations. Three major strata (major cities, other urban areas, and rural areas) were formed within each province and the sample clusters were allocated to these strata proportional to their 1971 census population counts, with the requirement that the minimum allocation to a stratum be two clusters.

In each of the major cities, a minimum of

two clusters were selected by simple systematic sampling. From the other urban areas stratum, a systematic sample of cities was selected within each province with probability proportional to their 1971 population. Each selected city was allocated one cluster. For both major cities and other urban areas, a cluster was defined as a group of city blocks, and a simple systematic sample of households was selected within each sample cluster. For the rural areas stratum, each province was sampled using a three-stage design with systematic sampling at each stage (see The Health of Canadians (1981), Appendix II for details). The response rates were 86% and 89% for the two parts of the interview component, and 72% and 80% for the two parts of the physical measures component, respectively.

## 2.2. Estimation of total counts and proportions

To provide estimates of total counts of categorical variables at the provincial level, some preliminary notation is required. We define the variable  ${}_a y_{i(htk)}$  for the  $k$ -th sample element in the  $t$ -th sample first-stage unit of the  $h$ -th stratum as one if the element belongs to the  $i$ -th category and the  $a$ -th age-sex group and zero otherwise ( $i=1, \dots, I+1$ ;  $a=1, \dots, A$ ;  $h=1, \dots, L$ ;  $t=1, \dots, r_h$ ;  $k=1, \dots, m_{ht}$ ). Similarly the indicator variable  ${}_a x_{htk}$  is defined as one if the  $(htk)$ -th sample element belongs to the  $a$ -th age-sex group and zero otherwise.

The basic sampling weight attached to the  $(htk)$ -th element, denoted by  $w_{htk}$ , is taken as the inverse of the probability of inclusion of the element in the sample (i.e., the estimate of a total is the usual Horvitz-Thompson estimate), but adjusted for nonresponse (see The Health of Canadians (1981) page 18 for details on the nonresponse adjustment). The estimates of total counts were also adjusted for post-stratification using the projected

census age-sex distribution at the provincial level. The adjusted estimate of total count in the  $i$ -th category at the provincial level is thus given by

$$\hat{N}_i = \sum_a ({}_a \hat{N}_i / {}_a \hat{N}) {}_a N, i=1, \dots, I+1, \quad (2.1)$$

where

$${}_a \hat{N}_i = \sum_h \sum_t \{ \sum_k w_{htk} {}_a y_{i(htk)} \} = \sum_h \sum_t {}_a B_{i(ht)}, \text{ say}$$

$${}_a \hat{N} = \sum_h \sum_t \{ \sum_k w_{htk} {}_a x_{htk} \} = \sum_h \sum_t {}_a B_{ht}, \text{ say}$$

and  ${}_a N$  is the projected census population of the province in the  $a$ -th age-sex group at the time of the survey. The estimate  $\hat{N}_i$  should be more efficient than the unadjusted Horvitz-Thompson estimate  $N_i^* = \sum_a {}_a \hat{N}_i = \sum_h \sum_t \sum_k w_{htk} y_{i(htk)}$  for characteristics closely related to age and sex, where  $y_{i(htk)}$  takes the value one if the  $(htk)$ -th element belongs to the  $i$ -th category and zero otherwise.

If  $\hat{N}_i(1), \dots, \hat{N}_i(m)$  denote the estimates (2.1) for the  $m$  provinces in a region (or in Canada), then the aggregate estimate is  $\hat{N}_i = \hat{N}_i(1) + \dots + \hat{N}_i(m)$ . The estimate of the proportion,  $p_i$ , in the  $i$ -th category is given by the combined ratio estimate

$$\hat{p}_i = \hat{N}_i / \hat{N}, \quad (2.2)$$

where  $\hat{N} = \sum_i \hat{N}_i$ . The unadjusted estimate of  $p_i$  is  $p_i^* = N_i^* / N^*$ , where  $N_i^* = N_i^*(1) + \dots + N_i^*(m)$  and  $N^* = \sum_i N_i^*$ .

## 2.3. Estimated variances and covariances

We calculated estimates of variances and covariances by assuming that the first-stage units within a stratum have been selected with replacement. This assumption leads to some overestimation of sampling variances, but these estimates include response variance. The estimated covariance of  $\hat{N}_i$  and  $\hat{N}_j$  is given by (see the Appendix on page 132):

$$\text{estcov}(\hat{N}_i, \hat{N}_l) =$$

$$\sum_{h=1}^L \frac{r_h}{r_h - 1} \sum_{i=1}^{r_h} (z_{iht} - \bar{z}_{ih})(z_{lht} - \bar{z}_{lh}), \quad (2.3)$$

where

$$z_{iht} = B_{i(ht)} - \sum_a (a \hat{N}_i / a \hat{N}) a B_{ht},$$

and

$$B_{i(ht)} = \sum_a a B_{i(ht)}, \quad \bar{z}_{ih} = \sum_i z_{iht} / r_h.$$

The estimated variance of  $\hat{N}_i$  is obtained from (2.3) by setting  $l=i$ .

Noting that  $\hat{p}_i$ , given by (2.2), is a combined ratio estimate, the covariance of  $\hat{p}_i$  and  $\hat{p}_l$  is estimated as

$$\text{estcov}(\hat{p}_i, \hat{p}_l) = \hat{N}^{-2} \sum_{f=1}^m \hat{\sigma}_{il}(f). \quad (2.4)$$

Here  $\hat{\sigma}_{il}(f)$ , for a province, is given by (2.3) with  $z_{iht}$  replaced by (see the Appendix):

$$[B_{i(ht)} - \hat{p}_i B_{(ht)}] - \sum_a (a B_{ht} / a \hat{N})(a \hat{N}_i - \hat{p}_i a \hat{N}),$$

where

$$B_{(ht)} = \sum_i B_{i(ht)} \text{ and } a \hat{N} = \sum_i a \hat{N}_i.$$

The estimated covariance of the unadjusted estimates  $p_i^*$  and  $p_l^*$  of  $p_i$  and  $p_l$  is calculated as

$$\text{estcov}(p_i^*, p_l^*) = N^{*-2} \sum_{f=1}^m \sigma_{il}^*(f), \quad (2.5)$$

where  $\sigma_{il}^*(f)$ , for a province, is given by (2.3) with  $z_{iht}$  replaced by  $B_{i(ht)} - p_i^* B_{(ht)}$ .

The estimates  $\hat{N}_i$  and  $\hat{p}_i$  and the associated variances and covariances provide the basic building blocks for constructing both the Wald test statistic and the adjustments to chi-squared tests.

### 3. Simple Goodness-of-Fit

#### 3.1. Concepts, theoretical results, and computational aspects

The goodness-of-fit problem involves testing the hypothesis that the population distribution of a characteristic in a specified domain is the same as a known distribution. For example, one might wish to test the agreement of the age distribution among smokers with the projected census age distribution.

