TESTING EQUALITY OF VECTORS OF PROPORTIONS FOR SEVERAL CLUSTER SAMPLES

Jeffrey R. Wilson, Oklahoma State University
Kenneth J. Koehler, Iowa State University

SUMMARY

A test statistic for comparing proportions should reflect the type of sampling scheme used to collect the data. This paper considers obtaining a test statistic for proportions under cluster sampling. Brier (1980) used the Dirichlet Multinomial model for contingency tables generated by cluster sampling schemes. Here the Dirichlet Multinomial is used for comparing vectors of proportions from different subpopulations. This model assumes that the covariance matrix of the estimated proportion under Dirichlet Multinomial sampling is a multiple of the covariance matrix of the estimated proportions under multinomial sampling. Several methods are suggested for estimating this multiplier. A test for the fit of the model is obtained. The Dirichlet Multinomial model is used to analyze data taken from Brier (1980).

1. INTRODUCTION

In the analysis of categorical data most methods have been developed extensively assuming multinomial sampling. Under this assumption it is well known that the usual goodness-of-fit statistics, namely the Pearson chi-squared X^2 and the likelihood ratio chi-squared statistic, G^2 have asymptotic chi-squared distributions if the fitted model is correct.

In this paper we develop a test of proportions for several subpopulations under a Dirichlet Multinomial model. This model assumes that the covariance matrix of the proportions under the Dirichlet Multinomial model is a multiple of the covariance matrix for the proportions under multinomial sampling. A test is obtained for the fit of the model and several methods are developed for estimating this multiplier.

2. DIRICHLET MULTINOMIAL DISTRIBUTION

Consider a population consisting of several clusters. A sample of s clusters is randomly chosen with replacement and with probability proportional to size (pps). A simple random sample of n secondary units is taken from each cluster, and the total sample size is N = ns.

Let $\underset{\sim}{p}_t = (p_{1t}, p_{2t}, \ldots, p_{It})'$ for t = 1,2, ...s be the vector of proportions for the t-th cluster of the population. Assume that the vector $\underset{\sim}{p}_t$ is distributed independently and identically with distribution function F(p). The distribution F(p) is the Dirichlet distribution. Good (1965) described the Dirichlet distribution as a conjugate prior distribution for the cell probabilities for multinomial models. It has density function

$$f(\underset{\sim}{p}_t|\underset{\sim}{\pi},k) = \frac{\Gamma(k)}{\prod\limits_{i=1}^{I} \Gamma(k\pi_i)} \prod_{i=1}^{I} p_i^{k\pi_i - 1} .$$

The parameters are the vector $\pi = (\pi_1, \pi_2, \ldots \pi_I)'$ and $k(> 0)$. k is a scaling parameter.

With a Dirichlet prior the unconditional distribution of $\underset{\sim}{X}_t - (X_{it}, X_{2t}, \ldots X_{It}),'$ is given by

$$f(\underset{\sim}{X}_t|\underset{\sim}{\pi}, k) = \binom{n}{X_{1t}, X_{2t}, \ldots, X_{It}} \; X \; \frac{\Gamma(k)}{\Gamma(n+k)} \prod_{i=1}^{I} \left\{ \frac{\Gamma(X_{it} + k\pi_i)}{\Gamma(k_{\pi i})} \right\}.$$

We denote this unconditional distribution by $DM_I(n, \underset{\sim}{\pi}, k)$. Within each cluster the vector of category counts

$\underset{\sim}{X}_t = (X_{1t}, X_{2t}, \ldots, X_{It})$, t = 1, 2, ...s;

has a multinomial distribution, with parameters n and $\underset{\sim}{p}_t$, conditional upon $\underset{\sim}{p}_t$. Wilks (1962) gives the moments of the Dirichlet distribution as

$E(p_i) = \pi_i$,

$E(p_i^2) = \pi_i (1 + k)^{-1} (1 + k\pi_i)$

and

$E(P_i p_i') = (1 + k)^{-1} \, k\pi_i \pi_i'$, $i \neq i'$.

The moments of the conditional distribution i.e. the multinomial distribution are well known. Thus the moments for the Dirichlet Multinomial are

$E(\underset{\sim}{X}_t) = n \, \underset{\sim}{\pi}$,

$V(\underset{\sim}{X}_t = n \, C(\Delta_\pi - \underset{\sim}{\pi} \, \underset{\sim}{\pi}')$,

where Δ_π is a diagonal matrix with elements π_i, and

$C = (1 + k)^{-1} (n + k)$.

