CHISQUARED TESTS FOR ANALYSIS OF CATEGORICAL DATA FROM COMPLEX SURVEYS

J.N.K. Rao* and A.J. Scott*
Carleton University and University of Auckland

SUMMARY

The effect of stratification and clustering on the asymptotic distributions of X^2, the standard goodness-of-fit chisquared statistic, and X_I^2, the chisquared statistic for testing independence in a two-way table, is investigated. It is shown that both X^2 and X_I^2 are asymptotically distributed as weighted sums of independent χ_1^2 random variables. The weights are then related to the familiar design effects (deffs) used by survey samplers. For the special case of stratified simple random sampling under proportional allocation, X^2 (or X_I^2) is shown to provide a conservation test without any correction. A simple correction to X^2, which requires only the knowledge of variance estimates (or deffs) for individual cells in the goodness-of-fit problem, is proposed and empirical results on the performance of corrected X^2 provided. Empirical work on X_I^2 indicated that the distortion of nominal significance levels is much smaller with X_I^2 than X^2.

1. INTRODUCTION

Methods for analysis of categorical data have been developed extensively under the assumption of multinomial sampling; in particular we have χ^2 tests for problems involving goodness-of-fit and tests of independence and homogeneity in two-way contingency tables. Recent extensions of these methods to multidimensional contingency tables using log linear models (e.g. Fienberg (1977)) have attracted considerable attention due to their close similarity to analysis of variance in providing systematically tests of various hypotheses.

Subject matter research workers have long been using these classical methods to analyze sample survey data, but most of the commonly used survey designs employ stratification or cluster sampling or both and hence do not satisfy the assumption of multinomial sampling. Operational and cost considerations often dictate the use of a complex clustered design and, therefore, it becomes important to assess the impact of survey design on the classical χ^2 tests.

In this paper we propose to investigate the effect of stratification and clustering on the asymptotic distributions of X^2, the goodness-of-fit χ^2 statistic, and X_I^2, the χ^2 statistic for testing independence in a two-way contingency table. In Section 2 we show that X^2 is distributed asymptotically as a weighted sum of independent χ_1^2 random variables, and then relate the weights to the familiar "design effects (deffs)" used by survey samplers. For the special case of stratified simple random sampling (srs) under proportional allocation, we show that X^2 provides a conservative test without any correction. Similar results for X_I^2 are obtained in Section 4.

Our investigation on the asymptotic distribution of X^2 suggests that a simple correction factor to X^2, which requires only the knowledge of variance estimates (or deffs) for individual cells in the goodness-of-fit problem, might be satisfactory for many designs. Empirical results on the performance of corrected statistic $\tilde{X}^2$ are given in Section 3. Results in Section 4, however, indicate that no such simple correction for X_I^2 is possible. Some empirical results in Section 5 indicate that X_I^2 is likely to be less affected by the survey design than X^2.

Cohen (1976), Altham (1976), and Brier (1978), among others have proposed simple models for clustering and derived appropriate χ^2 tests for goodness-of-fit and independence in the case of single-stage cluster sampling (with equal cluster sizes) and two-stage cluster sampling (with equal subsample sizes). They have shown that a simple correction to X^2 (or X_I^2) leads to asymptotically correct test statistic, under their models. Extensions of these model-based results to unequal subsample sizes, three-stage sampling, stratified two-stage sampling and testing finite population cell proportions, along with a study on the effect of model deviations on the corrected X^2 of Altham and Brier are given in the Appendix.

2. EFFECT OF SURVEY DESIGN ON X^2

2.1 X^2 Statistic

Consider the goodness-of-fit problem with k cells and associated probabilities $p_1,\ldots,p_k$ ($p_i \geq 0$, $\Sigma p_i = 1$). Let $n_1,\ldots,n_k$ denote the observed cell frequencies in a sample, s, of n ultimate units drawn according to a specified sampling design, $p(s)$. The conventional χ^2 statistic for testing $H_0 : p_i = p_{0i}$ $(i=1,\ldots,k)$ is then given by

$$X^2 = \sum_1^k (n_i - np_{0i})^2/(np_{0i}) \quad . \qquad (2.1)$$

The statistic (2.1) is appropriate when $E(n_i) = np_i$ which, for instance, is satisfied for self-weighting designs or under models for cluster sampling investigated in Section 4. For other cases we encounter difficulties with noncentral distributions, so it is natural to consider a more general statistic

$$X^2 = n \sum_1^k (\hat{p}_i - p_{0i})^2/p_{0i} \ , \qquad (2.2)$$

where $\hat{p}_i$ is an unbiased (or asymptotically unbiased) estimate of p_i. If $n\hat{p}_i = n_i$, (2.2) reduces to the usual statistic (2.1).

2.2 Asymptotic Distribution

We now look at the asymptotic distribution of X^2. To this end, we suppose that $n^{\frac{1}{2}}(\hat{\underset{\sim}{p}} - \underset{\sim}{p})$ is asymptotically (k-1)-variate normal with mean $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{V}$ say, where $\underset{\sim}{p} = (p_1,\ldots,p_{k-1})'$. If $\underset{\sim}{\ell}'\hat{\underset{\sim}{p}}$ for any

$\underset{\sim}{\ell} = (\ell_1, \ldots, \ell_{k-1})'$ can be written (at least asymptotically as $\sum_{t \in s} w_t(s) y_t$, an estimator of the population mean for some y_t and weights $w_t(s)$, then a suitable central limit theorem for means ensures that $n^{\frac{1}{2}}(\hat{\underset{\sim}{p}} - \underset{\sim}{p})$ is asymptotically normal.

If a consistent estimator of V, say $\hat{V}$, is available, the analysis is fairly straightforward. The natural test is based on the corresponding Wald statistic

$$X_W^2 = n(\hat{\underset{\sim}{p}} - \underset{\sim}{p}_0)' \hat{\underset{\sim}{V}}^{-1} (\hat{\underset{\sim}{p}} - \underset{\sim}{p}_0) \tag{2.3}$$

which will be distributed asymptotically as χ^2_{k-1} under H_0. A good illustration of this approach is given in Koch et al (1975), Nathan (1975) and Shuster and Downing (1976). Note that X^2 can be expressed in the form

$$X^2 = n(\hat{\underset{\sim}{p}} - \underset{\sim}{p}_0)' \underset{\sim}{P}_0^{-1} (\hat{\underset{\sim}{p}} - \underset{\sim}{p}_0) ,$$

where $P_0 = \text{diag.}(\underset{\sim}{p}_0) - \underset{\sim}{p}_0 \underset{\sim}{p}_0'$ is the covariance matrix for multinomial sampling under H_0, so that X^2 is essentially a special case of the Wald statistic. The exact analogue is the modified χ^2 statistic $n \sum_1^k (\hat{p}_i - p_{0i})^2 / \hat{p}_i$ which is asymptotically equivalent to X^2 under H_0 and under local alternatives.

Routine calculation of $\hat{\underset{\sim}{V}}$ is a very desirable practice, but unfortunately this is still the exception rather than the rule. When no estimator $\hat{\underset{\sim}{V}}$ is available, practitioners very often use the standard X^2 statistic (2.1) or (2.2) even when they realize that multinomial assumptions are inappropriate. The correct asumptotic distribution of X^2 follows directly from the asymptotic normality of $\hat{\underset{\sim}{p}}$ and standard results on quadratic forms (Johnson and Kotz (1970, p. 150)).

