Journal of Official Statistics
Vol. 9. No. 4, 1993, pp. 815–823

Transforming Hypotheses for Test of Homogeneity with Survey Data

Jeffrey R. Wilson and Mark Reiser[1]

Abstract: Hypotheses tested using X^2 (Pearson statistic) and G^2 (likelihood ratio) in SAS, BMDP, SPSSX and similar statistical programs assume that the data have been obtained from a simple random sampling scheme. In the use of survey data such an assumption is hardly ever met and adjustments must be made to the test statistic for the possible presence of extra-variation. Instead of making test statistic adjustments, this paper considers transformation of the original hypotheses and the construction of test statistics with familiar forms to investigate the transformed hypothesis. Thus the hypothesis is transformed when survey data are used and the form of the basic statistic is maintained.

Key words: Design effects; complex sampling; extravariation.

1. Introduction

In the analysis of survey data, which may include clustered and stratified data, the standard chi-squared X^2, and likelihood ratio G^2 test statistics as obtained from SAS, BMDP, SPSSX, or any modern statistical package greatly inflate the type I error rate when a strong, positive intra-cluster correlation is present. Thus, a researcher who is unaware of the effect of the design on the variance of estimates can easily produce meaningless results with X^2 and G^2. Recent works (Rao and Scott 1979, 1981; Brier 1980; Fay 1985; Wilson 1986; Koehler and Wilson 1986; Roberts, Rao and Kumar 1987; Anderson 1988; Wilson 1989; and Wilson and Koehler 1991), to name a few, have made adjustments to the X^2 and G^2 for testing hypotheses. Although some programs such as GLIM (Baker and Nelder 1978), CPLX (Fay 1988) and PCCARP (Fuller, Kennedy, Schnell, Sullivan, and Park 1987) have been adopted to deal with the adjusted statistics, the approach is still becoming popular. Unfamiliarity with these programs and for some the complexity involved in their use may force some researchers to use the unadjusted test statistics found in SAS, BMDP, and SPSSX.

When the variances of sample proportions exceed those implied by Poisson, binomial or multinomial distributions the sample proportions are often referred to as overdispersed. Some authors suggest that such a phenomenon may be caused by

[1]Department of Decision and Information Systems, Arizona State University, Tempe, AZ 85287-4206, U.S.A.
Acknowledgements: This research has been supported in part by a grant from the Faculty Grant-In-Aid Program at Arizona State University. An earlier draft of this paper was prepared for the Survey Research Methods Section, Annual Meetings of the American Statistical Association, Washington, D.C., August 1989. The authors would like to thank the Associate Editor and some anonymous reviewers for helpful hints on earlier drafts.

clumped sampling. Certainly it is the case that such extravariation may be caused by the use of complex sampling schemes to obtain sample data. Thus use of the Pearson chi-square test or likelihood ratio test for comparing vectors of proportions with such data should be avoided since such tests ignore any extravariation due to the complex sampling scheme. Any test that ignores overdispersion may cause a great inflation of the type I error rate.

Survey researchers in business and the social sciences are increasingly using complex sampling procedures in order to gain greater accuracy, to obtain more information, or to reduce the costs of estimating population characteristics. Examples include market research studies, and large nationwide surveys about labor force participation, health care, energy usage, and economic activity. Survey samples are now used in every field of scientific study including demography, transportation, health care, economics and so on. Indeed it is not an exaggeration to say that most of the data undergoing any form of statistical analysis are collected in surveys (Wolter 1985).

In order to make statistically valid conclusions from the results of complex surveys in the testing of hypotheses, it is necessary to use methods that have been developed during the last decade. Many survey researchers may not be familiar with these methods. The generally used statistical packages (SAS, SPSSX, and BMDP) do not include these methods, although some work has been done in SAS. The effects of using the present packages for complex surveys are well documented (Brier 1980; Rao and Scott 1979, 1981; Bedrick 1983; Wilson and Turner 1988; and Wilson 1989).

Different methods have been directed to finding appropriate test statistics for testing hypotheses when the data are obtained from complex sampling schemes. These methods can be summarized as (i) those making use of quadratic forms to construct Wald statistics (Wald 1943; Moore 1977; Forthofer and Koch 1973; Forthofer and Lehnen 1981; Grizzle, Starmer, and Koch 1969), (ii) those making use of probability models to describe the extravariation in the true vectors of proportions (Brier 1980; Wilson 1986; Koehler and Wilson 1986), (iii) those making use of partial information about the covariance matrix of the observed vectors of frequencies (Bedrick 1983; Rao and Scott 1979, 1981; Wilson 1989) and, (iv) the jackknifing of the usual Pearson X^2 statistic (Fay 1985). The performance of these methods was examined by Wilson and Turner (1988). In each of these methods the aim is to make adjustments or changes to the test statistic so as to better test the null hypothesis.