Brier (1980) gives some methods of estimating C. In section 3.4 we obtain other methods of estimating C and obtain a lack-of-fit statistic for the model.

2.2 Extension of the Dirichlet Multinomial Model

The description given so far concentrates on one subpopulation so as to give a basic idea of the Dirichlet Multinomial model. Consider extending this basic idea for J > 1 subpopulations. Assume that for j th subpopulation, j = 1, 2, ... J; the number of sampled clusters within that subpopulation is S_j, and a simple random sample of size n_j is taken from each of the sampled clusters with replacement. Let $\underset{\sim}{X}_{t_j}$ denote the vector of counts for the t th cluster with the j th subpopulation. Further, assume that

$\underset{\sim}{X}_{t_j} \sim$ iid $DM_I(n_j, \underset{\sim}{\pi}_j, k_j)$.

This model permits a different distribution for the vectors of proportions within each subpopulation. Define

$$\underset{\sim}{X}_j = \sum_{t=1}^{S_j} \underset{\sim}{X}_{t_j} \quad j = 1, 2, \ldots J;$$

Then, the covariance matrix for the vector of sample totals, $\underset{\sim}{X}_j$, is

$Var(\underset{\sim}{X}_j) = n_j S_j C_j (\Delta_{\underset{\sim}{\pi}_j} - \underset{\sim}{\pi}_j \, \underset{\sim}{\pi}_j')$

where

$C_j = (1 + k_j)^{-1} (n_j + k_j)$.

Let the total sample size for the j th subpopulation be denoted by

$N_j = n_j S_j$

then,

$Var(\underset{\sim}{X}_j) = N_j C_j (\Delta_{\underset{\sim}{\pi}} - \underset{\sim}{\pi}_j \, \underset{\sim}{\pi}_j')$.

Define a vector of observed proportions for the j th subpopulation as

$$\hat{\pi}_j = N_j^{-1} X_j,$$

then $\hat{\pi}_j$ is an unbiased estimator of π_j. The covariance matrix for $\hat{\pi}_j$ is

$$B_j = N_j^{-1} C_j (\Delta_{\pi_j} - \pi_j \pi_j').$$

The vector of deviations $(\hat{\pi}_j - \pi_j)$ has a mean vector 0 and covariance matrix $\hat{B}_j$. By the Central Theorem as $S_j \to \infty$, $\sqrt{N_j}(\hat{\pi}_j - \pi_j) \to \mathcal{N}(0, N_j B_j)$.

3. TESTING OF HYPOTHESES CONCERNING π_j

3.1 Test of Homogeineity

Suppose the interest is in testing the hypothesis

$$H_o: \pi_j = \pi_o \quad j = 1, 2, \ldots J; \tag{1}$$

where π_o is a known vector. Then, $\hat{\pi}_j - \pi_o$ is an unbiased estimator for the vector $\pi_j - \pi_o$, and the variance of $\hat{\pi}_j - \pi_o$ is given by B_j. Under the null hypothesis a consistent estimator of B_j is

$$\hat{B}_j = N_j^{-1} \hat{C}_j (\Delta_{\pi_o} - \pi_o \pi_o'), \tag{2}$$

where $\hat{C}_j$ is a consistent estimator for C_j. Then, a test of the form considered by Wald (1943) is

$$X^2_{DMH} = (\hat{\pi} - \pi_o)' \hat{B}^- (\hat{\pi} - \pi_o) \tag{3}$$

where

$$(\hat{\pi} - \pi_o) = (\hat{\pi}_1' - \pi_o', \hat{\pi}_2' - \pi_o', \ldots, \hat{\pi}_J' - \pi_o')'$$

and $\hat{B}$ is a consistent estimator of Var$(\hat{\pi} - \pi_o)$ and $\hat{B}^-$ is the generalized Moore-Penrose inverse of $\hat{B}$. $\hat{B}$ is a block diagonal matrix with entries given by $\hat{B}_j$, j = 1, 2, ...J; on the diagonal. The statistic X^2_{DMH} can be written as:

$$X^2_{DMH} = \sum_{j=1}^{J} (\hat{\pi}_j - \pi_o)' \hat{B}_j^- (\hat{\pi}_j - \pi_o) \tag{4}$$