Let  $q_i = N_{(1)i} / N_{(1)}$  be the proportion in a specified domain  $D_1$  belonging to  $i$ -th category, where  $N_{(1)i}$  is the population count in  $D_1$  belonging to  $i$ -th category and  $N_{(1)} = \sum_i N_{(1)i}$  is the domain size ( $q_i > 0, i=1, \dots, I+1; \sum_i q_i = 1$ ). It is convenient here to denote  $\underline{q} = (q_1 \dots q_I)'$ . The null hypothesis  $H_0$  is then given by

$$H_0 : \underline{q} = \underline{q}_0, \quad (3.1)$$

where  $\underline{q}_0 = (q_{01} \dots q_{0I})'$  is the known distribution ( $q_{0,I+1} = 1 - q_{01} - \dots - q_{0I}$ ). Using (2.1) and (2.2) with "i" changed to "(1)i," the estimate of  $\underline{q}$  is obtained as  $\hat{\underline{q}} = (\hat{q}_1 \dots \hat{q}_I)'$ , where

$$\hat{q}_i = \hat{N}_{(1)i} / \hat{N}_{(1)}, \quad i=1, \dots, I+1 \quad (3.2)$$

and  $\sum_i \hat{q}_i = 1$ . The unadjusted estimate of  $q_i$  is given by  $q_i^* = N_{(1)i}^* / N_{(1)}^*$ .

The estimated covariance of  $\hat{q}_i$  and  $\hat{q}_l$  is given by

$$\hat{\sigma}_{il} = \hat{N}_{(1)}^{-2} \sum_{f=1}^m \hat{\sigma}_{il}(f), \quad (3.3)$$

where  $\hat{\sigma}_{il}(f)$ , for a province, is given by (2.3) with  $z_{iht}$  replaced by

$$[B_{(1)i(ht)} - \hat{q}_i B_{(1)ht}] - \sum_a (a B_{ht} / a \hat{N}) [a \hat{N}_{(1)i} - \hat{q}_i a \hat{N}_{(1)}],$$

where  $B_{(1)ht} = \sum_i B_{(1)i(ht)}$  and  ${}_a\hat{N}_{(1)} = \sum_i {}_a\hat{N}_{(1)i}$  ( $i=1, \dots, I+1$ ). Note that  $\hat{\sigma}_{il}$  is computationally simple since it depends only on the first-stage unit totals  $B_{(1)i(ht)}$  and  ${}_aB_{ht}$  where  $B_{(1)i(ht)} = \sum_k w_{htk} y_{(1)i(htk)}$  and  $y_{(1)i(htk)} = 1$  if  $(htk)$ -th unit belongs to  $D_1$  and the  $i$ -th category, zero otherwise. Similarly, the estimated covariance of  $q_i^*$  and  $q_j^*$  is given by

$$\sigma_{il}^* = N_{(1)}^{-2} \sum_{f=1}^m \sigma_{il}^*(f), \tag{3.4}$$

where  $\sigma_{il}^*(f)$ , for a province, is given by (2.3) with  $z_{iht}$  replaced by  $B_{(1)i(ht)} - \hat{q}_i B_{(1)ht}$ . Formula (3.4) should be used if post-stratification adjustment cannot be implemented.

The customary Pearson chi-squared statistic for testing  $H_0$  is given by

$$X_G^2(P) = n_{(1)} \sum_{i=1}^{I+1} (\hat{q}_i - q_{oi})^2 / q_{oi}, \tag{3.5}$$

where  $n_{(1)}$  is the total number of sampled ultimate units in domain  $D_1$ . The statistic  $X_G^2(P)$  is asymptotically distributed as a weighted sum,  $\delta_1 W_1 + \dots + \delta_I W_I$ , of independent  $\chi^2$  variables  $W_i$ , under  $H_0$ , where the weights  $\delta_i$  are the eigenvalues of the "design effects matrix"  $\underline{D}$  (Rao and Scott (1981)). Under multinomial sampling, all the  $\delta_i$  are equal to 1 and  $\sum \delta_i W_i$  reduces to  $\chi_I^2$ , a  $\chi^2$  variable with  $I$  d.f. The estimated weights  $\hat{\delta}_i$  are the eigenvalues of  $\hat{\underline{D}} = n_{(1)} \underline{Q}_o^{-1} \hat{\underline{\Sigma}}$ , where  $\underline{Q}_o = \text{diag}(q_o) - q_o q_o'$  and  $\hat{\underline{\Sigma}} = (\hat{\sigma}_{il})$  is the estimated covariance matrix of  $\hat{\underline{q}}$ .

A first order correction to  $X_G^2(P)$ , requiring only the estimated cell design effects  $\hat{d}_i$ , is given by

$$X_G^2(\hat{\delta}_.) = X_G^2(P) / \hat{\delta}_., \tag{3.6}$$

where

$$I\hat{\delta}_. = \sum_{i=1}^{I+1} \frac{\hat{q}_i}{q_{oi}} (1-\hat{q}_i) \hat{d}_i, \tag{3.7}$$

and

$$\hat{d}_i = \hat{\sigma}_{ii} / [\hat{q}_i (1-\hat{q}_i) n_{(1)}^{-1}]. \tag{3.8}$$

Another first order correction is given by

$$X_G^2(\hat{d}_.) = X_G^2(P) / \hat{d}_., \tag{3.9}$$

(Fellegi (1980)), where  $\hat{d}_. = \sum \hat{d}_i / (I+1)$  is the average cell deff. Both  $X_G^2(\hat{\delta}_.)$  and  $X_G^2(\hat{d}_.)$  are treated as  $\chi_v^2$ , under  $H_0$ .

A second order correction to  $X_G^2$ , based on the Satterthwaite approximation to  $\sum \delta_i W_i$ , is obtained by treating

$$X_G^2(S) = \frac{X_G^2(P)}{\hat{\delta}_. (1 + \hat{C}_\delta^2)} \text{ as } \chi_v^2, v = \frac{I}{1 + \hat{C}_\delta^2}, \tag{3.10}$$

where  $\hat{C}_\delta$  is the estimated coefficient of variation of the  $\hat{\delta}_i$ :

$$\hat{C}_\delta^2 = \frac{I}{1} \frac{\sum \hat{\delta}_i^2 / (I\hat{\delta}_.)^2}{1} - 1,$$

and

$$\sum_{i=1}^I \hat{\delta}_i^2 = n_{(1)}^2 \sum_{i=1}^{I+1} \sum_{l=1}^{I+1} \hat{\sigma}_{il}^2 / (\hat{q}_i \hat{q}_l).$$

The second order correction takes account of the variability in the  $\delta_i$ , unlike the first order corrections (3.6) and (3.9). Note also that both  $\hat{\delta}_.$  and  $\hat{C}_\delta$  can be calculated without evaluating the individual eigenvalues  $\hat{\delta}_i$ . If the post-stratification adjustment is not made, then the adjusted test statistics are obtained from the above formulae by simply replacing  $\hat{\sigma}_{il}$  with  $\sigma_{il}^*$  given by (3.4) and by changing  $\hat{q}_i$  to  $q_i^*$  and  $\hat{d}_i$  to  $d_i^* = \sigma_{ii}^* / [q_i^* (1-q_i^*) n_{(1)}^{-1}]$ .