<u>Theorem 1</u>. Under the hypothesis $H_0 : \underset{\sim}{p} = \underset{\sim}{p}_0$, $X^2 = \sum_1^{k-1} \lambda_i Z_i^2$ where $Z_1, \ldots, Z_{k-1}$ are asymptotically independent $N(0,1)$ random variables and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{k-1}$ are the eigenvalues of $\underset{\sim}{D} = \underset{\sim}{P}_0^{-1} \underset{\sim}{V}$.

In general, then, X^2 is distributed asymptotically as a weighted sum of independent χ_1^2 random variables. The following results are immediate consequences of Theorem 1.

<u>Corollary 1</u>. $X^2/\lambda_1 \leq \sum_1^{k-1} Z_i^2$ where $\sum_1^{k-1} Z_i^2 = n(\hat{\underset{\sim}{p}} - \underset{\sim}{p}_0)' \underset{\sim}{V}^{-1} (\hat{\underset{\sim}{p}} - \underset{\sim}{p}_0)$ is distributed asymptotically ($\sim$) as χ^2_{k-1} under H_0.

If the largest eigenvalue can be specified (or a reasonable upper bound can be set) we can obtain, through this result, a conservative test by treating X^2/λ_1 as χ^2_{k-1}.

<u>Corollary 2</u>. $X^2/\lambda \sim \chi^2_{k-1}$ for some constant λ if and only if $\underset{\sim}{V} = \lambda \underset{\sim}{P}_0$ (i.e. $\text{Var}(\hat{p}_i) = \lambda p_{0i}(1 - p_{0i})/n$ and $\text{Cov}(\hat{p}_i, \hat{p}_j) = - \lambda p_{0i} p_{0j}/n$).

Cohen (1976), Altham (1976) and Brier (1979) have all proposed models with the property $\underset{\sim}{V} = \lambda \underset{\sim}{P}_0$ or $\lambda_1 = \ldots = \lambda_{k-1} = \lambda$.

These general results can also be framed in more familiar sampling terms. In the case of $k = 2$ categories, $\underset{\sim}{D}$ reduces to the ordinary design effect (deff), $n \text{Var}(\hat{p}_1)/[p_{01}(1-p_{01})]$. For general k, $\underset{\sim}{D}$ can be thought of as the natural multivariate extension of the design effect. In particular,

$$\lambda_1 = \sup_{\underset{\sim}{c}} [\underset{\sim}{c}' \underset{\sim}{V} \underset{\sim}{c} / \underset{\sim}{c}' \underset{\sim}{P} \underset{\sim}{c}]$$

$$= \sup_{\underset{\sim}{c}} [\text{Var}(\sum_1^{k-1} c_i \hat{p}_i) / V_{srs}(\sum_1^{k-1} c_i \hat{p}_i)],$$

where V_{srs} denotes the variance under srs. Thus λ_1 represents the largest possible design effect taken over all individual cell proportions and over all possible linear combinations of the cell proportions. In the same way λ_{k-1} can be shown to be the smallest possible design effect taken over all possible linear combinations of the cell proportions. The other λ_i's also represent design effects for special linear combinations of the $\hat{p}_i$'s. Thus the λ_i's may be termed "generalised design effects".

Corollary 1 enables us to construct a conservative test, if we can find an upper bound for the design effects of linear combinations of cell proportions. Corollary 2 says that to have $X^2 \sim \lambda \chi^2_{k-1}$ requires not only that all the individual cells have the same design effect λ but that the deff for each of the covariance terms must also be equal to λ. This condition is rather stringent in practice.

2.3 <u>Special Cases</u>

We now consider two special cases of the general result of Section 2.2.

(i) <u>Stratified srs (Proportional Allocation)</u>

Consider L strata and a stratified sample $s = (s_1, \ldots, s_L)$ where s_h is a random sample of size m_h drawn with replacement from stratum h. Let W_h and P_{ih} respectively denote the proportion of units in stratum h and the proportion of units from stratum h belonging to category i. Then $p_i = \sum_h W_h P_{ih}$ and $m_h = nW_h$ for proportional allocation. With this design we have

$$\underset{\sim}{V} = \underset{\sim}{P} - \sum_1^L W_h (\underset{\sim}{p}_h - \underset{\sim}{p})(\underset{\sim}{p}_h - \underset{\sim}{p})'$$

where $\underset{\sim}{p}_h = (P_{1h}, \ldots, P_{k-1,h})'$. Therefore

$$0 \leq [\underset{\sim}{c}' \underset{\sim}{V} \underset{\sim}{c} / \underset{\sim}{c}' \underset{\sim}{P} \underset{\sim}{c}] = 1 - \sum_h W_h [\underset{\sim}{c}'(\underset{\sim}{p}_h - \underset{\sim}{p})]^2 / (\underset{\sim}{c}' \underset{\sim}{P} \underset{\sim}{c}) \leq 1$$

or $$\lambda_1 \leq 1$$

and
$$0 \leq X^2 \leq \Sigma\, z_i^2 \sim \chi^2_{k-1} .$$

Thus the ordinary χ^2 test is always conservative.

(ii) Two Stage Sampling

Suppose we have R primary sampling units (psu's) with M_t secondary units in the t^{th} p.s.u. $(t = 1,\ldots,R;\ \Sigma M_t = M_0)$. Consider the following two-stage design which is commonly used: r psu's selected with probability proportional to size (pps) of M_t with replacement and subsamples each of size m drawn srs with replacement independently from each selected psu $(n = mr)$. In this case $E(n_i) = np_i$ where $p_i = p_i = \Sigma_t M_t P_{it}/M_0$ and P_{it} is the proportion in the t^{th} psu belonging to category i. We also have

$$\underset{\sim}{V} = \underset{\sim}{P} + (m-1)\Sigma_t W_t (\underset{\sim}{p}_t - \underset{\sim}{p})(\underset{\sim}{p}_t - \underset{\sim}{p})' = \underset{\sim}{P} + (m-1)\underset{\sim}{A}_1 \text{ (say)},$$

where $W_t = M_t/M_0$ and $\underset{\sim}{P}_t = (P_{1t},\ldots,P_{k-1,t})'$.
Let $\rho_1 \geq \ldots \geq \rho_{k-1}$ denote the eigenvalues of $\underset{\sim}{P}^{-1}\underset{\sim}{A}_1$, then $\lambda_i = 1 + (m-1)\rho_i$ and

$$X^2 = \sum_{1}^{k-1} [1 + (m-1)\rho_i] z_i^2 . \qquad (2.4)$$

Also $(\underset{\sim}{c}'\underset{\sim}{A}_1\underset{\sim}{c})/(\underset{\sim}{c}'\underset{\sim}{P}\underset{\sim}{c}) \leq 1$ so that $\rho_1 \leq 1$ and we get

$$\sum_{1}^{k-1} z_i^2 \leq [1 + (m-1)\rho_{k-1}] \sum_{1}^{k-1} z_i^2 \leq X^2$$
$$\leq [1 + (m-1)\rho_1] \sum_{1}^{k-1} z_i^2 \leq m \sum_{1}^{k-1} z_i^2$$

since $\rho_i \geq 0$. Thus X^2/m gives a conservative test whatever the ρ_i's , but we can get a better conservative test if a value for ρ_1 can be specified. We call the ρ_i's "generalised measures of homogeneity" analogous to the measure of homogeneity based on the intraclass correlation ρ (Kish et al, 1976 and Kalton, 1977). The measure ρ is "portable" in the sense that it is relatively insensitive to cluster sizes unlike the deff.