$$= \sum_{j=1}^{J} \hat{C}_j^{-1} X^2_{mj}$$

where

$$X^2_{mj} = N_j (\hat{\pi}_j - \pi_o)' (\Delta_{\pi_o} - \pi_o \pi_o')^- (\hat{\pi}_j - \pi_o) \tag{5}$$

denotes a Pearson goodness-of-fit statistic for the j th subpopulation and $\hat{C}_j$ is a consistent estimator for C_j, j = 1, 2, ...J. Since $\sqrt{N_j}(\hat{\pi}_j - \pi_o)$ has a limiting normal distribution with mean vector 0 and covariance matrix B_j, a test statistic based on a consistent estimator of B has an asymptotic chi-square distribution, (Moore, 1977).

Therefore, X^2_{DMH} has a limiting chi-square distribution with J(I-1) degress of freedom since B has a rank of J(I-1).

3.2 A Test of Independence

Suppose the vector π_o is unknown in the hypothesis

$$H_o: \pi_j = \pi_o \tag{6}$$

and is estimated by a linear combination of the J unbiased estimators obtained from each of the J subpopulations. Then, an estimator for π_o is

$$\hat{\pi}_o = \sum_{j=1}^{J} \alpha_j \pi_j, \tag{7}$$

where the α_j's are positive and sum to one. The estimator $(\hat{\pi}_j - \hat{\pi}_o)$ is an unbiased estimator of $\pi_j - \pi_o$ if the α's are fixed. Let T_{jj} denote the variance of $\hat{\pi}_j - \hat{\pi}_o$ and $T_{jj'}$, denote the covariance between $(\hat{\pi}_j - \hat{\pi}_o)$ and $(\hat{\pi}_{j'} - \hat{\pi}_o)$. Then

$$T_{jj} = B_j - 2\alpha_j^{-1} B_j + \sum_{\ell=1}^{J} \alpha_\ell^{-2} B_\ell \tag{8}$$

and

$$T_{jj'} = -\alpha_j^{-1} B_j - \alpha_{j'}^{-1} B_{j'} + \sum_{\ell=1}^{J} \alpha_\ell^{-2} B_\ell,$$

a test statistic for the hypothesis in (2) where π_o is an unknown vector, can be constructed using a consistent estimator of the covariance matrix for

$$(\hat{\pi} - \hat{\pi}_o) = (\hat{\pi}_1' - \hat{\pi}_o', \hat{\pi}_2' - \hat{\pi}_o', \ldots, \hat{\pi}_I' - \hat{\pi}_o')'.$$

Let M denote the covariance matrix for $(\hat{\pi} - \hat{\pi}_o)$. Then M has T_{jj} as the j th diagonal block and $T_{jj'}$ as the corresponding off diagonal block. Let M^- denote the generalized Moore-Penrose inverse of M. In estimating M, B_j is replaced by

$$\hat{B}_j = N_j^{-1} C_j (\Delta_{\hat{\pi}_o} - \hat{\pi}_o \hat{\pi}_o').$$

By Wilson (1984) a Wald test statistic as considered by Wald (1943) is given by

$$X^2_{DMI} = (\hat{\pi} - \hat{\pi}_o)' \hat{M}^- (\hat{\pi} - \hat{\pi}_o) = \sum_{j=1}^{J} N_j \hat{C}_j^{-1} \left\{ \sum_{i=1}^{I} \frac{(\hat{\pi}_{ij} - \hat{\pi}_{io})^2}{\hat{\pi}_{io}} \right\} \tag{9}$$

where

$$\hat{\pi}_{io} = \sum_{j=1}^{J} \alpha_j \pi_{ij}$$

and the α_j's are chosen inversely proportional to the variance, such that

$$\alpha_j = N_j \hat{C}_j^{-1} \left[\sum_{\ell=1}^{J} N_\ell C_\ell^{-1} \right]^{-1}. \tag{10}$$

The Wald test statistic X^2_{DMI} resembles the usual Pearson statistic for a test of independence under multinomial sampling. However, it differs from the Pearson statistic in that the multiplier $\sum_{i=1}^{I} (\hat{\pi}_{ij} - \hat{\pi}_{io})^2 \hat{\pi}_{io}^{-1}$ is $N_j \hat{C}_j^{-1}$ and not N_j, also $\hat{\pi}_{io}$ is not the maximum likelihood estimator obtained under multinomial sampling. If $C_j \equiv 1$ for all j, then X^2_{DMI} reduces to the usual Pearson statistic

$$X^2_{MI} = \sum_{j=1}^{J} N_j \sum_{i=1}^{I} \frac{(\hat{\pi}_{ij} - \hat{\pi}_{io})^2}{\hat{\pi}_{io}}. \tag{11}$$