It is convenient to use the critical point  $\chi_v^2(\alpha)$ , the customary upper  $\alpha$ -point of  $\chi_v^2$ , rather than  $\chi_v^2(\alpha)$ , in which case  $X_G^2(S)$  should be modified to

$$X_G^2(S, \alpha) = X_G^2(S) [\chi_v^2(\alpha) / \chi_v^2(\alpha)]. \tag{3.11}$$

The null hypothesis is rejected at the  $\alpha$ -level if  $X_G^2(S, \alpha)$  exceeds  $\chi^2_I(\alpha)$ . Assuming that the Satterthwaite approximation is accurate, i.e.,  $Pr[X_G^2(S, \alpha) > \chi^2_I(\alpha) | H_0] \doteq \alpha$ , the type I error rates of customary  $X^2$  and the first order correction  $X_G^2(\hat{\delta}_.)$  are estimated as  $Pr[\chi^2_v \geq \chi^2_I(\alpha) / \{\hat{\delta}_. (1 + \hat{C}^2_{\hat{\delta}})\}]$  and  $Pr[\chi^2_v \geq \chi^2_I(\alpha) / (1 + \hat{C}^2_{\hat{\delta}})]$  respectively, for nominal level  $\alpha$ . Similarly, the type I error rate of Fellegi's correction  $X_G^2(\hat{d}_.)$  is estimated as  $Pr[\chi^2_v \geq \{\hat{d}_. \chi^2_I(\alpha)\} / \{\hat{\delta}_. (1 + \hat{C}^2_{\hat{\delta}})\}]$ .

A test which is asymptotically a  $\chi^2_I$  under  $H_0$  is given by the Wald statistic

$$X_G^2(W) = (\hat{q} - q_o)' \hat{\Sigma}_q^{-1} (\hat{q} - q_o), \tag{3.12}$$

where  $\hat{\Sigma}_q$  is the estimated covariance matrix of  $\hat{q}$  with elements  $\hat{\sigma}_{ii}$ . If the degrees of freedom,  $r$ , for  $\hat{\Sigma}$  is not large, an improvement to  $X_G^2(W)$  is obtained by treating

$$F_G(W) = \frac{(r-I+1)}{rI} X_G^2(W) \tag{3.13}$$

as an  $F$  variable with  $I$  and  $r-I+1$  d.f. respectively, under  $H_0$  (Fellegi (1980), Hidiroglou et al. (1980)). 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,  $\hat{q}_i - q_{oi}$ , is useful for detecting deviations from  $H_0$ . The standardized residuals

$$\hat{e}_i = e_i / \hat{d}_i^{1/2}, i=1, \dots, I+1 \tag{3.14}$$

are approximately  $N(0,1)$  under  $H_0$ , where the  $e_i$  denote the standardized residuals under the assumption of simple random sampling:

$$e_i = (\hat{q}_i - q_{oi}) / [\hat{q}_i (1 - \hat{q}_i) / n_{(1)}]^{1/2}. \tag{3.15}$$

Deviations from  $H_0$  are indicated by cells with large  $|\hat{e}_i|$ -values. Ignoring the design, and hence using the  $e_i$ , could be misleading if the cell deffs  $\hat{d}_i$  are large.

3.2. Example

Suppose that we wish to test the agreement of age distribution ( $q_i$ ) among those consuming 1–6 drinks per week ( $D_1$ ) with the projected census age distribution ( $q_{oi}$ ), at the national level. Table 1 gives  $q_{oi}$ , the adjusted and the unadjusted estimated cell proportions  $\hat{q}_i$  and  $q_i^*$ , and the corresponding estimated cell deffs  $\hat{d}_i$  and  $d_i^*$ , for  $I+1=7$  age categories.

Table 1. Age Distribution Among those Consuming 1–6 Drinks per Week (National Level:  $n_{(1)}=5,204$ ): Cell Proportions and Cell Design Effects

	Age						
	15–19	20–24	25–34	35–44	45–54	55–64	65+
$q_{oi}$	0.133	0.127	0.218	0.152	0.140	0.115	0.115
$\hat{q}_i$	0.117	0.150	0.265	0.175	0.148	0.093	0.053
$\hat{d}_i$	1.36	1.17	2.07	1.06	0.60	1.09	0.98
$q_i^*$	0.120	0.138	0.265	0.182	0.153	0.090	0.051
$d_i^*$	2.58	2.44	7.02	1.66	3.61	2.14	2.70

It is clear from Table 1 that the post-stratification adjustment has led to a substantial reduction in cell deffs, the average deff  $\hat{d}_i$  of the  $\hat{q}_i$  being 1.19 compared to  $d^*=3.16$ , the average deff of the  $q_i^*$ . Also, the  $\hat{d}_i$  vary considerably across the categories, ranging from 0.6 to 2.2 (similarly the  $d_i^*$  vary from 1.66 to 7.02). The calculated values of  $\hat{\delta}_i$  and  $\hat{C}_\delta$  were 1.14 and 0.67 respectively, while  $\delta^*=3.04$  and  $C_\delta^*=0.92$ .

The value of  $X_G^2(P)=298$  is so large here, compared to  $\chi_6^2(0.05)=12.6$ , that it is not necessary to do any correction to  $X_G^2(P)$  to conclude that  $H_0 : q=q_0$  is not tenable at  $\alpha=0.05$  level. We have, however, given the values of corrected  $X^2$  and the Wald statistic below for comparison. In the adjusted for age-sex case, we obtained the following values:

$$X_G^2(P)=298, X_G^2(\hat{\delta}_i)=261, X_G^2(\hat{d}_i)=243,$$

with estimated type I error rates 0.115, 0.076 and 0.065 respectively ( $\alpha=0.05$ ), and

$$X_G^2(S, 0.05)=234, X_G^2(W)=513.$$

Taking  $r$  as  $100-31=69$ , the value of  $F_G(W)$  is calculated as 79 and compared to  $F_{6,64}(0.05)=2.25$ . The values of  $X_G^2(S, 0.05)$  and  $X_G^2(W)$  are much larger than the upper 5% point of  $\chi_6^2$ ,  $\chi_6^2(0.05)=12.6$ , so that the null hypothesis is not tenable. The effect of sample design, after post-stratification

adjustment, is not substantial in this example.