3. MODIFICATION TO X^2

If the λ_i's were known we could get accurate approximation for the percentage points of the true asymptotic distribution of X^2 , using the methods in Solomon and Stephens (1977), for example. However, knowledge of the λ_i's is essentially equivalent to knowing the full covariance matrix $\underset{\sim}{V} = (v_{ij})$ and if we have this we can always construct an asymptotically correct Wald statistic. In practice we would like a simple approximation to the distribution of X^2 that requires only very limited information about V . One very simple approach is to treat the modified statistic

$$\tilde{X}^2 = X^2/\bar{\lambda} = \sum_{1}^{k-1} (\lambda_i/\bar{\lambda}) z_i^2$$

as a χ^2_{k-1} random variable, where $\bar{\lambda} = \Sigma\lambda_i/(k-1)$. Obviously $\tilde{X}^2$ has the same expected value as χ^2_{k-1} but its asymptotic variance is

$$V(\tilde{X}^2) = 2(k-1) + 2 \sum_{1}^{k-1} (\lambda_i - \bar{\lambda})^2/\bar{\lambda}^2$$

which is larger than $V(\chi^2_{k-1}) = 2(k-1)$ unless all λ_i's are equal.

The important point to note about the modification is that $\bar{\lambda}$ only depends on the cell variances v_{ii} (or equivalently the cell design effects $d_1,\ldots,d_k$) since

$$\bar{\lambda} = \mathrm{tr}(\underset{\sim}{P}^{-1}\underset{\sim}{V})/(k-1) = \sum_{1}^{k} v_{ii}/[p_i(k-1)]$$
$$= (k-1)^{-1} \sum_{1}^{k} (1-p_i)d_i .$$

Some knowledge about design effects for individual cells in a goodness-of-fit problem is often available (Kish et al (1976) for example), whereas information about effects for the covariance terms is much less common. Note that $\bar{\lambda}$ is not in general the same as $\bar{d}$, the average cell design effect, except when $p_i = 1/k$ or $d_i = d$ $(i = 1,\ldots, k)$.

Empirical Results. Table 1, taken from Ewings (1979), shows the significance levels of the ordinary X^2 test and of the modified test, $\tilde{X}^2$, for a nominal level of 0.05 for various items taken from the 1971 General Household Survey (GHS) of the U.K. This is a stratified three-stage sample of approximately 13,000 households. Here $\bar{b}$ denotes the average cluster size and d.f. is an abbreviation for degrees of freedom. The modified test gives very good results in all cases of Table 1. In contrast, if we naively use the ordinary chi-squared test without modification the true significance is as high as 0.41 which is obviously unacceptable.

Table 1: Significance Levels (SL) for Ordinary X^2 and Modified Test, $\tilde{X}^2$

Variable	d.f.	$\bar{b}$	$\bar{\lambda}$	$Var(\lambda_i)$	$SL(X^2)$	$SL(\tilde{X}^2)$
Age of Bldg	2	33.1	3.42	1.65	0.41	0.05
Home (own/rent)	3	34.4	2.54	1.82	0.37	0.06
No. of bedrooms	6	34.7	1.29	0.20	0.15	0.06
Bedroom Standard	4	34.5	1.01	0.18	0.06	0.055
No.of rooms	9	34.6	1.19	0.25	0.14	0.06
No.of cars	3	34.6	1.48	0.85	0.16	0.06
Household gross weekly income	3	26.6	1.40	0.41	0.14	0.055

4. EFFECT OF SURVEY DESIGN ON X_I^2

Consider the problem of testing for independence in a two-way table. Suppose the $k = rc$ categories are set out in a table with r rows and c columns. Let $\hat{p} = (\hat{p}_{11}, \hat{p}_{12}, \ldots, \hat{p}_{rc})'$ be the vector of estimated cell proportions and assume as before that $\sqrt{n}(\hat{p} - p)$ is asymptotically normal with mean vector and covariance matrix V. (Note that V will be singular here since it is convenient to modify the definition of p to include all rc cells in this section.) Let $p_r = (p_{1+}, \ldots, p_{r+})'$ denote the vector of marginal row proportions, and $p_c = (p_{+1}, \ldots, p_{+c})'$ the vector of column proportions, with $p_{i+} = \sum_{j=1}^{c} p_{ij}$ and $p_{+j} = \sum_{i=1}^{r} p_{ij}$. We are interested in the null hypothesis of independence between rows and columns, i.e.

$$H_0 : p_{ij} = p_{i+}p_{+j} \quad (i=1,\ldots,r;\ j=1,\ldots,c)$$

or

$$H_0 : p = p_r \otimes p_c ,$$

i.e., p is the direct product of p_r and p_c. The usual chi-squared statistic for testing independence is

$$X_I^2 = n \sum_1^r \sum_1^c \frac{(\hat{p}_{ij} - \hat{p}_{i+}\hat{p}_{+j})^2}{\hat{p}_{i+}\hat{p}_{+j}}$$

which, after some manipulation, can be written in the form

$$X_I^2 = n\, h'(p)\ \hat{P}_r^{-1} \otimes \hat{P}_c^{-1}\ h(\hat{p})$$

where $h(\hat{p})$ denotes the $(r-1)(c-1)$-dimensional vector with components $h_{ij} = \hat{p}_{ij} - \hat{p}_{i+}\hat{p}_{+j}$ and $P_r = \text{diag.}(\tilde{p}_r) - \tilde{p}_r\tilde{p}_r'$ where $\tilde{p}_r = (p_{1+}, \ldots, p_{(r-1)+})'$ with similar definition for P_c and $\tilde{p}_c$.

It can be shown that h_{ij} is asymptotically equivalent to $h^*_{ij} = \hat{p}_{ij} - \hat{p}_{i+}p_{+j} - p_{i+}\hat{p}_{+j} + p_{i+}p_{+j}$ under the conditions above, so that h has the same asymptotic distribution as $h^* = H(\hat{p} - p_r \otimes p_c)$ say, where H is the $(r-1)(c-1) \times rc$ matrix

$$H = J_r \otimes J_c - (\tilde{p}_r 1_r') \otimes J_c - J_r \otimes (\tilde{p}_c 1_c')$$

where $J_r = (I_{r-1})|0)$ and 1_r is an $r \times 1$ vector of ones. This means that h is asymptotically normal with mean $H(p - p_r \otimes p_c)$ and covariance matrix HVH'/n. As before, if we have a consistent estimator, $\hat{V}$, of V available a natural test for H_0 would be based on the Wald statistic

$$X_W^2 = nh'(p)(\hat{H}\,\hat{V}\,\hat{H}')^{-1}\,h(\hat{p})$$

where $\hat{H} = H(\hat{p})$, which is asymptotically $\chi^2_{(r-1)(c-1)}$ under H_0, or on a modified version of the Wald statistic in which V (or H) is estimated assuming H_0 is true. With multinomial sampling $H\,V\,H'$ reduces to $P_r \otimes P_c$ when $p = p_r \otimes p_c$ so that the ordinary X_I^2 is equivalent to the corresponding modified Wald statistic.