If the C_j's are all equal to some value C then the statistic reduces to a multiple of the usual Pearson statistic,

$$X^2_{DMI} = \hat{C}\, X^2_{MI}$$

3.3 Estimation of C.

To be able to use X^2_{DMI} an estimate of each C_j is required. Altham (1976) and Cohen (1976) point out the fact that $1 \leq C_j \leq n_j$, and this can be used to obtain a conservative test. Brier (1980) gave a possible method of estimating C_j based on the method of moments, as a multiple of the Pearson Chi-square statistic for testing equality of the S_j probability vectors in the I x S_j table formed by classifying units by clusters and by categories. Here we consider several different estimators for which we compute in the example given below.

Under the Dirichlet Multinomial model each of the I^2 elements of the covariance matrix is C times the corresponding elements under the multinomial model. Recall that the variance of $\hat{\underset{\sim}{\pi}}_j$ under the Dirichlet Multinomial model is

$$\mathrm{Var}_{DM}(\hat{\underset{\sim}{\pi}}_j) = n_j^{-1} C_j (\Delta_{\underset{\sim}{\pi}_j} - \underset{\sim}{\pi}_j\, \underset{\sim}{\pi}_j{}') \qquad (12)$$

and under the multinomial model the variance is

$$\mathrm{Var}_M(\hat{\underset{\sim}{\pi}}_j) = n_j^{-1} (\Delta_{\underset{\sim}{\pi}_j} - \underset{\sim}{\pi}_j\, \underset{\sim}{\pi}_j{}'). \qquad (13)$$

A simple moment estimator for $\mathrm{var}_{DM}(\hat{\underset{\sim}{\pi}}_j)$ is

$$\mathrm{Var}_{DM}(\hat{\underset{\sim}{\pi}}_j) = n_j^{-1}(s_j-1)^{-1} \sum_{t=1}^{S_j} (\hat{\underset{\sim}{\pi}}_{tj} - \hat{\underset{\sim}{\pi}}_j)(\hat{\underset{\sim}{\pi}}_{tj} - \hat{\underset{\sim}{\pi}}_j)', \qquad (14)$$

where

$$\hat{\underset{\sim}{\pi}} = S_j^{-1} \sum_{t=1}^{S_j} \hat{\underset{\sim}{\pi}}_j .$$

The sample covariance matrix in (14) can be expressed in a vector form by writing

$$\mathrm{vech}[\mathrm{var}(\hat{\underset{\sim}{\pi}}_j)] = n_j^{-1}(S_j-1)^{-1} \begin{bmatrix} \sum_{t=1}^{S_j} (\hat{\pi}_{1tj}-\hat{\pi}_{1j})2 \\ \sum_{t=1}^{S_j} (\hat{\pi}_{1tj}-\hat{\pi}_{1j})(\hat{\pi}_{2tj}-\hat{\pi}_{2j}) \\ \vdots \quad \vdots \quad \vdots \\ \sum_{t=1}^{S} (\hat{\pi}_{J-1,tj}-\hat{\pi}_{(I-1)j})^2 \end{bmatrix}$$

$$= n_j^{-1} \hat{\underset{\sim}{V}}_j . \qquad (15)$$

The expected value for vech $[\mathrm{var}\,(\hat{\underset{\sim}{\pi}}_j)]$ is

$$E\{n_j^{-1}\hat{\underset{\sim}{V}}_j\} = n_j^{-1} \begin{bmatrix} \pi_{ij}^2 \\ \pi_{1j}\,\pi_{2j} \\ \vdots \\ \pi^2_{I-1j} \end{bmatrix}$$