In the unadjusted for age-sex, we obtained the following values:

$$X_G^2(P)=315, X_G^2(\delta^*)=104, X_G^2(d^*)=100,$$

with estimated type I error rates 0.563, 0.095 and 0.085 respectively ( $\alpha=0.05$ ), and

$$X_G^2(S, 0.05)=85, X_G^2(W)=281.$$

The effect of sample design is substantial without post-stratification adjustment since the type I error rate is reduced from 0.60 to about 0.09 by using a first order correction to  $X_G^2(P)$ . Post-stratification adjustment clearly provides a more powerful test: compare  $X_G^2(S, 0.05)=229$  in the adjusted case to  $X_G^2(S, 0.05)=86$  in the unadjusted case. The Wald statistic seems to be somewhat unreliable. Its value 281 is closer to the uncorrected  $X^2$  value of 315, despite the substantial design effect in the unadjusted case ( $d^*=3.2$ ).

The standardized residuals  $\hat{e}_i$  and  $e_i$  for the age-sex adjusted case are given in Table 2.

The residuals  $e_i$  under the assumption of simple random sampling are comparable to the  $\hat{e}_i$  since the deffs here are small in the adjusted case. Inspection of the  $\hat{e}_i$  clearly shows a large deviation from  $H_0$  in most of the age categories, especially in the 65+ age group with  $\hat{q}_i$  much smaller than the corresponding  $q_{oi}$ .

Table 2. Standardized Residuals  $\hat{e}_i$  and  $e_i$

	Age						
	15-19	20-24	25-34	35-44	45-54	55-64	65+
$\hat{e}_i$	-3.06	4.28	5.06	4.20	2.09	-5.32	-19.67
$e_i$	-3.59	4.65	7.53	4.36	1.62	-5.46	-19.96



#### 4. Homogeneity in a Two-Way Table

##### 4.1. Concepts, theoretical results, and computational aspects

The test of the homogeneity problem involves testing of the hypothesis that the  $I+1$  category proportions, in a specified domain  $D_1$ , are homogeneous across  $R+1$  regions (provinces). Extending the notation of Section 3 by adding a subscript “ $r$ ” ( $r=1, \dots, R+1$ ) to denote a region, the null hypothesis of interest is

$$H_o : q_1 = q_2 = \dots = q_{R+1}. \quad (4.1)$$

A Pearson chi-squared statistic for testing  $H_o$  is given by

$$X_H^2(P) = \sum_{r=1}^{R+1} n_{r(1)} \sum_{i=1}^{I+1} (\hat{q}_{ri} - \hat{q}_{+i})^2 / \hat{q}_{+i}, \quad (4.2)$$

where  $\hat{q}_{+i} = \sum_r n_{r(1)} \hat{q}_{ri} / n_{(1)}$  with  $n_{(1)} = \sum_r n_{r(1)}$ . The statistic  $X_H^2(P)$  is asymptotically distributed as a weighted sum,  $\delta_1 W_1 + \dots + \delta_{RI} W_{RI}$ , of independent  $\chi_1^2$  variables  $W_i$ , under  $H_o$ , where the weights  $\delta_s$  are estimated by  $\hat{\delta}_s$ , the eigenvalues of estimated deff matrix

$$\hat{D}_H = n_{(1)} (F \otimes \hat{Q}^{-1}) \hat{\Delta} \quad (4.3)$$

(Scott and Rao (1981)). Here

$$F = \text{diag}(f) - f f', f = (n_{1(1)}/n_{(1)} \dots n_{R(1)}/n_{(1)})',$$

$$\hat{Q} = \text{diag}(\hat{q}) - \hat{q} \hat{q}', \hat{q} = (\hat{q}_{+1} \dots \hat{q}_{+R})',$$

and

$$\hat{\Delta} = \bigoplus_{r=1}^R \hat{\Sigma}_r + \hat{\Sigma}_{R+1} \otimes J_R,$$

where  $\hat{\Sigma}_r = [\hat{\sigma}_{rij}]$  is estimated covariance matrix of  $\hat{q}_r = (q_{r1} \dots q_{rI})'$  obtained from (3.3), using the sample data from the  $r$ -th region,  $J_R$  is an  $R \times R$  matrix of 1's and  $\oplus$  and  $\otimes$  respectively denote the direct sum and direct product operators. Note that the estimates  $\hat{q}_1, \dots, \hat{q}_{R+1}$  are stochastically independent since the samples within regions are drawn independently.

As before, the corrections to  $X_H^2(P)$  can be obtained without evaluating the individual eigenvalues,  $\hat{\delta}_s$ . A first order correction to  $X_H^2(P)$ , requiring only the estimated cell deffs  $\hat{d}_{ri}$ , in each region, is given by

$$X_H^2(\hat{\delta}_s) = X_H^2(P) / \hat{\delta}_s, \quad (4.4)$$

where

$$IR \hat{\delta}_s = \sum_{r=1}^{R+1} (1-f_r) \left\{ \sum_{i=1}^{I+1} \frac{\hat{q}_{ri}}{\hat{q}_{+i}} (1-\hat{q}_{ri}) \hat{d}_{ri} \right\},$$

with

$$f_r = n_{r(1)} / n_{(1)},$$

and

$$\hat{d}_{ri} = \hat{\sigma}_{rii} / [\hat{q}_{ri} (1-\hat{q}_{ri}) n_{r(1)}^{-1}].$$

The statistic  $X_H^2(\hat{\delta}_s)$  is treated as a  $\chi_{IR}^2$  under  $H_o$ . The second order corrections  $X_H^2(S)$  and  $X_H^2(S, \alpha)$  are obtained from (3.10) and (3.11) with  $X_G^2(P)$  replaced by  $X_H^2(P)$ ,  $I$  by  $IR$  and  $\Sigma \hat{\delta}_i^2$  by  $\text{tr}(\hat{D}_H)$ , where  $\hat{D}_H$  is given by (4.3) and “tr” denotes the trace operator.

A Wald statistic, which is asymptotically a  $\chi_{IR}^2$  under  $H_o$ , is given by

$$X_H^2(W) = \hat{q}_H' \hat{\Delta}^{-1} \hat{q}_H, \quad (4.5)$$

where

$$\hat{q}_H = (\hat{q}_1' - \hat{q}_{R+1}' \dots \hat{q}_R' - \hat{q}_{R+1}').$$

The standardized residuals

$$\hat{e}_{ri} = (\hat{q}_{ri} - \hat{q}_{+i}) / \{ \text{var}(\hat{q}_{ri} - \hat{q}_{+i}) \}^{1/2}, \quad r=1, \dots, R+1$$
$$i=1, \dots, I+1 \tag{4.6}$$

are approximately  $N(0,1)$  under  $H_o$ , where

$$n_{(1)}^2 \text{var}(\hat{q}_{ri} - \hat{q}_{+i}) =$$
$$[\hat{q}_{+i} (1 - \hat{q}_{+i})] [n_{(1)} (n_{(1)} - 2n_{r(1)}) \hat{d}_{ri} / n_{r(1)}$$
$$+ \sum_{s=1}^{R+1} n_{s(1)} \hat{d}_{si}].$$

Analogous results are obtained from  $\Sigma_r^*$  if adjustment for post-stratification is not used,

where  $\Sigma_r^*$  is obtained from (3.4) using sample data from the  $r$ -th region.