This paper examines the use of test statistics for comparing transformed vectors of proportions. These vectors are transformed based on an estimated function of the design effects matrix. Then the common procedure of comparing vectors by forming quadratic forms using the inverse of the covariance matrix is pursued. The work will assist survey researchers in better understanding how their survey results can defensibly be interpreted. Section 2 introduces the test of homogeneity and reviews the general construction of quadratic forms for checking the fit of categorical linear models. Transformation of hypotheses is considered in Section 3. A numerical example analyzed by Wilson (1989) is revisited in Section 4.

2. Test of Homogeneity

2.1. *Simple random sampling*

Consider comparing vectors of proportions from m subpopulations. This procedure was considered by Koehler and Wilson (1986) but reviewed here for ease of reference and to serve as a natural building block. The members of each subpopulation are classified into the same set of I mutually exclusive and exhaustive categories. For the jth subpopulation, the vector of true proportions for members in the various categories is denoted by the vector $\boldsymbol{\pi}_j = (\pi_{1j}, \pi_{2j}, \ldots, \pi_{Ij})'$. Estimates of the true vectors of proportions are obtained from independent two-stage cluster samples from each population. A sample of K_j clusters is randomly selected with replacement and with probability proportional to size (pps) from the jth subpopulation. Furthermore, a random sample of n_{jk} secondary units is selected with replacement from the kth cluster selected from the jth subpopulation and each sampled unit is classified into one of the I mutually exclusive categories. Conditionally on the cluster selected, the vector of observed frequencies for the kth cluster selected from the jth subpopulation, $\mathbf{X}_{jk} = (X_{1jk}, X_{2jk}, \ldots, X_{Ijk})'$, has a multinomial distribution with parameters n_{jk} and $\mathbf{p}_{jk} = (p_{1jk}, p_{2jk}, \ldots, p_{Ijk})'$, the true vector of proportions for the particular cluster selected.

A two-dimensional table of frequency totals can be constructed in which the rows correspond to the I categories and the columns correspond to the m subpopulations. The jth column of this table consists of the vector $\mathbf{X}_j = \sum_{k=1}^{Kj} \mathbf{X}_{jk}$ of total frequencies for the jth subpopulation. For the sampling scheme considered here, an unbiased estimator for $\boldsymbol{\pi}_j$ is

$$\hat{\boldsymbol{\pi}}_j = n_j^{-1}\mathbf{X}_j$$

where n_j is the total number of observations in the jth subpopulation.

Consider testing the null hypothesis

$$H_0 : \boldsymbol{\pi}_j = \boldsymbol{\pi}_0, \quad j = 1, 2, \ldots, m; \tag{1}$$

for some unknown vector of proportions $\boldsymbol{\pi}_0$, against the general alternative. When the null hypothesis is true an unbiased estimator of $\boldsymbol{\pi}_0$ is

$$\hat{\boldsymbol{\pi}}_0 = \sum_{j=1}^{m} \alpha_j \hat{\boldsymbol{\pi}}_j$$

where the weights α_j are known constants such that $1 = \sum_{j=1}^{m} \alpha_j$ and $\alpha_j \geq 0$, for $j = 1, 2, \ldots, m$. It can be shown (Wilson and Koehler 1984) that the covariance matrix for $\hat{\boldsymbol{\pi}}_j - \hat{\boldsymbol{\pi}}_0$ has diagonal elements

$$\mathbf{V}_{jj} = \mathbf{S}_j - 2\alpha_j\mathbf{S}_j + \sum_{t=1}^{m} \alpha_t^2 \mathbf{S}_t \tag{2}$$

where

$$\mathbf{S}_j = \mathbf{V}_{\mathrm{srs}(j)} + n_j^{-2}\left(\sum_{k=1}^{K_j} n_{jk}^2 - n_j\right) \times \sum_{l} w_{jl}(\mathbf{p}_{jl} - \boldsymbol{\pi}_j)(\mathbf{p}_{jl} - \boldsymbol{\pi}_j)'. \tag{3}$$

$\mathbf{V}_{\mathrm{srs}(j)}$ is the covariance matrix for the jth subpopulation under multinomial sampling and $\sum_l$ denotes the sum across all clusters in the jth subpopulation with weights w_{jl} equal to the proportion of the population in the lth cluster. The off-diagonal elements in the covariance matrix for $\hat{\boldsymbol{\pi}}_j - \hat{\boldsymbol{\pi}}_0$ is

$$\mathbf{V}_{ij} = -\alpha_i\mathbf{S}_i - \alpha_j\mathbf{S}_j + \sum_{t=1}^{m} \alpha_t^2 \mathbf{S}_t. \tag{4}$$