As in the case of X^2, the true asymptotic distribution of X_I^2 follows directly from standard results on quadratic forms:

<u>Theorem 2</u>. $X_I^2 = \sum_I^{(r-1)(c-1)} \delta_i Z_i^2$ where $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_{(r-1)(c-1)}$ are the eigenvalues of $(P_r^{-1} \otimes P_c^{-1})(H\,V\,H')$ and $Z_1, Z_2, \ldots, Z_{(r-1)(c-1)}$ are asymptotically independent $N(0,1)$ random variables under the null hypothesis of independence.

The Z_i's are now linear combinations of the h^*_{ij}'s, which are themselves linear combinations of the $\hat{p}_{ij}$'s, and δ_i is the deff for Z_i. Since Z_i is a particular linear combination of the $\hat{p}_{ij}$'s, we have

$$\lambda_1 \geq \delta_1 \geq \lambda_{rc-1}$$

where λ_1 is the largest possible deff among all linear combinations of the cell proportions and λ_{rc-1} is the smallest. (More precisely, it follows from results in Anderson and Das Gupta (1953) that $\lambda_i \geq \delta_i \geq \lambda_{i+r+c-2}$.) Clearly $X_I^2/\delta_1 \leq \Sigma Z_i^2$ so that we can get a conservative test whenever we can specify δ_1 (or a reasonable upper bound for δ_1).

The most important application of this result is to stratified srs with proportional allocation. Since we must have $0 \leq \delta_i \leq 1$ for all i, $0 \leq X_I^2 \leq \Sigma Z_i^2 \sim \chi^2_{(r-1)(c-1)}$ so that the ordinary chi-squared test is conservative, just as in the goodness-of-fit case. The gains from stratification are often rather small and the usual test should work very well in these cases.

We could follow the approach of Section 2.2 to obtain an approximate chi-squared statistic by dividing X_I^2 by $\bar{\delta}$. Unfortunately this requires information on the design effects of the h_{ij}'s (or h^*_{ij}'s) and such information is rarely available at present except in the 2×2 case (Kish and Hess (1959)). We could perhaps divide by $\bar{\lambda}$, in the hope that $\bar{\lambda}$ is close to $\bar{\delta}$. As before $\bar{\lambda}$ can be calculated from the value of the design effects for the cells (i.e. from the diagonal elements of V) but information on design effects for individual cells is not likely to be available either. (Fellegi (1978) suggests dividing by the average cell deff $\bar{d}$ which will usually be close to $\bar{\lambda}$.) Perhaps the best that we can reasonably hope for in most cases is to have some information about design effects for the marginal row and column totals, (the sort of information given in Kish, Groves and Krotki (1976), for example), and it would be desirable to have an approximation based on the marginal design effects. A simple special case where this is possible is when $\lambda_1 = \lambda_2 = \ldots = \lambda_{rc-1} = \lambda$ and hence $\delta_i = \lambda$ for $i = 1,\ldots,(r-1)(c-1)$. All marginal design effects are also equal to λ in this case and we can find the appropriate multiplier from either margin. Cohen (1976), Altham (1976) and Brier (1979) have all proposed models with this property. However, the assumption of a common design effect throughout the table seems less realistic here than in the goodness-of-fit case, especially if the row and column variables are of quite different types. For example, average design effects for socio-economic classifications reported in Kish, Groves and Krotki (1976) range from 4 to 8 while average design effects for demographic variables range from 1.0 to 1.6. Furthermore, the limited amount of empirical evidence available suggests that $\bar{\delta}$ will tend to be less than the average design effect of either margin in general (Kish and Hess (1959), Kish and Frankel (1974), Nathan (1975)).

5. EMPIRICAL RESULTS FOR X_I^2

An extensive empirical investigation based on the results of Section 4 has been carried out by Holt, Scott and Ewings (1979) using data from two large-scale U.K. surveys. Some results of that study are summarized here for illustration. In all the examples the variance of the δ_i's is small and $X^2/\bar{\delta}$ can be regarded as a $\chi^2_{(r-1)(c-1)}$ random variable under H_0.

Example 1. The first set of variables comes from the British Election Study (BES), which is a stratified three-stage sample, like GHS, (voters within wards without constituencies) of about 2500 British voters. The sample is approximately self-weighting. We are grateful to the SSRC Survey Archive at the University of Essex for making the data available. We give results for cross-classifications of the four variables set out in Table 2, along with the marginal design effects. Deffs are reasonably small for all variables in this study.

Table 2: Description of Variables for BES

Variable	No. of Categories	Deff
B1: Sex	2	1.06
B2: "Election campaign gave people facts"	2	1.29
B3: "USA looks at World Politics as we do"	3	1.32
B4: Social Grade	3	1.59
B5: Length of residence	4	1.55

In Table 3 we give values of $\bar{\lambda}$ and $\bar{\delta}$ for the cross-classifications, along with the value of the Wald statistic to give some feeling for the relative degree of dependence.

Table 3. Deffs for Cross-classifications (BES)

Variables	r × c	$\bar{\lambda}$	$\bar{\delta}$	X_W^2
B1 x B2	2 x 2	1.14	1.02	2.39
B1 x B3	2 x 3	1.15	1.02	18.68
B1 x B4	2 x 3	1.30	1.12	146.75
B1 x B5	2 x 4	1.25	1.02	2.52
B2 x B3	2 x 3	1.13	0.97	4.94
B2 x B4	2 x 3	1.21	0.84	2.24
B2 x B5	2 x 4	1.26	1.00	0.58
B3 x B4	3 x 3	1.24	1.07	19.74
B3 x B5	3 x 4	1.19	1.03	17.89
B4 x B5	3 x 4	1.19	0.97	20.59

The outstanding feature of the results (and of all the results for other variables in the study) is the small value of $\bar{\delta}$; the ordinary chi-squared needs no modification in any of these examples. Modified statistics based on $\bar{\lambda}$ or the average of the marginal deffs will be conservative in every case (and in every case in the larger study), with the one based on $\bar{\lambda}$ performing slightly better. The unmodified chi-squared statistic, however, is clearly the best choice in all cases.

Example 2. The second set of variables comes from the General Household Survey (GHS) of 1971. As can be seen from Table 4, the deffs are rather higher for the variables considered here.

Table 4. Description of Variables for GHS

Variable	No. of Categories	Deff
G1: Rented, owned	2	3.28
G2: Age of dwelling	2	4.28
G3: Age of head	3	1.26
G4: Type of Accommodation	3	2.83

Values of $\bar{\lambda}$ and $\bar{\delta}$ for the 6 cross-classifications are given in Table 5.