4.2. Example

Suppose that we wish to test the homogeneity of age distribution among current smokers ( $D_1$ ) across the following  $R+1=4$  regions in Canada: Atlantic Provinces, Quebec, Ontario, and Western Provinces. Table 3 gives the adjusted and unadjusted age distributions among current smokers in the four regions, i.e.,  $\hat{q}_{ri}$  and  $q_{ri}^*$ ,  $r=1, \dots, 4$  and  $i=1, \dots, 7$ . Here  $n_{1(1)}=1\,991$ ,  $n_{2(1)}=1\,919$ ,  $n_{3(1)}=1\,415$ ,  $n_{4(1)}=2\,574$  ( $\Sigma n_{r(1)}=n_{(1)}=7\,899$ ) and the age categories ( $I+1=7$ ) are the same as in the example of Section 3.2.

Table 3. Age Distribution Among Current Smokers in the Four Regions in Canada: Cell Proportions

	Age						
	15-19	20-24	25-34	35-44	45-54	55-64	65+
Atlantic							
$\hat{q}_{1i}$	0.149	0.135	0.217	0.138	0.122	0.114	0.123
$q_{1i}^*$	0.162	0.115	0.184	0.159	0.132	0.124	0.124
Quebec							
$\hat{q}_{2i}$	0.135	0.129	0.223	0.157	0.142	0.110	0.103
$q_{2i}^*$	0.144	0.128	0.199	0.167	0.159	0.116	0.087
Ontario							
$\hat{q}_{3i}$	0.129	0.122	0.215	0.156	0.145	0.116	0.116
$q_{3i}^*$	0.132	0.113	0.217	0.157	0.145	0.108	0.129
West							
$\hat{q}_{4i}$	0.132	0.127	0.216	0.148	0.136	0.118	0.124
$q_{4i}^*$	0.127	0.114	0.247	0.152	0.136	0.109	0.114

The values of  $\hat{d}_\cdot$ ,  $\hat{\delta}_\cdot$  and  $\hat{C}_\delta$  were computed as 0.96, 0.96, and 0.78 respectively, while  $d^*=3.3$ ,  $\delta^*=3.2$ , and  $C_\delta^*=1.4$ . The post-stratification adjustment has led to small deffs ( $\hat{d}_\cdot=0.96$  vs.  $d^*=3.3$ ) so that any correction of  $X_{ij}^2(P)$  is not necessary here. We have,

however, given the values of corrected  $X^2$  and the Wald statistic below for comparison. The type I error rates were estimated as in Section 3, assuming that the Satterthwaite approximation is accurate.

In the adjusted for age-sex case, we obtained the following values:

$$X_H^2(P)=30.8, X_H^2(\hat{\delta}_.)=32.1 = X_H^2(\hat{d}.) \\ = X_H^2(P)/\hat{d}.$$

with estimated type I error rates 0.07 and 0.09 respectively ( $\alpha=0.05$ ) and

$$X_H^2(S, 0.05)=29.0, X_H^2(W)=75.3.$$

Note that the type I error rate for the first order correction, in fact, is slightly larger than that for the uncorrected  $X^2$  since  $\hat{\delta}_.$  is close to 1 and  $\hat{C}_\delta$  is large. In the unadjusted for age-sex case, we obtained the following values:

$$X_H^2(P)=41.5, X_H^2(\delta^*)=13.5, X_H^2(d^*)=12.5$$

with estimated type I error rates 0.82, 0.14, and 0.12 respectively ( $\alpha=0.05$ ), and

$$X_H^2(S, 0.05)=10.0, X_H^2(W)=15.6.$$

Both the first order corrections reduced the type I error rate dramatically in the unadjusted case (0.82 to about 0.14), but they are not totally satisfactory here due to a large value of  $C_\delta^*$ . Post-stratification has provided a more powerful test of  $H_0$ . In fact,  $X_H^2(S, 0.05)=10.0 < \chi_{18}^2(0.05)=28.9$  in the unadjusted case indicating the tenability of  $H_0$ , while  $X_H^2(S, 0.05)=29.0$  in the adjusted case indicating some evidence against  $H_0$ .

Turning to the residuals, only one residual  $e_{11}^*=3.51$  (out of 28), corresponding to the cell (Atlantic Provinces, 15–19) in the two-way table, indicated significant deviation from  $H_0$  in the unadjusted case. On the other hand, three out of seven residuals  $\hat{e}_{1i}$  for the Atlantic region exceeded 3.0 in absolute value under post-stratification adjustment. Therefore it appears that the age distribution among smokers in the Atlantic region might

be significantly different from the corresponding distributions in the remaining three regions.

## 5. Independence in a Two-Way Table

### 5.1. Concepts, theoretical results, and computational aspects

Suppose that a two-way table has  $I+1$  rows (variable  $A$ ) and  $J+1$  columns (variable  $B$ ) and  $p_{ij} = N_{ij}/N$  denotes the population proportion in the  $(i,j)$ -th cell, where  $N_{ij}$  is the total count in the  $(i,j)$ -th cell and  $N = \sum_i \sum_j N_{ij}$ ,  $i=1, \dots, I+1; j=1, \dots, J+1$ . Let  $\hat{p}_{ij} = \hat{N}_{ij}/\hat{N}$  be the sample estimate of  $p_{ij}$ , obtained from (2.2) with “ $i$ ” replaced by “ $ij$ ”. Finally, let  $\hat{p}_{i+} = \sum_j \hat{p}_{ij}$  and  $\hat{p}_{+j} = \sum_i \hat{p}_{ij}$  be the sample estimates of marginal proportions  $p_{i+}$  and  $p_{+j}$  respectively, and  $\hat{p}_{i+} = (\hat{p}_{1+} \dots \hat{p}_{I+})'$ ,  $\hat{p}_{+j} = (\hat{p}_{+1} \dots \hat{p}_{+J})'$ .

The hypothesis of independence is given by

$$H_0: p_{ij} = p_{i+} p_{+j}, i=1, \dots, I+1; j=1, \dots, J+1. \quad (5.1)$$

The Pearson statistic for testing  $H_0$  is

$$X_I^2(P) = n \sum_{i=1}^{I+1} \sum_{j=1}^{J+1} (\hat{p}_{ij} - \hat{p}_{i+} \hat{p}_{+j})^2 / (\hat{p}_{i+} \hat{p}_{+j}), \quad (5.2)$$

where  $n$  is the total sample size. If the hypothesis of independence is to be tested within a domain  $D_1$ , then  $n$  and  $\hat{p}_{ij}$  in (5.2) should be replaced by  $n_{(1)}$  and  $\hat{q}_{ij} = \hat{N}_{(1)ij} / \hat{N}_{(1)}$  respectively. The statistic  $X_I^2(P)$  is asymptotically distributed as a weighted sum,  $\delta_1 w_1 + \dots + \delta_J w_J$ , of independent  $\chi^2$  variables  $W_s$ , under  $H_0$ , where the weights  $\delta_s$  are estimated by  $\hat{\delta}_s$ , the eigenvalues of estimated deff matrix

$$\hat{D}_I = n(\hat{P}_{+}^{-1} \otimes \hat{P}_{+}^{-1}) \hat{\Gamma}, \quad (5.3)$$