The matrix $\mathbf{V}$ with diagonal blocks given by (2) and off-diagonal blocks given by (4) is the covariance matrix for the vector of random deviations

$$\hat{\mathbf{d}} = (\hat{\boldsymbol{\pi}}_1' - \hat{\boldsymbol{\pi}}_0', \hat{\boldsymbol{\pi}}_2' - \hat{\boldsymbol{\pi}}_0', \ldots, \hat{\boldsymbol{\pi}}_m' - \hat{\boldsymbol{\pi}}_0')'.$$

A consistent and nearly unbiased estimator of $\mathbf{V}$ is obtained by replacing $\mathbf{S}_j$ in (2) and (4) with

$$\hat{\mathbf{S}}_j = t_j n_j^{-1}(K_j - 1)^{-1} \times \sum_{i=1}^{K_j} n_{ij}(\hat{\mathbf{p}}_{jk} - \hat{\boldsymbol{\pi}}_j)(\hat{\mathbf{p}}_{jk} - \hat{\boldsymbol{\pi}}_j)' + (1 - t_j)\hat{\mathbf{V}}_{\text{srs}(j)}$$

where

$$\hat{\mathbf{V}}_{\text{srs}(j)} = n_j^{-1}[\text{diag}(\hat{\boldsymbol{\pi}}_j) - \hat{\boldsymbol{\pi}}_j\hat{\boldsymbol{\pi}}_j']$$

$$t_j = (K_j - 1)\frac{\left(\sum_{k=1}^{K_j} n_{jk}^2 - n_j\right)}{\left(n_j^2 - (K_j - 1) - \sum_{k=1}^{K_j} n_{jk}^2\right)}.$$

We see that $\text{diag}(\hat{\boldsymbol{\pi}}_j)$ is a diagonal matrix with elements $\hat{\boldsymbol{\pi}}_j$, and $\hat{\mathbf{p}}_{jk} = n_{jk}^{-1}\mathbf{X}_{jk}$ is the vector of observed proportions for the kth cluster sampled from the jth subpopulation. The estimate of the covariance matrix for $\hat{\mathbf{d}}$ is denoted by $\hat{\mathbf{V}}$.

Clearly, $\mathbf{V}$ and $\hat{\mathbf{V}}$ are singular matrices. Nonsingular covariance matrices can be obtained by deleting some elements from $\hat{\mathbf{d}}$, but it is notationally more convenient to retain redundant differences in $\hat{\mathbf{d}}$ and use a generalized inverse of $\hat{\mathbf{V}}$ in the definition of test statistics. Consequently, a Wald statistic for testing the equality of the vectors of subpopulation proportions is

$$X_W^2 = \hat{\mathbf{d}}'\hat{\mathbf{V}}^-\hat{\mathbf{d}}$$

where $\hat{\mathbf{V}}^-$ is the generalized inverse of $\hat{\mathbf{V}}$. Following Moore (1977), this statistic has a limiting central chi-square distribution with degrees of freedom equal to the rank of $\mathbf{V}$ when the null hypothesis is correct. The statistic, X_W^2, reduces to the Pearson statistic

$$X_P^2 = \sum_{j=1}^{m} n_j \sum_{i=1}^{I} \hat{\pi}_i^{-1}(\hat{\pi}_{ij} - \hat{\pi}_i)^2$$

when $\hat{\mathbf{V}}$ is the usual estimate of the covariance matrix of $\hat{\mathbf{d}}$ for simple random sampling.

The accuracy of the large sample chi-square approximation for the null distribution of X_W^2 is greatly influenced by the accuracy of $\hat{\mathbf{V}}$ as an estimator of $\mathbf{V}$. A substantial number of sampled clusters is required to accurately estimate large covariance matrices. When a large number of clusters cannot be sampled from each population, it may be advantageous to describe variation among clusters within populations with a more parsimonious model. Kohler and Wilson (1986) used the Dirichlet-multinomial model with one parameter to account for among cluster variation within each population.

2.2 Extravariation model

Assume that the sample vector of observed proportions $\hat{\boldsymbol{\pi}}_j$ is distributed with mean vector $\boldsymbol{\pi}_j$ (vector of subpopulation proportions) and covariance $\mathbf{B}_j$, where $\mathbf{B}_j$ is a function of $\boldsymbol{\pi}_j$. Let n denote the overall sample size and n_j for $j = 1, \ldots, m$; denote the sample size for the jth subpopulation such that $n = \sum_{j=1}^{m} n_j$.