Table 5. Deffs for Cross-classifications (GHS)

Variables	r × c	$\bar{\lambda}$	$\bar{\delta}$	X_W^2
G1 x G2	2 x 2	3.18	1.99	22.87
G1 x G3	2 x 3	1.61	1.13	20.17
G1 x G4	2 x 3	2.50	1.94	592.36
G2 x G3	2 x 3	1.75	0.97	156.39
G2 x G4	2 x 3	2.36	1.97	651.33
G3 x G4	3 x 3	1.51	0.96	50.83

Again $\bar{\delta}$ is always much smaller than $\bar{\lambda}$, which is slightly smaller in turn than the average of the two marginal deffs. Thus modifications based on $\bar{\lambda}$ or the average marginal deffs will be too conservative for the GHS data too, although $\bar{\delta}$ is too large for some of the cross-classifications for the unmodified chi-squared test to be satisfactory.

These empirical results give extra support to the hypothesis that the ordinary chi-squared statistic for independence is very much less affected in a complex survey than the corresponding statistics for goodness-of-fit, although both these surveys are of very similar structure (approximately self-weighting). The results also indicate that modifications based on the average deff or the average marginal deff may be too severe. However, the values of $\bar{\delta}$ in the GHS results mean that blind application of the naive test still has dangers. We need models which explain the difference between $\bar{\lambda}$ and $\bar{\delta}$ and which leads to a smaller divisor than $\bar{\lambda}$. A simple model for two-stage sampling will be explored elsewhere.

REFERENCES

Altham, P.A.E. (1976), "Discrete Variable Analysis for Individuals Grouped Into Families", Biometrika, 63, 263-9.

Anderson, T.W. and Das Gupta, S. (1963). "Some Inequalities on Characteristic Roots of Matrices", Biometrika, 50, 522-4.

Brier, S.S. (1978), Categorical Data Models for Complex Data Structures, Unpublished Ph.D. Dissertation, School of Statistics, University of Minnesota.

Cohen, J.E. (1976), "The Distribution of the Chi-squared Statistic under Cluster Sampling from Contingency Tables", J. Amer. Statist. Assoc., 71, 665-70.

Ewings, P.D. (1979), The Effect of Complex Sample Designs on the Chi-squared Goodness of Fit Statistic, Unpublished M.Sc. Dissertation, University of Southampton.

Fellegi, I.P. (1978) "Approximate Tests of Independence and Goodness of Fit based on Stratified and Multi-stage Samples", Survey Methodology, 4, No. 2, 29-56.

Fienberg, S.E. (1977). The Analysis of Cross-classified Categorical Data, New York: MIT Press.

Holt, D., Scott, A.J. and Ewings, P.D. (1979), "The Behaviour of Chi-squared Tests with Complex Survey Data", Unpublished Manuscript.

Johnson, N.L., and Kotz, S. (1970), Continuous Univariate Distributions - 2, Boston: Houghton Mifflin Co.

Kalton, G. (1977), "Practical Methods for Estimating Sampling Errors", Paper Presented at the Meetings of the Int. Statist. Inst., New Delhi.

Kish, L., Groves, R.M., and Krotki, K.P. (1976), Sampling Errors for Fertility Survey, Occasional Paper No. 17. London: World Fertility Survey.

Kish, L. and Frankel, M.R. (1974), "Inference from Complex Samples", J. R. Statist. Soc. B, 36, 1-37.

Kish, L. and Hess, I. (1959), "On Variances of Ratios and their Differences in Multi-stage samples", J. Amer. Statist. Assoc., 54, 416-46.

Koch, G.C., Freeman, D.H. Jr. and Freeman, J.L. (1975), "Strategies in the Multivariate Analysis of Data from Complex Surveys," Int. Stat. Rev., 43, 59-78.

Nathan, G. (1975), "Tests for Independence in Contingency Tables from Stratified Samples," Sankhya C, 37, 77-87.

Shuster, J.J. and Downing, D.J. (1976), "Two-way Contingency Tables for Complex Sampling Schemes," Biometrika 63, 271-6.

Solomon, H., and Stephens, M.A. (1977), "Distribution of a Sum of Weighted Chi-square Variables, J. Amer. Statist. Assoc., 72, 881-5.

FOOTNOTE

* The names of the authors are listed in alphabetical order. Most of this work was carried out at the University of Southampton under Grant No. HR4973/1 from the U.K. Social Science Research Council. The authors are grateful for the stimulus and help given by D. Holt, G. Nathan, T.M.F. Smith and G. Kalton.

APPENDIX

X^2 UNDER MODELS FOR CLUSTER SAMPLING

A.1. Two-stage Sampling (Altham's Model)

The population consists of R psu's with M_t secondaries in the t^{th} psu. A two stage sample s is denoted by $(s_1,\ldots,s_r)$ where s_t is a subsample of size m_t ($\sum_1^r m_t = n$ and $m_t \le M_t$) and r is the number of sampled psu's. Let

$$z_{t\lambda i} = \begin{cases} 1 \text{ if } \lambda^{th} \text{ population element of } i^{th} \text{ psu is in category } i \\ 0 \text{ otherwise; } \lambda=1,\ldots,M_t;\ t=1,\ldots,R;\ i=1,\ldots,k. \end{cases}$$

Then

$$n_i = \sum_{t=1}^{r} \sum_{\lambda=1}^{m_t} z_{t\lambda i} = \sum_{t=1}^{r} m_{ti}\ , \text{ say.}$$

In the model approach, we regard the $Z_{t\lambda i}$'s as random variables with assumed distributional properties and expectations etc. are taken with respect to the assumed model. Suppose (Altham, 1976)

$$E(Z_{t\lambda i}) = p_i \quad \text{and} \quad Cov(Z_{t\lambda i}, Z_{t\lambda' j}) = b_{ij} \qquad (A.1)$$

for $\lambda \neq \lambda'$. Note that these values do not depend on the cluster label, so we are assuming a form of weak exchangeability between clusters. This model will not be suitable if the clusters are obviously different (as in the case of different strata for example). If we also assume that values in different clusters are uncorrelated then it is straightforward to show that, conditional on the realized sample sizes $(m_1,\ldots,m_r)$, the covariance matrix of $\underset{\sim}{n} = (n_1,\ldots,n_{k-1})'$ can be expressed in the form $n\underset{\sim}{V} = n\underset{\sim}{P} + (\Sigma m_t^2 - n)\underset{\sim}{B}$, where $\underset{\sim}{P} = \text{diag}(\underset{\sim}{p}) - \underset{\sim}{p}\,\underset{\sim}{p}'$ as before and $\underset{\sim}{B} = (b_{ij})$, or as

$$\underset{\sim}{V} = \underset{\sim}{P} + (m_0 - 1)\underset{\sim}{B} \ , \qquad (A.2)$$

where $m_0 = \sum_1^r m_t^2/n$.

Notice the close similarity to the expression for pps sampling given in (2.4). The difference is that (2.4) applies only to a specific design but requires no assumptions about the population elements, while (A.2) is valid for any two-stage design but requires the weak exchangeability assumptions of the model. Of course, the choice of design and model are usually intimately related and we would normally feel more secure about results based on an exchangeable model if the sample was chosen with the self-weighting design of Section 2.3.