(Rao and Scott (1981)). Here

$$\hat{P}_{I+} = \text{diag}(\hat{p}_{I+}) - \hat{p}_{I+} \hat{p}'_{I+}, \hat{P}_{+J} = \text{diag}(\hat{p}_{+J}) - \hat{p}_{+J} \hat{p}'_{+J}$$

and  $\hat{\Gamma}$  is the estimated covariance matrix of  $\hat{h}_{ij} = \hat{p}_{ij} - \hat{p}_{i+} \hat{p}_{+j}$  ( $i=1, \dots, I; j=1, \dots, J$ ) obtained from (2.4) with  $z_{iht}$  replaced by (see the Appendix)

$$z_{iht} = \{B_{ij(ht)} - \hat{p}_{i+} B_{+j(ht)} - \hat{p}_{+j} B_{i+(ht)} + \hat{p}_{i+} \hat{p}_{+j} B_{(ht)}\} - \Sigma_a ({}_a B_{ht} / {}_a \hat{N})$$

$$\{{}_a \hat{N}_{ij} - \hat{p}_{i+} {}_a \hat{N}_{+j} - \hat{p}_{+j} {}_a \hat{N}_{i+} + \hat{p}_{i+} \hat{p}_{+j} {}_a \hat{N}\}, (5.4)$$

with obvious extension of notation, where  $B_{i+(ht)} = \sum_{j=1}^{J+1} B_{ij(ht)}$ , etc. Note that the elements of  $\hat{\Gamma}$ , say  $\hat{\gamma}_{ij, i'j'}$ , depend only on the first-stage unit totals  $B_{ij(ht)}$  and their marginal totals.

As before, the corrections to  $X^2_I(P)$  can be obtained without evaluating the individual eigenvalues,  $\hat{\delta}_s$ . A first order correction, requiring only the estimated cell deffs,  $\hat{d}_{ij}$ , and the estimated deffs of margins,  $\hat{d}_{A(i)}$  and  $\hat{d}_{B(j)}$ , is given by

$$X^2_I(\hat{\delta}_s) = X^2_I(P) / \hat{\delta}_s, (5.5)$$

where

$$IJ\hat{\delta}_s = \sum_{i=1}^{I+1} \sum_{j=1}^{J+1} \frac{\hat{p}_{ij}(1-\hat{p}_{ij})}{\hat{p}_{i+} \hat{p}_{+j}} \hat{d}_{ij}$$

$$- \sum_{i=1}^{I+1} (1-\hat{p}_{i+}) \hat{d}_{A(i)}$$

$$- \sum_{j=1}^{J+1} (1-\hat{p}_{+j}) \hat{d}_{B(j)},$$

$$\hat{d}_{ij} = \text{estvar}(\hat{p}_{ij}) / [\hat{p}_{ij}(1-\hat{p}_{ij}) n^{-1}],$$

and

$$\hat{d}_{A(i)} = \text{estvar}(\hat{p}_{i+}) / [\hat{p}_{i+}(1-\hat{p}_{i+}) n^{-1}], \hat{d}_{B(j)} =$$

$$\text{estvar}(\hat{p}_{+j}) / [\hat{p}_{+j}(1-\hat{p}_{+j}) n^{-1}].$$

The correction (5.5) can be simply implemented from published tables providing the cell deffs and marginal deffs. The statistic  $X^2_I(\hat{\delta}_s)$  is treated as a  $\chi^2_{IJ}$  under  $H_0$ .

The second order corrections  $X^2_I(S)$  and  $X^2_I(S, \alpha)$  are obtained from (3.10) and (3.11) with  $X^2_O(P)$  replaced by  $X^2_I(P)$ ,  $I$  by  $IJ$  and  $\Sigma \delta_i^2$  by  $n^2 \Sigma \hat{\gamma}_{ij, i'j'}' / (\hat{p}_{i+} \hat{p}_{+j}) (\hat{p}'_{i+} \hat{p}'_{+j})$ , where the summation is taken over  $i, i' = 1, \dots, I+1$  and  $j, j' = 1, \dots, J+1$ .

A Wald statistic can be constructed from the deviations  $\hat{h}_{ij}$  as follows:

$$X^2_I(W1) = \hat{h}' \hat{\Gamma}^{-1} \hat{h}, (5.6)$$

where  $\hat{h} = (\hat{h}_{11} \dots \hat{h}_{IJ}; \dots; \hat{h}_{1I} \dots \hat{h}_{IJ})'$  with associated covariance matrix  $\hat{\Gamma}$ . Under  $H_0$ ,  $X^2_I(W1)$  is asymptotically a  $\chi^2_{IJ}$ . An alternative Wald statistic is obtained by expressing  $H_0$  as  $H_0 : u_{12(ij)} = 0$  for all  $(i, j)$  in the saturated loglinear model

$$\ln p_{ij} = \mu_{ij} = \bar{u} + u_{1(i)} + u_{2(j)} + u_{12(ij)},$$

$$i=1, \dots, I+1; j=1, \dots, J+1, (5.7)$$

where the parameters  $u_{1(i)}$ ,  $u_{2(j)}$  and  $u_{12(ij)}$  are constrained by  $\Sigma_i \mu_{1(i)} = 0$ ,  $\Sigma_j \mu_{2(j)} = 0$ ,  $\Sigma_i \mu_{12(ij)} = 0$  for all  $j$ ,  $\Sigma_j \mu_{12(ij)} = 0$  for all  $i$ , and  $\bar{u}$  is a normalizing factor to ensure that  $\Sigma_i \Sigma_j p_{ij} = 1$ . In matrix notation, (5.7) may be expressed as

$$\underline{\mu} = \bar{u} \underline{1} + X_1 \underline{\theta}_1 + X_2 \underline{\theta}_2; X_2' X_1 = 0, X_2' \underline{1} = 0, (5.8)$$

where  $\underline{\mu}$  is the  $(I+1)(J+1)$ -vector of the  $\mu_{ij}$  (in lexicographical order),  $\underline{\theta}_1$  is the  $(I+J)$ -vector of parameters  $u_{1(1)}, \dots, u_{1(I)}; u_{2(1)}, \dots, u_{2(J)}$  with associated model matrix  $X_1$  consisting of +1's, 0's and -1's,  $\underline{\theta}_2$  is the  $(IJ)$ -vector of parameters  $u_{12(11)}, \dots, u_{12(1J)}; \dots; u_{12(I1)}, \dots, u_{12(IJ)}$  with associated model matrix  $X_2$  similar to  $X_1$ , and  $\underline{1}$  is the vector of 1's. Noting that  $H_0 : \underline{\theta}_2 = \underline{0}$  is equivalent to testing  $\phi = X_2' \underline{\mu} = \underline{0}$ , the alternative Wald statistic is given by

$$X^2_1(W_2) = \hat{\phi}' \hat{\Sigma}^{-1}_{\hat{\phi}} \hat{\phi}, \tag{5.9}$$

where  $\hat{\Sigma}_{\hat{\phi}}$  is the estimated covariance matrix of  $\hat{\phi} = \hat{X}'_2 \hat{\mu}$ :

$$\hat{\Sigma}_{\hat{\phi}} = X'_2 D_{\hat{p}}^{-1} \hat{\Sigma} D_{\hat{p}}^{-1} X_2.$$