The multinomial model implies that the covariance matrix for the vector of proportions for randomly chosen units within the jth subpopulation is $\mathbf{V}_{\text{srs}(j)} = n_j^{-1}(\text{diag}\,\boldsymbol{\pi}_j - \boldsymbol{\pi}_j\boldsymbol{\pi}_j')$. Wilson (1989) referred to this as the simple model. Wilson defines the multinomial extravariation model variance for $\hat{\boldsymbol{\pi}}_j$ as

$$\mathbf{B}_j = \mathbf{H}_j^{1/2}\mathbf{V}_{\text{srs}(j)}\mathbf{H}_j^{1/2}$$

where $\mathbf{H}_j^{1/2} = \text{diag}\,\mathbf{h}_j^{1/2}$ and $\mathbf{h}_j = (h_{1j}, h_{2j}, \ldots, h_{Ij})'$ is a vector of unknown extravariation parameters of dimension I. This model suggests that the covariance matrix may be written as

$$\mathbf{B}_j = \mathbf{V}_{\text{srs}(j)} + \boldsymbol{\Delta}_j$$

where $\boldsymbol{\Delta}_j$ is the added variation beyond the multinomial assumption. Also assume that a central limit theorem for the specified design is available which ensures that $n_j^{1/2}(\hat{\pi}_j - \pi)$ converges in distribution to a normally distributed random variable with a zero vector and covariance matrix $\mathbf{B}_j$.

To derive a test for the null hypothesis $H_0 : \pi_j = \pi_0$, $j = 1, 2, \dots, m$; where π_0 is unknown, and estimated by the average vector of proportions $\hat{\pi} = \sum_{j=1}^{m} \alpha_j \hat{\pi}_j$, the vector of differences $\hat{\mathbf{d}} = (\hat{\pi}_1 - \hat{\pi}_0, \dots, \hat{\pi}_m - \hat{\pi}_0)'$ is used. For the extravariation model the covariance matrix $\boldsymbol{\Omega}$ for $\hat{\mathbf{d}}$ is composed of diagonal elements

$$\boldsymbol{\Omega}_{jj} = \mathbf{B}_j - 2\alpha_j \mathbf{B}_j + \sum_{t=1}^{m} \alpha_t^2 \mathbf{B}_t$$

and off-diagonal elements

$$\boldsymbol{\Omega}_{ij} = -\alpha_i \mathbf{B}_i - \alpha_j \mathbf{B}_j + \sum_{t=1}^{m} \alpha_t^2 \mathbf{B}_t.$$

A consistent estimator of the covariance matrix can be obtained by substituting $\hat{\pi}_j$ for π_j in $\mathbf{B}_j$ and using a consistent estimator for $\mathbf{H}_j^{1/2}$. Denoting the estimator for the covariance matrix by $\hat{\boldsymbol{\Omega}}$ and noting that the elements of $\hat{\mathbf{d}}$ sum to zero, a large sample chi-square test statistic for testing the equality of the vectors of population proportions is given by the quadratic form

$$X_{\Omega}^2 = \hat{\mathbf{d}}' \hat{\boldsymbol{\Omega}}^{-} \hat{\mathbf{d}}$$

where the degrees of freedom are equal to the rank of $\hat{\boldsymbol{\Omega}}$ (the estimated covariance of $\hat{\mathbf{d}}$). When the weights used for $\hat{\pi}$ are chosen inversely proportional to the size of the elements of the covariance matrices for the $\hat{\pi}_j'$s under H_0, i.e.,

$$\alpha_j = n_j \sum_{i}^{I} h_{iij}^{-1} \left(\sum_{t=1}^{m} n_t \sum_{l}^{I} h_{iil}^{-1} \right)^{-1} \tag{5}$$

where h_{iij} is the iith element of $\mathbf{H}_j$, then X_{Ω}^2 reduces to

$$X_{\Omega}^2 = \sum_{j=1}^{m} n_j \sum_{i=1}^{I} (h_{iij} \hat{\pi}_i)^{-1} (\hat{\pi}_{ij} - \hat{\pi}_i)^2. \tag{6}$$

The test statistic X_{Ω}^2 is equivalent to X_p^2 when $h_{iij} = 1$ for all i, $j = 1, 2, \dots, I$. So when the test statistic, X_p^2, is used to test $H_0 : \pi_j = \pi_0$ and $h_{iij} \neq 1$ for at least one of the categories then it is not the most suitable test provided in those programs. In such cases, the covariance matrix under the design is not the simple random sampling matrix $\mathbf{V}_{\text{srs}}$.

3. Test Statistics for Transformed Hypotheses

3.1 Transformed hypotheses

Consider a linear combination of π_j such that the transformed vector

$$\mathbf{q}_j = \mathbf{Q}_j \pi_j$$

where $\mathbf{Q}_j$ is a square matrix of constants. Then the hypothesis of interest $H_0 : \mathbf{q}_j = \mathbf{q}_0$ for $j = 1, 2, \dots, m$; can be tested using techniques similar to the ones adopted in Section 2. Let $\hat{\mathbf{q}}_j$ be an unbiased estimator for $\mathbf{q}_j$ and let $\hat{\mathbf{q}}_0$ be a linear combination of $\hat{\mathbf{q}}_j$, $j = 1, 2, \dots, m$. Then a Wald test statistic