If we make the stronger assumption of independence between different clusters, then $\underset{\sim}{n}$ is approximately normal for large r and, from Theorem 1, $X^2 \sim \sum_1^{k-1} (1 + (m_0-1)\rho_i)Z_i^2$ where $\rho_1,\ldots,\rho_{k-1}$ are the eigenvalues of $\underset{\sim}{P}^{-1}\underset{\sim}{B}$.

<u>Lemma</u>. $\underset{\sim}{P} - \underset{\sim}{B}$ is non-negative definite.

<u>Proof</u>. Let $\underset{\sim}{B}^* = \underset{\sim}{P} - \underset{\sim}{B}$ and let $p_{ij} = E(Z_{t\lambda i}Z_{t\lambda' j})$. Then, noting that $p_i = p_{ii} + \sum_{j\neq i}^{k} p_{ij}$, we can write

$$\underset{\sim}{c}'\underset{\sim}{B}^*\underset{\sim}{c} = \sum_1^{k-1} p_{ik}c_i^2 + \sum_{i<j}^{k-1}\sum p_{ij}(c_i - c_j)^2 \geq 0$$

since $p_{ij} \geq 0$.

It follows from the lemma that $\rho_i \leq 1$ for $i=1,\ldots,k-1$ so that $X^2/m_0 \leq \sum_1^{k-1} Z_i^2$. Hence we get a conservative test if we treat X^2/m_0 as a χ^2_{k-1} random variable.

A.2. <u>Special Case: Constant Deffs</u>

In the special case when $\underset{\sim}{B} = \rho\underset{\sim}{P}$ we have $\rho_i = \rho$ $(i=1,\ldots,k-1)$ so that $X^2 \sim (1+(m_0-1)\rho) \sum_1^{k-1} Z_i^2$ and the asymptotic distribution of the modified statistic $\tilde{X}^2 = X^2/(1+(m_0-1)\rho)$ is exactly χ^2_{k-1} under H_0. The parameter ρ is like a generalized measure of homogeneity. This is an example of the situation in Corollary 2 of Section 2.2 and, as we noted there, the model is fairly restrictive since it implies the same deff for all individual cells and for all linear combinations of the cell totals. This model was introduced for single-stage sampling by Cohen (1976) for $M_t = M = 2$ and Altham (1976) for general M , and they suggest several ways in which it might arise naturally.

Brier (1978) considered a model for two-stage sampling (for the case $m_t = m$) with multinomial sampling within each sampled cluster. Thus $\underset{\sim}{m}_t = (m_{t1},\ldots,m_{t,k-1})'$ is multinomial with probabilities $\underset{\sim}{p}_t = (p_{t1},\ldots,p_{t,k-1})'$ and the $\underset{\sim}{p}_t$'s are assumed to be sampled independently from a Dirichlet distribution with density function

$$f(\underset{\sim}{p}_t|\underset{\sim}{p},\nu) = [\Gamma(\nu)/\Gamma(\nu p_i)]\prod_i^k p_{ti}^{\nu p_i - 1} \ , \quad \nu>0,\ p_i>0,\ \sum_1^k p_i = 1.$$

The marginal distribution of $\underset{\sim}{m}_t$ is compound multinomial with $E(\underset{\sim}{m}_t) = m\underset{\sim}{p}$ and $Cov(\underset{\sim}{m}_t) = [m(\nu+m)/(\nu+1)]\underset{\sim}{P}$. Using these moments, it is readily seen that $E(\underset{\sim}{n}) = n\underset{\sim}{p}$ and

$$\underset{\sim}{V} = [(m+\nu)/(1+\nu)]\underset{\sim}{P} = [1 + (m-1)\rho]\underset{\sim}{P}$$

where $\rho = 1/(1+\nu)$. Thus this model is an example of the constant deff model and $\tilde{X}^2 = [(1+\nu)/(m+\nu)]X^2 \sim \chi^2_{k-1}$ under H_0 .

We can extend these results easily to the more general case of unequal subsample sizes and finite M_t , assuming srs without replacement in each sampled cluster. The distribution of $\underset{\sim}{m}_t$ for given $\underset{\sim}{M}_t = (M_{tk},\ldots, M_{t,k-1})'$ is now a $(k-1)$-dimensional hypergeometric distribution with density function

$$f(m_t|M_t) = \prod_1^k \binom{M_{ti}}{m_{ti}} \Big/ \binom{M_t}{m_t}$$

where $M_t = \sum_1^k M_{ti}$. If we regard the $\underset{\sim}{M}_t$'s as independent random variables having compound multinomial distributions with $E(\underset{\sim}{M}_t) = M_t\underset{\sim}{p}$ and $Cov(\underset{\sim}{M}_t) = [M_t(M_t+\nu)/(1+\nu)]\underset{\sim}{P}$, then the marginal distribution of $\underset{\sim}{m}_t$ is again compound multinomial with $E(\underset{\sim}{m}_t) = m_t\underset{\sim}{p}$ and $Cov(\underset{\sim}{m}_t) = [m_t(m_t+\nu)/(1+\nu)]\underset{\sim}{P}$ and the $\underset{\sim}{m}_t$'s are independent. We have $E(\underset{\sim}{n}) = n\underset{\sim}{p}$ as above and $\underset{\sim}{V} = [(\Sigma m_t^2/n + \nu)/(1+\nu)]\underset{\sim}{P} = [1 + (m_0-1)\rho]\underset{\sim}{P}$ where $m_0 = \Sigma m_t^2/n$ and $\rho = 1/(1+\nu)$ as before. Again we have a constant design effect model and the modified statistic $\tilde{X}^2 = [(1+\nu)/(m_0+\nu)]X^2 \sim \chi^2_{k-1}$ under H_0. These results hold for any design for psu selection provided the $\underset{\sim}{m}_t$'s are uncorrelated and the compound multinomial model is valid.

To use the modified statistic $\tilde{X}^2$ we need a value for ρ or alternatively for $C = [(m_0+\nu)/(1+\nu)] = 1 + (m_0-1)\rho$. If the cell frequencies $\underset{\sim}{m}_t$ in each cluster are observed, it is possible to construct estimators of C (or ρ) which are unbiased. Since $E(\underset{\sim}{m}_t) = m_t\underset{\sim}{p}$ and $n\underset{\sim}{V} = Cov(\underset{\sim}{n}) = \sum_1^r Cov(\underset{\sim}{m}_t)$, an unbiased estimator of $\tilde{V}$ is given by

$$n\hat{\tilde{V}} = \frac{r}{r-1}\left[\sum_1^r (\tilde{m}_t-\bar{\tilde{m}})(\tilde{m}_t-\bar{\tilde{m}})' - \sum_1^r (m_t-\bar{m})^2 \; \tilde{p}_0\tilde{p}_0'\right]$$

where $\bar{\tilde{m}} = \sum_1^r \tilde{m}_t/r$ and $\bar{m} = \sum_1^r m_t/r$. Noting that $v_{ii}/p_i(1-p_i) = v_{ij}/p_ip_j = C$, we can construct several unbiased estimators of C. Using only the diagonal elements of $\hat{\tilde{V}}$ we could use

$$\hat{C}_1 = (k-1)^{-1}\sum_1^k \hat{v}_{ii}/p_{0i}\,,$$

$$\hat{C}_2 = k^{-1}\sum_1^k \hat{v}_{ii}/p_{0i}(1-p_{0i}) \qquad \text{(A.3)}$$

or

$$\hat{C}_3 = \sum_1^k \hat{v}_{ii}/(1-\sum_1^k p_{0i}^2)$$

which are all unbiased for C under $H_0: p_i = p_{0i}$. Approximately unbiased estimators of C which are valid for all $\tilde{p}$, are obtained from (A.3) if p_{0i} is replaced by its estimator $\hat{p}_i = n_i/n$.