A graph of the elements of the covariance matrix in (14) versus the corresponding elements of the covariance matrix in (13) should resemble a straight line when the model is true. This straight line must pass through the origin. Then, a generalized least squares estimator for C_j is given by

$$\hat{C}_{jW} = (\hat{\underset{\sim}{w}}_j{}'\hat{\Sigma}_j^{-1}\hat{\underset{\sim}{w}}_j)^{-1} w_j{}'\hat{\Sigma}_j^{-1}\hat{\underset{\sim}{V}}_j \qquad (16)$$

where $\hat{\Sigma}_j$ is a consistent estimator of the covariance matrix for $\hat{\underset{\sim}{V}}_j$ and

$$\hat{\underset{\sim}{w}}_j = (\hat{\pi}^2{}_{ij},\ \hat{\underset{\sim}{\pi}}_{ij}\,\hat{\underset{\sim}{\pi}}_{2j},\ \ldots\ldots,\ \hat{\pi}^2{}_{I-1,j})'. \qquad (17)$$

The generalized least squares technique for estimating C_j also provides an approximate goodness-of-fit test for the model. This estimation procedure assumes that the $\underset{\sim}{W}_j$'s are fixed. An alternate procedure taking into account the random $\underset{\sim}{W}_j$ is given in the appendix.

Three possible estimators for C_j are now given. The first is a non parametric estimator associated with $\hat{\underset{\sim}{V}}_j$. Because $\hat{\underset{\sim}{V}}_j$ is expressed as a mean, an estimator of the covariance matrix for $\hat{\underset{\sim}{V}}_j$ is

$$\hat{\Sigma}_{j1} = (S_j-1)^{-2} \sum_{t=1}^{S_j} (\hat{\underset{\sim}{V}}_{t_j} = \hat{\underset{\sim}{V}}_j)(\hat{\underset{\sim}{V}}_{t_j} - \hat{\underset{\sim}{V}}_j)' \qquad (18)$$

where

$$\hat{\underset{\sim}{V}}_{t_j} = (\hat{\pi}^2{}_{1t}, \hat{\pi}_{1tj}\hat{\pi}_{2tj}, \ldots\ldots\ldots \hat{\pi}^2{}_{(I-1)tj}).$$

This estimator requires a few assumptions, but the number of clusters must exceed $2^{-1}(I-1)I$ if $\hat{\Sigma}_{j1}$ is to be nonsingular.

The second estimator of C_j uses the Dirichlet Multinomial model and assumes that the cluster sizes are large enough so that the normal distribution can be used to approximate the covariance matrix of the sample variances. We consider the transformed observations

$$\underset{\sim}{Y}_{tj} = n_j^{-1}(\underset{\sim}{X}_{tj} - \bar{\underset{\sim}{X}})R_j{}' \qquad (19)$$

$$= (\underset{\sim}{\pi}_{tj} - \bar{\underset{\sim}{\pi}}_j)R_j{}';$$

where

$$\bar{\underset{\sim}{\pi}}_j = \sum_{t=1}^{S_j} \underset{\sim}{\pi}_{t_j},$$

R_j is the matrix such that

$$R_j\, \Sigma_{mj}\, R_j{}' = \begin{bmatrix} \mathcal{I}_{I-1} & \underset{\sim}{0} \\ \underset{\sim}{0}' & 0 \end{bmatrix} \qquad (20)$$

and Σ_{mj} is the multinomial covariance matrix for a sample of size n_j. Under the model the covariance matrix of the first I-1 elements of $\underset{\sim}{Y}_{tj}$ is a multiple of the identity matrix. Let

$$(S_j - 1)^{-1} \sum_{t=1}^{S_j} \underset{\sim}{Y}'_{tj}\, \underset{\sim}{Y}_{tj} = \begin{bmatrix} \hat{V}_{jYY} & \underset{\sim}{0} \\ \underset{\sim}{0}' & 0 \end{bmatrix}. \qquad (21)$$

Then under the model

$$E(\hat{V}_{YY}) = V_{jYY} = c_j \mathcal{I} . \qquad (22)$$

If $\underset{\sim}{Y}_{tj}$ is normally distributed,

$$\mathrm{Var}\{\mathrm{vech}\ \hat{V}_{jYY}\} = (S_j-1)^{-1}\ C_j^{\ 2}\ D_B, \qquad (23)$$

where

$$D_B = \mathrm{diag}\ (2,\ 1,\ 1,\ \ldots\ldots,\ 2,\ 1,\ 2) \qquad (24)$$

and the element of the diagonal matrix D_B for an estimated variance is two and the element for an estimated covariance is one, Anderson (1958). Because Σ_m is unknown it is necessary to replace Σ_{mj} with Σ_{mj} where