$$X_Q^2 = \sum_{j=1}^{m} (\hat{\mathbf{q}}_j - \hat{\mathbf{q}}_0)' \times [\mathrm{Var}(\hat{\mathbf{q}}_j - \hat{q}_0)]^{-1} (\hat{\mathbf{q}}_j - \hat{\mathbf{q}}_0) = \hat{\mathbf{c}}_q' \hat{\mathbf{V}}_Q^{-1} \hat{\mathbf{c}}_q$$

where $\hat{\mathbf{c}}_q = \hat{\mathbf{q}}_j - \hat{\mathbf{q}}_0$ where $\hat{\mathbf{q}}_0 = \sum_{j=1}^{m} \gamma_j \hat{\mathbf{q}}_j$, and $\mathbf{V}_Q$ is the variance-covariance matrix of $\hat{\mathbf{q}}_j - \hat{\mathbf{q}}_0$ with diagonal and off-diagonal elements given respectively as

$$\mathbf{V}_{jjQ} = E_j - 2\gamma_j E_j + \sum_{t=1}^{m} \gamma_t^2 E_t$$

and

$$\mathbf{V}_{ijQ} = -\gamma_i E_i - \gamma_j E_j + \sum_{t=1}^{m} \gamma_t^2 E_t$$

where $E_j = \mathrm{Var}(\mathbf{Q}_j \hat{\pi}_j)$, Var denotes variance, and $0 < \gamma_i < 1$ such that $1 = \sum_{i=1}^{I} \gamma_i$. Under the extravariation model, the covariance matrix for the linear combination $\mathbf{Q}_j\hat{\boldsymbol{\pi}}_j$ is given as

$$\begin{aligned}\mathrm{Var}(\hat{\mathbf{q}}_j) &= \mathrm{Var}(\mathbf{Q}_j\hat{\boldsymbol{\pi}}_j) = \mathbf{Q}_j\mathbf{B}_j\mathbf{Q}_j' \\ &= n_j^{-1}\mathbf{Q}_j\mathbf{H}_j^{1/2}\mathbf{B}_{\mathrm{srs}(0)}\mathbf{H}_j^{1/2}\mathbf{Q}_j' \qquad (7)\end{aligned}$$

where $\mathbf{Q}_j$ represents a matrix of constants for the jth subpopulation, and $n_j^{-1}\mathbf{B}_{\mathrm{srs}(0)}$ is the covariance matrix under H_0 for simple random sampling such that

$$\mathbf{B}_{\mathrm{srs}(0)} = \mathrm{diag}(\boldsymbol{\pi}_0) - \boldsymbol{\pi}_0\boldsymbol{\pi}_0'.$$

When $\mathbf{Q}_j = \mathbf{H}_j^{-1/2}$, then $\mathbf{V}_{jjQ}$ takes on the value

$$\mathbf{V}_{\mathrm{srs}(j)} - 2\gamma_j\mathbf{V}_{\mathrm{srs}(j)} + \sum_{t=1}^{m}\gamma_t^2\mathbf{V}_{\mathrm{srs}(t)}$$

and $\mathbf{V}_{ijQ}$ takes the value

$$-\gamma_i\mathbf{V}_{\mathrm{srs}(i)} - \gamma_j\mathbf{V}_{\mathrm{srs}(j)} + \sum_{t=1}^{m}\gamma_t^2\mathbf{V}_{\mathrm{srs}(t)}.$$

Furthermore, examining $\mathbf{V}_Q$ for the case of $m = 3$ subpopulations which can be easily generalized to $m > 3$, (under H_0) the diagonal elements are

$$\left((1 - 2\gamma_1)n_1^{-1} + \sum_{t=1}^{m} n_t^{-1}\gamma_t^2\right)\mathbf{B}_{\mathrm{srs}(0)}$$

and off-diagonal elements are

$$\left(-(\gamma_1 n_1^{-1} + \gamma_2 n_2^{-1}) + \sum_{t=1}^{m} n_t^{-1}\gamma_t^2\right)\mathbf{B}_{\mathrm{srs}(0)}$$

When the γ_ts are chosen proportional to the sample size the matrix $\mathbf{V}_Q$ reduces to

$$\hat{\mathbf{M}}_{HO} = (\mathrm{Inv}(\mathrm{diag}\,\mathbf{n}) - n^{-1}\mathfrak{J})\otimes\mathbf{B}_{\mathrm{srs}(0)}$$

where Inv denotes inverse, $\mathbf{n} = (n_1, n_2)$, $\mathfrak{J}$ is a square matrix with elements equal to one and $\otimes$ represents the direct product between matrices. Then a test statistic for $H_0 : \mathbf{q}_j = \mathbf{q}_0$ where $\mathbf{q}_0$ is unknown is given by