The modified statistic, $\tilde{X}^2$, derived above does not involve the finite population sizes M_t and R since we are dealing with model parameters p_i. If our interest is in the finite population proportions $P_i = \sum\sum Z_{t\lambda i}/N = N_i/N$, then it can be shown, after some algebra we omit, that $\tilde{X}^2 = X^2/A$ is asymptotically χ^2_{k-1} when $\tilde{B} = \rho\,\tilde{P}$, where

$$A=(1-\frac{n}{N})(1+(m_0-1)\rho)-2\,\frac{\rho}{n}\sum_1^r m_t(M_t-m_t)+\frac{n}{N}(M_0-m_0)\rho, \qquad \text{(A.4)}$$

$m_0 = \sum_1^r m_t^2/n$ as before and $M_0 = \sum_1^R M_t^2/N$. In the special case of $m_t = m$, $M_t = M$ for all t, (A.4) reduces to

$$A = (1-\frac{mr}{MR})(1+(m-1)\rho) - \frac{mr}{MR}(M-m)\rho. \qquad \text{(A.5)}$$

If $\rho = 1/(1+\nu)$ as in Brier's model, then (A.5) reduces to

$$A = C[(1-\frac{r}{R}) + \frac{\nu r}{(m+\nu)R}(1-\frac{m}{M})]\,.$$

A.3 Distribution of $\tilde{X}^2$ Under a Deviation from Constant Deffs Model

Since the assumption of constant deffs is rather restrictive, it is of interest to study the asymptotic distribution of $\tilde{X}^2$ under model deviations. A simple model deviation, which leads to nonconstant deffs, is given by the following "mixture" assumptions:

$$E(Z_{t\lambda i}) = \sum_{h=1}^L W_h p_{hi} = p_i,\; W_h \geq 0,\; \sum W_h = 1$$

$$\text{Cov}(Z_{t\lambda i}, Z_{t\lambda' j}) = \sum_{h=1}^L W_h b_{hij} = b_{ij},\; \lambda \neq \lambda' \qquad \text{(A.6)}$$

$$\tilde{B}_h = (b_{hij}) = \rho\,\tilde{P}_h,\; h=1,\ldots,L$$

where $\tilde{P}_h = \text{diag}(\tilde{p}_h) - \tilde{p}_h\tilde{p}_h'$ and $\tilde{p}_h = (p_{h1},\ldots,p_{h,k-1})'$. The model (A.6) is a special case of the general model (A.1) of Altham. It follows that

$$\tilde{V} = [1+(m_0-1)\rho]\tilde{P} + (m_0-1)(1-\rho)\sum W_h(\tilde{p}_h-\tilde{p})(\tilde{p}_h-\tilde{p})',$$

so that deffs are nonconstant unless $\tilde{p}_h = \tilde{p}$ for all h.

In the special case of $L = 2$, we can evaluate the eigenvalues, λ_i, of $\tilde{P}^{-1}\tilde{V}$ explicitly, where

$$\tilde{P}^{-1}\tilde{V} = [1+(m_0-1)\rho]\tilde{I} + (m_0-1)(1-\rho)W_1W_2\tilde{P}^{-1}(\tilde{p}_1-\tilde{p}_2)(\tilde{p}_1-\tilde{p}_2)'$$

and $\tilde{I}$ is the identity matrix. Since the rank of $\tilde{P}^{-1}(\tilde{p}_1-\tilde{p}_2)(\tilde{p}_1-\tilde{p}_2)'$ is one, k-2 of its eigenvalues are zero $(k > 2)$ and the nonzero eigenvalue is given by its trace: $\sum_1^k (p_{1i}-p_{2i})^2/p_i$. Consequently, the eigenvalues λ_i are given by $\lambda_2=\ldots=\lambda_{k-1} = 1+(m_0-1)\rho$; $\lambda_1 = 1+(m_0-1)\rho + (m_0-1)(1-\rho)\delta$ with $\delta = W_1W_2\sum_1^k (p_{1i}-p_{2i})^2/p_i$. Since the model (A.6) is a special case of (A.1) it follows from Section A.1 that $\lambda_1 \leq m_0$ or $0 \leq \delta \leq 1$.

The asymptotic distributions of $X^2/\hat{C}_1$ and $X^2/\bar{\lambda}$ $(= \tilde{X}^2)$ are the same under (A.6) so using the above λ_i we have

$$\frac{X^2}{\hat{C}_1} \sim (a+\frac{b}{k-1})^{-1}(a\,\chi^2_{k-1} + b\,\chi^2_1) \qquad \text{(A.7)}$$

where $a = 1+(m_0-1)\rho$, $b = (m_0-1)(1-\rho)\delta$.

The coefficient of variation of the λ_i's is given by $c = b(k-2)^{1/2}/(b+(k-1)a)$ which tends to $(k-2)^{1/2}(m_0-1)/(m_0+k-2)$ as $\delta \to 1$ and $\rho \to 0$. As $\rho \to 1$ or $\delta \to 0$, (A.7) approaches χ^2_{k-1}. Using (A.7) and Satterthwaite's approximation (or the methods of Solomon and Stephens (1977)) we can compute the true significance level of $X^2/\hat{C}_1$ for any desired combination (m_0,ρ,k,δ).

A.4 Three-stage Sampling

Suppose the population consists of R psu's, M_t second-stage units (ssu's) in the t^{th} psu, and $K_{t\lambda}$ elements in the λ^{th} ssu of the t^{th} psu $(\sum_t\sum_\lambda M_{t\lambda} = N)$. The sample has r psu's, m_t ssu's from the t^{th} sampled psu and $k_{t\lambda}$ elements in the λ^{th} ssu of the t^{th} psu $(\sum_t\sum_\lambda k_{t\lambda} = n)$. Let $k_t = \sum_\lambda k_{t\lambda}$ be the number of sampled elements in the t^{th} psu. As in the case of two-stage sampling, we let $Z_{t\lambda si} = 1$ if the s^{th} element of the $(t,\lambda)^{th}$ ssu is in category i and $Z_{t\lambda si} = 0$ otherwise and we assume that

$$E(Z_{t\lambda si}) = p_i,\; \text{Cov}(Z_{t\lambda si},Z_{t\lambda s'j}) = b_{ij}\;\; (s \neq s'),$$

$$\text{Cov}(Z_{t\lambda si},Z_{t\lambda' s'j}) = d_{ij}\;\; (\lambda \neq \lambda'),$$

$$\text{Cov}(Z_{t\lambda si}, Z_{t'\lambda' s'j}) = 0 \quad (t \neq t').$$

It is easily shown that the covariance matrix of $\tilde{n}$ is given by

$$n\tilde{V} = n\tilde{P} + (\sum_t\sum_\lambda k_{t\lambda}^2 - n)\tilde{B} + (\sum_t k_t^2 - \sum_t\sum_\lambda k_{t\lambda}^2)\,\tilde{D}$$

where $\tilde{B} = (b_{ij})$ and $\tilde{D} = (d_{ij})$. We can show that

$\underset{\sim}{P} - \underset{\sim}{B}$ and $\underset{\sim}{P} - \underset{\sim}{D}$ are non-negative definite as in Section A.1, so that the eigenvalues of $\underset{\sim}{P}^{-1}\underset{\sim}{V}$ are all bounded above by $k_0 = \sum_t k_t^2/n$. Thus treating X^2/k_0 as a χ^2_{k-1} random variable provides a conservative test for H_0. In the special case with $m_t = m$, $k_{t\lambda} = k$ this reduces to X^2/mk. Thus the modification to X^2 that is required gets worse as we switch from two-stage sampling to three-stage sampling.