$$\hat{R}_j\ \hat{\Sigma}_{mj}\ \hat{R}_j' = \mathcal{J}.$$

Then the estimated generalized least squares estimator C_j is

$$\hat{C}_{Bj} = (\underset{\sim}{A}'D_B^{\ -1}\underset{\sim}{A})^{-1}\underset{\sim}{A}'D_B^{\ -1}\underset{\sim}{H}_j, \qquad (25)$$

where

$$A = \mathrm{vech}\,\mathcal{J},$$

$$H_j = \mathrm{vech}(\hat{V}_{jYY}), \qquad (26)$$

which is identical to the estimator proposed by Brier (1980). Under the normal assumption, and with Σ_{mj} known, an estimator of the variance of $\hat{C}_{Bj}$ is

$$\hat{V}\ \{\hat{C}_{Bj}\} = (I-1)^{-1}\ \hat{C}^2_{B_j}\ (\underset{\sim}{A}'D_B^{\ -1}\underset{\sim}{A})^{-1}. \qquad (27)$$

A lack-of-fit statistic for the model is

$$\chi^2_{Bj} = (I-1)\hat{C}_{Bj}^{\ -2}\ [\underset{\sim}{H}_j'D_B^{\ -1}\underset{\sim}{H}_j - \underset{\sim}{H}_j'D_B^{\ -1}\underset{\sim}{A}\ \hat{C}_{Bj}]. \qquad (28)$$

If Σ_{mj} is known and the model is true, the large sample distribution of χ^2_{Bj} is approximately that of a chi-square random variable with $2^{-1}(I-1)I-1$ degrees of freedom. An alternative estimator of the variance $\hat{C}_{Bj}$ and alternative lack-of-fit statistics are considered in the example of section 4.

A third estimator of C_j falls between the previous two in the amount of model information used in the construction. Under the model the covariance matrix of vech $\hat{V}_{jYY}$ is a diagonal matrix. An estimator of the variance of the rs-th element is

$$D_{wjrs} = (s_j-1)^{-2}\sum_{t=1}^{S_j}(y_{jrt}y_{jst} - \hat{V}_{jYYrs})^2 \qquad (29)$$

where $\hat{V}_{jYYrs}$ is the rs-th element of $\hat{V}_{jYY}$ defined in (22). Then a generalized least squares estimator of C_j is

$$\hat{C}_{wj} = (\underset{\sim}{A}'D_{wj}^{\ -1}\underset{\sim}{A})^{-1}\underset{\sim}{A}'D_{wj}^{\ -1}\underset{\sim}{H}_j \qquad (30)$$

where

$$D_w = \mathrm{diag}(D_{W11},\ D_{W21},\ \ldots\ldots,\ D_{W,I-1,I-1}). \qquad (31)$$

The associated test statistic is

$$X^2_{wj} = \underset{\sim}{H}_j\ D_{Wj}^{\ -1}\underset{\sim}{H}_j - \underset{\sim}{H}_j\ D_{wj}^{\ -1}\underset{\sim}{A}\ \hat{C}_{wj}. \qquad (32)$$

If Σ_{mj} is known and the model is true, the large sample distribution of X^2_{wj} is that of a chi-square random variable with $2^{-1}(I-1)I-1$ degrees of freedom.

4. A NUMERICAL EXAMPLE

Data for this example were taken from Brier (1980). The data pertain to the manner in which the residents of Minnesota perceive the quality of their housing and their community housing. The variables of interest in this survey are the opinions of families about their homes and about their neighborhood. In each community, five homes were randomly selected and the families were questioned about two items: satisfaction with the housing in the neighborhood as a whole (unsatisfied, satisfied, very satisfied) and satisfaction with their own home. In this example, we examine only the data on the owners' satisfication with their own homes. The groups of five homes are the clusters. There are 18 clusters in the metropolitan area and 17 clusters in the non-metropolitan area. These are the clusters with complete responses. Those clusters with less than 5 homes were deleted from the data. There were 2 such clusters in the metropolitan area and 3 in the non-metropolitan area.