$$X_q^2 = \hat{\mathbf{c}}_q'\hat{\mathbf{M}}_{HO}^{-1}\hat{\mathbf{c}}_q$$

and $\hat{\mathbf{M}}_{HO}$ is a consistent estimator of $\mathbf{M}_{HO}$. One consistent estimator can be obtained by replacing $\boldsymbol{\pi}_0$ with $\hat{\boldsymbol{\pi}}_0$ in $\mathbf{M}_{HO}$. Then for $\hat{\mathbf{B}}_{\mathrm{srs}(0)} = \mathrm{diag}(\hat{\boldsymbol{\pi}}_0) - \hat{\boldsymbol{\pi}}_0\hat{\boldsymbol{\pi}}_0'$

$$\begin{aligned}\hat{\mathbf{M}}_{HO}^{-1} &= (\mathrm{diag}\,\mathbf{n} + n_3^{-1}\mathbf{n}\mathbf{n}')\otimes\hat{\mathbf{B}}_{\mathrm{srs}(0)}^{-1} \\ &= \begin{bmatrix} [n_1 + n_3^{-1}n_1^2]\hat{\mathbf{B}}_{\mathrm{srs}(0)}^{-1} & n_3^{-1}n_1n_2\hat{\mathbf{B}}_{\mathrm{srs}(0)}^{-1} \\ n_3^{-1}n_1n_2\hat{\mathbf{B}}_{\mathrm{srs}(0)}^{-1} & [n_2 + n_3^{-1}n_2^2]\hat{\mathbf{B}}_{\mathrm{srs}(0)}^{-1}\end{bmatrix}\end{aligned}$$

$$X_q^2 = \sum_{j=1}^{3} n_j(\hat{\mathbf{q}}_j - \hat{\mathbf{q}}_0)'\hat{\mathbf{B}}_{\mathrm{srs}(0)}^{-1}(\hat{\mathbf{q}}_j - \hat{\mathbf{q}}_0).$$

For $m > 3$ the generalization is obvious. Additionally, the test statistic for $H_0 : \mathbf{q}_j = \mathbf{q}_0$ reduces to

$$X_q^2 = \sum_{j=1}^{m} n_j \sum_{i=1}^{I}(\hat{q}_{ij} - \hat{q}_i)^2\hat{\pi}_i^{-1} \qquad (8)$$

where $\hat{\mathbf{q}}_j = (\hat{q}_{1j}, \ldots, \hat{q}_{Ij})'$, and $\hat{\mathbf{q}}_0 = (\hat{q}_{10}, \ldots, \hat{q}_{I0})'$ are unbiased estimates of $\mathbf{q}_j$ and $\mathbf{q}_0$, respectively. Note that the form of X_q^2 is similar in structure to X_p^2, the Pearson statistic, and to X_Ω^2 in (6). The mean and variance for X_q^2 were obtained empirically based on a simulation of data for a stratified two-stage cluster sampling scheme. The simulation was conducted based on procedures used by Wilson and Turner (1988). The test statistic X_q^2 performed satisfactorily when compared to a chi-square random variable with the associated degrees of freedom.

Consider $n_j^{1/2}(\hat{\boldsymbol{\pi}}_j - \boldsymbol{\pi})$ as an asymptotically $(\mathbf{I} - 1)$ multivariate normal random vector with mean 0 and covariance matrix $\mathbf{D}_{(j)}$ say. Suitable central limit theorems (Fuller 1975; Krewski and Rao 1978) ensure that $n_j^{1/2}(\hat{\boldsymbol{\pi}}_j - \boldsymbol{\pi})$ is asymptotically normal. So for fixed $\mathbf{Q}_j$, the linear combination $n_j^{1/2}\mathbf{Q}_j(\hat{\boldsymbol{\pi}}_j - \boldsymbol{\pi})$ is also asymptotically normally distributed. The asymptotic distribution of X_q^2 follows (Moore 1977).

Testing $H_0 : \mathbf{q}_j = \mathbf{q}_0$ using X_q^2 is asymptotically equivalent to testing $H_0 : \boldsymbol{\pi}_j = \boldsymbol{\pi}_0$

with X^2_Ω. Thus a researcher can examine the transformed hypothesis and use X^2_q to obtain information of the homogeneity of $\boldsymbol{\pi}_j, j = 1, 2, \ldots, m$.

3.2 Estimating generalized design matrix

Computing X^2_q requires an estimate of $\mathbf{Q}_j$, $j = 1, 2, \ldots, m$; or equivalently the matrix $\mathbf{H}_j$. An estimate of $\mathbf{H}_j$ can be obtained from the product of $\boldsymbol{\Omega}_j$ and $\mathbf{V}^{-1}_{\text{srs}(j)}$, i.e., the design matrix and the inverse of matrix under simple random sampling. A priori knowledge of the design effects may be available and can be used in estimating the generalized design matrix.