Here  $\hat{\mu}$  is the vector of logprobabilities  $\hat{\mu}_{ij} = \ln \hat{p}_{ij}$ ,  $D_{\hat{p}} = \text{diag}(\hat{p})$ ,  $\hat{p}$  is the  $(I+1)(J+1)$ -vector of estimated cell proportions  $\hat{p}_{ij}$  and  $\hat{\Sigma}$  is the estimated covariance matrix of  $\hat{p}$  obtained from (2.4). Under  $H_0$ ,  $X^2_1(W_2)$  is also asymptotically a  $\chi^2_{IJ}$ .

Analogous results are obtained in the unadjusted case from  $\Gamma^*$ , where  $\Gamma^*$  is obtained from (2.5) with  $z_{iht}$  replaced by  $B_{ij(hi)} - p^*_{i+} B_{+j(hi)} - p^*_{+j} B_{i+(ht)} + p^*_{i+} p^*_{+j} B_{(ht)}$ . Similarly, for testing independence within a domain  $D_1$ , a subscript (1) is added to terms in (5.4) involving the subscripts  $i$  or  $j$  or both, and  $n$  and  $\hat{p}_{ij}$  are replaced by  $n_{(1)}$  and  $\hat{q}_{ij}$  respectively.

The standardized residuals

$$\hat{e}_{ij} = e_{ij} / \{\hat{d}_{ij}(h)\}^{1/2} \tag{5.10}$$

are approximately  $N(0,1)$  under  $H_0$ , where

$$e_{ij} = \hat{h}_{ij} / [\hat{p}_{i+} \hat{p}_{+j} (1-\hat{p}_{i+}) (1-\hat{p}_{+j}) n^{-1}]^{1/2},$$

are the standardized residuals under the assumption of simple random sampling (Haberman, (1973)) and  $\hat{d}_{ij}(h) = \hat{\gamma}_{ij,ij} / [\hat{p}_{i+} \hat{p}_{+j} (1-\hat{p}_{i+}) (1-\hat{p}_{+j}) n^{-1}]^{1/2}$  is the estimated deff of  $\hat{h}_{ij}$  under  $H_0$ .

5.2. Example

The estimated proportions,  $\hat{p}_{ij}$ , given in Table 2, are cross-classified by drug use (four categories: 0,1,2,3 + drug variates in a two-day period) and sex (male, female). We test the hypothesis of independence in this two-way table.

Table 4. Estimated Cell Proportions in a 4×2 Table (Canada Level): Drug Use × Sex (n=31 668)

Number of drug varieties	Sex		Total
	Male	Female	
0	0.2936	0.2277	0.5213
1	0.1338	0.1589	0.2917
2	0.0478	0.0725	0.1203
3+	0.0207	0.0450	0.0657
Total	0.4959	0.5041	1.0000

We present here results only for the age-sex adjusted case. The value of Pearson statistic  $X^2_1(P)=774$  is so large here, compared to  $\chi^2_3(0.05)=7.8$  ( $IJ=3$ ), that it is not necessary to do any correction to  $X^2_1(P)$  to conclude that the hypothesis of independence is not tenable at  $\alpha=0.05$  level. We have, however, given the values of corrected  $X^2$  and the Wald statistics below for comparison:

$$\hat{d}_{..} = \Sigma \Sigma \hat{d}_{ij} / (I+1)(J+1) = 2.37, \hat{\delta}_{..} =$$

$$1.77, \hat{C}_{\delta} = 0.47$$

and

$$X^2_1(P) = 774, X^2_1(\hat{\delta}_{..}) = 437, X^2_1(\hat{d}_{..}) =$$

$$X^2_1(P) / \hat{d}_{..} = 327$$

with estimated type I error rates 0.226, 0.062, and 0.023 respectively ( $\alpha=0.05$ ), and

$$X^2_1(S,0.05) = 408, X^2_1(W1) = 538,$$

$$X^2_1(W2) = 617.$$

The effect of survey design is quite substantial since the estimated type I error rate is reduced from 0.226 to 0.062 by using the first order correction,  $X^2_1(\hat{\delta}_{..})$ . The Fellegi correction,  $X^2_1(\hat{d}_{..})$ , is somewhat conservative: estimated type I error rate of 0.023 compared to

nominal level 0.05. Note that  $X^2_1(\hat{d}..)$  does not depend on  $H_o$  unlike other corrections. The Satterthwaite correction and both Wald statistics clearly indicate a strong association between drug use and sex, although we should bear in mind the possibility of extraneous association due to collapsing over other related variables. The  $F$ -versions of Wald statistics are obtained by multiplying  $X^2_1(W1)$  and  $X^2_1(W2)$  by the factor  $(r-IJ+1)/(rIJ) = (69-3+1)/(69 \times 3) = 0.3236$  and treating the resulting statistics as  $F$  variables with  $IJ$  and  $r-IJ+1$  d.f. respectively, under  $H_o$ . The resulting  $F$ -values  $F_I(W1) = 174.1$  and  $F_I(W2) = 199.7$  are compared to  $F_{3,67}(0.05) \doteq 2.74$ , again indicating a strong association.

The design-based standardized residuals,  $\hat{e}_{ij}$ , are given in Table 5. All the  $\hat{e}_{ij}$  are very large in absolute value (note that  $\hat{e}_{i1} + \hat{e}_{i2} = 0$ ), and we may conclude

Table 5. Standardized Residuals  $\hat{e}_{ij}$

$i \backslash j$	1	2
1	19.7	-19.7
2	-7.7	7.7
3	-7.6	7.6
4	-16.4	16.4

from the large positive values  $\hat{e}_{22} = 7.7$ ,  $\hat{e}_{32} = 7.6$  and  $\hat{e}_{42} = 16.4$ , that a significantly greater proportion of women use drugs compared to men.

Since the independence hypothesis (the model (5.7) with  $u_{12(ij)} = 0$ ) is rejected and since the drug use variable is ordinal, it may be possible to find a more complex nonsaturated loglinear model that provides an adequate fit to the data (see Agresti (1983)). First order and second order corrections to the resulting Pearson statistic can be obtained using the general results of Rao and Scott (1984), but we have not investigated this extension.

6. References

Agresti, A. (1983): A Survey of Strategies for Modelling Cross-Classifications Having Ordinal Variables. *Journal of the American Statistical Association*, 78, pp. 184-198.