If $\underset{\sim}{B} = \rho_1 \underset{\sim}{P}$ and $\underset{\sim}{D} = \rho_2 \underset{\sim}{P}$, $\underset{\sim}{V}$ reduces to $\underset{\sim}{V} = C\underset{\sim}{P}$ where $C = 1 + \rho_1(\sum_t\sum_\lambda k_{t\lambda}^2/n - 1) + \rho_2(\sum_t k_t^2/n - \sum_t\sum_\lambda k_{t\lambda}^2/n)$, so we have another example of a model with constant design effect and $\tilde{X}^2 = X^2/C \sim \chi^2_{k-1}$ under H_0.

Estimators of C can be obtained from cluster totals in the same way as in Section A.2. For example,

$$\hat{C}_1 = (k-1)^{-1} \sum_1^k \hat{v}_{ii}/p_{0i}$$

where

$$\hat{v}_{ii} = \frac{r}{r-1} \sum_{t=1}^r (k_{ti} - \frac{n_i}{r})^2 - \frac{r}{r-1} p_{0i}^2 \sum_{t=1}^r (k_t - \frac{n}{r})^2 ,$$

with $k_{ti} = \sum_\lambda k_{t\lambda i}$, is an unbiased estimator.

A.5 Stratified Two-stage Sampling

We have R_h psu's in stratum h with M_{ht} ssu's in the t^{th} psu $(h=1,\ldots,L)$. A stratified two-stage sample consists of r_h psu's and m_h ssu's in the t^{th} sampled psu of stratum h. Let $Z_{ht\lambda i} = 1$ if the $(h,t,\lambda)^{th}$ element is in category i and assume the model of Section 4.1 holds in each stratum. Then, if $\underset{\sim}{n}_h = (n_{h1},\ldots, n_{h,k-1})'$ is the vector of category totals for stratum h , we have $E(\underset{\sim}{n}_h) = \tilde{n}_h \underset{\sim}{p}_h$, $\text{Cov}(\underset{\sim}{n}_h) = \tilde{n}_h\underset{\sim}{P}_h + \tilde{n}_h(m_h-1)\underset{\sim}{B}_h$ where $\tilde{n}_h = r_h m_h$. Let $\underset{\sim}{p} = \sum_1^L W_h \underset{\sim}{p}_h$ and $\hat{\underset{\sim}{p}} = \sum_1^L W_h \underset{\sim}{n}_h/\tilde{n}_h$ where $W_1,\ldots,W_L$ are the strata weights. We have $E(\hat{\underset{\sim}{p}}) = \underset{\sim}{p}$ and

$$\text{Cov}(\hat{\underset{\sim}{p}}) = \sum_1^L W_h^2(\underset{\sim}{P}_h + (m_h-1)\underset{\sim}{B}_h)/\tilde{n}_h = \underset{\sim}{V}/n .$$

In the case of proportional allocation with $\tilde{n}_h/n = W_h$, $\underset{\sim}{V}$ reduces to $\underset{\sim}{V} = \sum W_h(\underset{\sim}{P}_h + (m_h-1)\underset{\sim}{B}_h)$. Note that $\underset{\sim}{P}_h - \underset{\sim}{B}_h$ is non-negative definite as in Section A.1, so that $\underset{\sim}{V} \leq \sum W_h m_h \underset{\sim}{P}_h \leq m^*\sum W_h \underset{\sim}{P}_h = m^*[\underset{\sim}{P} - \sum W_h(\underset{\sim}{p}_h-\underset{\sim}{p})(\underset{\sim}{p}_h-\underset{\sim}{p})'] \leq m^*\underset{\sim}{P}$, where $m^* = \max(m_h)$. Thus the largest eigenvalue of $\underset{\sim}{P}^{-1}\underset{\sim}{V}$, $\lambda_1 \leq m^*$ and X^2/m^* provides a conservative test.

An important special case occurs when $\underset{\sim}{B}_h = \rho\underset{\sim}{P}_h$. Assuming $m_h = m$, we get

$$\underset{\sim}{V} = [1 + (m-1)\rho][\underset{\sim}{P} - \sum_1^L W_h(\underset{\sim}{p}_h-\underset{\sim}{p})(\underset{\sim}{p}_h-\underset{\sim}{p})'].$$

For the special case $L = 2$, we can obtain the eigenvalues explicitly. We have

$$\underset{\sim}{P}^{-1}\underset{\sim}{V} = [1 + (m-1)\rho][\underset{\sim}{I} - W_1W_2\underset{\sim}{P}^{-1}(\underset{\sim}{p}_1-\underset{\sim}{p}_2)(\underset{\sim}{p}_1-\underset{\sim}{p}_2)'].$$

Since the rank of $\underset{\sim}{P}^{-1}(\underset{\sim}{p}_1-\underset{\sim}{p}_2)(\underset{\sim}{p}_1-\underset{\sim}{p}_2)'$ is one, (k-2) of its eigenvalues roots are zero and the non-zero eigenvalue is $\sum_1^k (p_{1i} - p_{2i})^2/p_i$. Hence $\lambda_i = 1 + (m-1)\rho$, $i=1,\ldots,k-2$; $= [1 + (m-1)\rho](1-\delta)$, $i=k-1$, where $\delta = W_1W_2 \sum_1^k (p_{1i}-p_{2i})^2/p_i$ and $0 \leq \delta \leq 1$. Therefore

$$\frac{X^2}{\bar{\lambda}} \sim \frac{\chi^2_{k-2}}{1-\delta/(k-1)} + \frac{(1-\delta)}{1-\delta/(k-1)} \chi^2_1$$

where $\bar{\lambda} = (k-1)^{-1} \sum_1^k v_{ii}/p_i$. As $\delta \to 1$, $X^2/\bar{\lambda}$ approaches $[(k-1)/(k-2)]\chi^2_{k-2}$, $(k > 2)$. The coefficient of variation of the λ_i's is given by $c = (k-2)^{\frac{1}{2}}\delta/(k-1-\delta)$ which tends to $1/(k-2)^{\frac{1}{2}}$ as $\delta \to 1$.

If a value for ρ can be obtained, an appealing alternative test statistic is $X^{*2} = X^2/[1 + (m-1)\rho] \sim \chi^2_{k-2} + (1-\delta^*)\chi^2_1$. This provides a conservative test and should perform better than $X^2/\bar{\lambda}$.