The covariance matrix $\boldsymbol{\Omega}_j$ depends on $\boldsymbol{\pi}_j$ and on the unknown vector of parameters $\mathbf{h}$. Consider for the jth subpopulation so with what follows we ignore the index j. The diagonal elements of $\boldsymbol{\Omega}$ are $\sigma_{ii} = h_i(\pi_i - \pi_i^2)$ for $i = 1, 2, \ldots, I$; and off-diagonal elements are $\sigma_{it} = -h_i^{1/2} h_t^{1/2} \pi_i \pi_t$ for $i \neq t = 1, 2, \ldots, I$. Thus, if a consistent estimator $\hat{\boldsymbol{\Omega}} = (\hat{\sigma}_{it})$ is available, then a consistent estimator for h_i is $\hat{h}_i = n\hat{\sigma}_{ii}/\hat{\pi}_i(1 - \hat{\pi}_i)$.

4. Numerical Example

Wilson (1989) using a vector of design effects analyzed data from a study of housing satisfaction performed for the U.S. Department of Agriculture (Brier 1980). Households in the vicinity of Monte-video, Minnesota were stratified into two populations: those in the metropolitan area and those outside of the metropolitan area. A random sample of 20 neighborhoods was taken from each population, and five households were randomly selected from each of the sampled neighborhoods. One response was obtained from the residents of each household concerning their satisfaction with their home. The possible responses were "unsatisfied" (US), "satisfied" (S), and "very satisfied" (VS). Only data from neighborhoods in which responses were obtained from each of the five sampled households are used here to illustrate the evaluation of the test statistics. This reduces the original data set to $K_1 = 18$ households from the nonmetropolitan area and $K_2 = 17$ households from the metropolitan area. These data were also analyzed by Koehler and Wilson (1986). They used the statistic $X^2_{\text{DMB}} = \sum_{j=1}^m n_j \hat{e}_j^{-1} \sum_{i=1}^I \hat{\pi}_i^{-1} (\hat{\pi}_{ij} - \hat{\pi}_i)^2$ where $\hat{e}_j$ is an estimate of the overdispersion. The estimated vector of proportions for the nonmetropolitan area is $\hat{\boldsymbol{\pi}}_1 = (.522, .422, .056)'$, and the corresponding estimated vector of proportions for the metropolitan area is $\hat{\boldsymbol{\pi}}_2 = (.353, .506, .141)'$.

There is interest in comparing the vectors of probabilities for the two subpopulations. In fact there is some indication that residents of the metropolitan area are less satisfied with their homes. The value of the usual Pearson chi-square (Koehler and Wilson 1986) is $X^2_p = 6.81$ on two degrees of freedom. Wilson (1989) reports that the test statistic is substantially reduced when the effects of the design are accounted for. The statistic X^2_Ω has a value of 4.43. Consider testing the transformed hypothesis $H_0 : \mathbf{Q}_1 \boldsymbol{\pi}_1 = \mathbf{Q}_2 \boldsymbol{\pi}_2$. From the design effects obtained from previous investigations for these data, an estimate of $\mathbf{Q}_j = \mathbf{H}_j^{-1/2}$ is the diagonal matrix

$$\hat{\mathbf{H}}_1^{-1/2} = \text{diag}[(2.088)^{-1/2}, (1.990)^{-1/2}, (1.134)^{-1/2}]$$

and $\mathbf{H}_2^{-1/2}$ is estimated by the diagonal matrix

$$\hat{\mathbf{H}}_2^{-1/2} = \text{diag}[(2.357)^{-1/2}, (1.812)^{-1/2}, (0.982)^{-1/2}].$$

It follows that if one was interested in test-

Table 1. Comparison of test statistics for hypotheses in survey data

Test statistic	Covariance matrix	Hypothesis	Comments
X_p^2 (Pearson)	Multinomial model	$H_0 : \pi_j = \pi_0$	No adjustment
X_Ω^2 (Wilson 1989)	Multinomial extravariation model	$H_0 : \pi_j = \pi_0$	Adjustment to covariance
X_q^2 (Section 3)	Same form as multinomial model	$H_0 : \mathbf{q}_j = \mathbf{q}_0$	Adjustment to hypothesis
X_{DMB}^2 (Koehler and Wilson 1986)	Heterogeneity factor times multinomial model	$H_0 : \pi_j = \pi_0$	Adjustment to covariances

ing the hypothesis $H_0 : \pi_1 = \pi_2$ (with the available design effect matrix) then one may test the transformed hypothesis

$$H_0 : \begin{pmatrix} 0.69\pi_{11} \\ 0.71\pi_{12} \\ 0.94\pi_{13} \end{pmatrix} = \begin{pmatrix} 0.65\pi_{21} \\ 0.74\pi_{22} \\ 1.01\pi_{23} \end{pmatrix}$$

using the test statistic X_q^2. The statistic has a value of 5.75 on 2 degrees of freedom. Thus there is no significant difference in the transformed vectors. The testing of this transformed hypothesis uses a test that does not inflate the type I error. The form is very familiar and the computations are quite simple.