Bedrick, E.J. (1983): Adjusted Goodness-of-Fit Tests for Survey Data. *Biometrika*, 70, pp. 591-595.

Fay, R.E. (1979): On Adjusting the Pearson Chi-Square Statistic for Cluster Sampling. *Proceedings of the American Statistical Association, Social Statistics Section*, pp. 402-405.

Fay, R.E. (1985): A Jack-Knifed Chi-Squared Test for Complex Samples. *Journal of the American Statistical Association*, 80, pp. 148-157.

Fellegi, I.P. (1980): On Adjusting the Pearson Chi-Square Statistic for Cluster Sampling. *Journal of the American Statistical Association*, 71, pp. 665-670.

Gross, W.F. (1984): A Note on "Chi-Squared Tests with Survey Data." *Journal of the Royal Statistical Society, Series B*, 46, pp. 270-272.

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

Hidirolou, M.A., Fuller, W.A., and Hickman, R.D. (1980): MINI CARP: A Program for Estimating Simple Descriptive Statistics and Their Variances for Multi-Stage Stratified Designs. Iowa State University, Ames, Iowa.

Koch, G.G., Freeman, D.H. Jr., and Freeman, J.L. (1975): Strategies in the Multivariate Analysis of Data from Complex Surveys. *International Statistical Review*, 43, pp. 59-78.

Nathan, G. (1975): Tests of Independence in Contingency Tables from Stratified Proportional Samples. *Sankhya, Series C*, 37, pp. 77-87.

Rao, J.N.K. and Scott, A.J. (1979): Chi-

- Squared Tests for Analysis of Categorical Data from Complex Surveys. Proceedings of the American Statistical Association, Section on Survey Research Methods, pp. 58–66.
- 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. *Journal of the American Statistical Association*, 76, pp. 221–230.
- Rao, J.N.K. and Scott, A.J. (1982): On Chi-Squared Tests for Multi-Way Tables with Cell Proportions Estimated from Survey Data. Paper presented at the Israeli Statistical Association International Meetings on Analysis of Sample Survey Data and on Sequential Analysis, Jerusalem, June 14–18, 1982.
- 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.
- Satterthwaite, F.E. (1946): An Approximate Distribution of Estimates of Variance Components. *Biometrics*, 2, pp. 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.). Academic Press, New York, pp. 247–266.
- Thomas, D.R. and Rao, J.N.K. (1984): A Monte Carlo Study of Exact Levels of Goodness-of-Fit Statistics Under Cluster Sampling. Proceedings of the American Statistical Association, Section on Survey Research Methods, pp. 207–211.
- The Health of Canadians (1981): Report of the Canada Health Survey. Health and Welfare and Statistics Canada, Catalogue 82–538E.
- Williams, W.H. (1962): The Variance of an Estimator with Post-Stratified Weighting. *Journal of the American Statistical Association*, 57, pp. 622–627.

Received January 1986  
Revised April 1987

# Appendix

## Derivation of Estimates of Variances and Covariances

At the provincial level the unadjusted estimate  $N_i^* = \sum_h \sum_t \{ \sum_k w_{htk} y_{i(htk)} \} = \sum_h \sum_t B_{i(ht)}$ . If the first-stage units are assumed to be sampled with replacement within strata, then it is well-known that

$$\text{estcov}(N_i^*, N_l^*) = \sum_{h=1}^L \frac{r_h}{r_h - 1} \sum_{t=1}^{r_h} (B_{i(ht)} - \bar{B}_{i(h)})(B_{l(ht)} - \bar{B}_{l(h)}), \quad (\text{A.1})$$

where  $\bar{B}_{i(h)} = \sum_t B_{i(ht)} / r_h$ . The estimated covariance of the post-stratified estimators  $\hat{N}_i$  and  $\hat{N}_l$  is simply obtained from (A.1) by changing  $y_{i(htk)}$  to  $y_{i(htk)} - \sum_a (a \hat{N}_i / a \hat{N}) a x_{htk}$ , i.e.,  $B_{i(ht)}$  is replaced by

$$\sum_k w_{htk} [y_{i(htk)} - \sum_a (a \hat{N}_i / a \hat{N}) a x_{htk}] = B_{i(ht)} - \sum_a (a \hat{N}_i / a \hat{N}) a B_{ht},$$

which gives (2.3).

The estimate  $\hat{p}_i (= \hat{N}_i / \hat{N})$  is the ratio of two post-stratified estimates. Hence  $\text{estcov}(\hat{p}_i, \hat{p}_l)$  is given by (2.4) with  $\hat{o}_{ih}(f)$  for a province, obtained from (A.1) by changing  $y_{i(htk)}$  to

$$\begin{aligned} \bar{y}_{i(htk)} &= [y_{i(htk)} - \sum_a (a \hat{N}_i / a \hat{N}) a x_{htk}] \\ - \hat{p}_i^* [y_{+(htk)} - \sum_a (a \hat{N}_i / a \hat{N}) a x_{htk}], \end{aligned} \quad (\text{A.2})$$

where  $y_{+(htk)} = \sum_i y_{i(htk)}$ , i.e.,  $B_{i(ht)}$  is replaced by  $\sum_k w_{htk} \bar{y}_{i(htk)} = [B_{i(ht)} - \hat{p}_i B_{(ht)}] - \sum_a (a B_{ht} / a \hat{N}) (a \hat{N}_i - \hat{p}_i a \hat{N})$  which leads to (2.4).

Turning to two-way tables, the covariance matrix of  $\hat{h}_{ij} = \hat{p}_{ij} - \hat{p}_{i+} \hat{p}_{+j}$  is approximately the same as that of  $\tilde{h}_{ij} = \hat{p}_{ij} - p_{i+} \hat{p}_{+j} - \hat{p}_{i+} p_{+j}$ . Hence, using obvious extension of the notation in (A.2) and noting that  $p_{ij} = p_{i+} p_{+j}$  under  $H_0$ , the estimated covariance of  $\hat{h}_{ij}$  and  $\hat{h}_{i'j'}$  is obtained from (2.3) and (2.4) by changing  $y_{i(htk)}$  to

$$\begin{aligned} \bar{y}_{ij(htk)} &= [y_{ij(htk)} \\ - \hat{p}_{i+} y_{+j(htk)} \\ - \hat{p}_{+j} y_{i+(htk)} + \hat{p}_{i+} \hat{p}_{+j} y_{++(htk)}] \\ - \sum_a (a x_{htk} / a \hat{N}) [a \hat{N}_{ij} - \hat{p}_{i+} a \hat{N}_{+j} - \hat{p}_{+j} a \hat{N}_{i+} \\ + \hat{p}_{i+} \hat{p}_{+j} a \hat{N}], \end{aligned} \quad (\text{A.3})$$

where  $y_{+j(htk)} = \sum_i y_{ij(htk)}$ , etc., i.e.,  $B_{i(htk)}$  is replaced by  $\sum_k w_{htk} \bar{y}_{ij(htk)} = z_{ijht}$  (equation (5.4)).