5. Discussion

The research presented thus far regarding extravariation and overdispersion for data obtained from complex sampling designs suggests that some adjustments must be made to the covariance matrix to account for the extravariation. It is generally agreed that the usual Pearson statistic X_p^2 is not adequate for testing hypotheses based on survey data. It suffers from inflation of type I error level.

The present research suggests obtaining a transformed hypothesis through the use of a function of the cell design effects and then using a form of Wald statistic. Table 1 gives a summary of some test statistics used with extravariation. The test statistic, X_Ω^2, makes adjustments to the covariance matrix. The X_q^2 developed in Section 3 makes adjustments to the hypothesis. This paper suggests that the X_q^2 for testing the adjusted hypothesis does not suffer from the effects associated with the use of certain test statistics on overdispersed data.

6. References

Anderson, D.A. (1988). Some Models for Overdispersed Binomial Data. Australian Journal of Statistics, 30, 125–148.

Baker, R.J. and Nelder, J.A. (1978). General Linear Interactive Modelling (GLIM). Release 3. Oxford: Numerical Algorithms Group.

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

Brier, S.S. (1980). Analysis of Contingency Tables Under Cluster Sampling. Biometrika, 67, 591–596.

Fay, R.E. (1985). A Jackknifed Chi-Square Test for Complex Samples. Journal of the American Statistical Association, 80, 148–157.

Fay, R.E. (1988). CPLX Contingency Table Analyses for Complex Sample Designs. Program Documentation, U.S. Bureau of the Census, Washington, D.C.

Forthofer, R.N. and Koch, G.G. (1973). A Program for Compounded Functions of Categorical Data. Biometrics, 2, 143–157.

Forthofer, R.N. and Lehnen, R.G. (1981). Public Program Analysis. A New Categorical Data Approach. California: A Division of Wadsworth, Inc.

Fuller, W.A. (1975). Regression Analysis for Sample Surveys. Sankhya C, 37, 117–132.

Fuller, W.A., Kennedy, W., Schnell, D., Sullivan, G., and Park, H.J. (1987). PC CARP. Ames, Iowa: Statistical Laboratory, Iowa State University.

Grizzle, J.E., Starmer, C.F., and Koch, C.G. (1969). Analysis of Categorical Data by Linear Models. Biometrics, 33, 133–158.

Koehler, K.J. and Wilson, J.R. (1986). Chi-Square Tests for Comparing Vectors of Proportions for Several Cluster Samples. Communications in Statistics, A15, 2977–2990.

Krewski, D. and Rao, J.N.K. (1978). Inference from Stratified Samples I: Large Sample Properties of the Linearization, Jackknife and Balanced Repeated Replication Methods. Carleton Math., Ser. No. 155, Carleton University, Ottawa.

Moore, D.S. (1977). Generalized Inverses, Wald's Method, and the Construction of Chi-Squared Tests-of-Fit. Journal of the American Statistical Association, 72, 131–137.

Rao, J.N.K. and Scott, A.J. (1979). Chi-Squared Test for Analysis of Categorical Data from Computer Survey Data. Proceedings of the Section on Survey Research Methods, American Statistical Association, 58–66.

Rao, J.N.K. and Scott, A.M. (1981). The Analysis of Categorical Data from Complex Surveys. Journal of the American Statistical Association, 76, 221–230.

Roberts, G., Rao, J.N.K., and Kumar, S. (1987). Logistic Regression Analysis of Sample Survey Data. Biometrika, 74, 1–12.

Wald, A. (1943). Tests of Statistical Hypotheses. Transaction of American Society, 54, 426–482.

Wilson, J.R. and Koehler, J. (1984). Testing Equality of Vectors of Proportions for Several Cluster Samples. Proceedings of the Section on Survey Research Methods, American Statistical Association, 201–206.

Wilson, J.R. (1986). Approximation Distribution and Test of Fit for the Clustering Effect in the Dirichlet-Multinomial Model. Communications in Statistics-Theory and Methods, 15, 1235–1249.

Wilson, J.R. (1989). Chi-square Tests for Over-Dispersion with Multiparameter Estimates. Applied Statistics, 38, 441–453.

Wilson, J.R. and Koehler, K.J. (1991). Hierarchical Models for Cross-Classified Overdispersed Multinomial Data. Journal of Business and Economics Statistics, 9, 103–110.

Wilson, J.R. and Turner, D. (1988). A Simulated Comparison of Test Statistics for Comparing Vectors of Proportions. DIS Technical Report, Arizona State University, Arizona.

Wolter, K.M. (1985). Introduction to Variance Estimation. New York: Springer-Verlag.

Received May 1990
Revised June 1993