In this analysis, the interest is in the distribution of the responses for the two areas of satisfaction categories. The hypothesis is

$$H_o\colon \underset{\sim}{\pi}_j = \underset{\sim}{\pi}_o \qquad j = 1,\ 2;$$

where $\underset{\sim}{\pi}_o$ is an unknown probability vector of dimension 3 and $\underset{\sim}{\pi}_j$ is the probability vector of the j th subpopulation. The estimated vectors are

$$\hat{\underset{\sim}{\pi}}_1 = (.5222,.4222,.0556)'$$

and

$$\hat{\underset{\sim}{\pi}}_2 = (.3529.5059,.1412)'.$$

The estimator of the covariance matrix for the metropolitan area is Σ_{m1} and for the non metropolitian area is Σ_{m2}. The vector $\hat{\underset{\sim}{w}}_1$ formed from Σ_{m1} in (17) is

$$\hat{\underset{\sim}{w}}_1 = (499.0,-441.0,487.9) \times 10^{-4}.$$

It is the right side of the regression equation

$$\hat{\underset{\sim}{V}}_1 = C_1\ \hat{\underset{\sim}{w}}_1 + \underset{\sim}{\varepsilon}_1,$$

where $\hat{\underset{\sim}{V}}_1$ is formed from the estimated covariance matrix constructed using the cluster variance formula,

$$\Sigma_{DMI} = 17^{-1}\sum_{k=1}^{18}(\underset{\sim}{\pi}_{1k} - \hat{\underset{\sim}{\pi}}_1)(\underset{\sim}{\pi}_{1k} - \hat{\underset{\sim}{\pi}}_1)'.$$

The estimated vector $\hat{\underset{\sim}{V}}_1$ is

$$\hat{\underset{\sim}{V}}_1 = (1041.83,-899.35,888.89) \times 10^{-4},$$

and the Cholesky decomposition of Σ^{-1}_{m1} is given by $\Sigma^{-1}_{m1} = \hat{R}_1\hat{R}_1'$, where

$$\hat{R}_1 = \begin{bmatrix} 9.9787 & 9.0192 \\ 0 & 4.5273 \end{bmatrix}.$$

The generalized least squares estimator defined in (25) is

$$\hat{C}_{B1} = 1.61918$$

and an estimated variance, of this estimator is

0.14548. The estimated lack-of-fit statistic using (28) is X^2_{B1} - .2288.

The estimated generalized least squares estimator defined in (30) is

$$\hat{C}_{w1} = 1.662$$

with an estimated variance of $\hat{C}_{w1}$ = .1716 using (31). The associated lack-of-fit statistic from (32) is X^2_{w1} = 3.1194. The values of the lack-of-fit statistics indicate that the model is a good approximation. Because Σ_{m1} is estimated for the transformation, the lack-of-fit statistic is biased.

Similar results were obtained for the metropolitan area. $\hat{C}_{B2} = 1.632$, $\hat{V}\{\hat{C}_{B2}\} = .0520$, $\hat{C}_{w2} = 1.563$ and $\hat{V}\{C_{w2}\} = .0774$. The associated lack-of-fit statistics are $X^2_{B2} = 5.24$ and $X^2_{w2} = 6.35$. On the basis of these lack-of-fit statistics the model is not rejected at the one percent level. Since $\hat{\Sigma}_j$ is non singular, another method of estimating C_j based on (18) is

$$\hat{C}_{jwls} = (\underset{\sim}{w}_j{}' \hat{\Sigma}_j^{-1} \hat{\underset{\sim}{w}}_j)^{-1} \hat{\underset{\sim}{w}}_j{}' \hat{\Sigma}_j^{-1} \hat{\underset{\sim}{V}}_j .$$

Then for the non metropolitan data the estimator is

$$C_{1wls} = 1.96148$$

and for the metropolitan data the estimator is

$$C_{2wls} = 1.19843.$$

A test-of-fit for the model, X^2_{jwls} which is approximately distributed as a chi square random variable with two degrees of freedom is equal to .0788 for the non metropolitan data and .7452 for the metropolitan data.

A summary of the C estimators for the two areas and the value of the statistic X^2_{DMI} in (9) is given in the following table.

TABLE 4.1 A SUMMARY OF STATISTICS

C Estimators	Metro	Non Metro	X^2_{DMI}	p-value
$\hat{C}_B$	1.6192	1.6320	4.1881	.10<p<.20
$\hat{C}_w$	1.6617	1.5634	4.2079	.10<p<.20
$\hat{C}_{wls}$	1.9615	1.1984	5.373	.05<p<.10
$\hat{C}_p$	1.0	1.0	6.8077	.02<p<.05

The value of X^2_{DMI} when the C estimators are one is equivalent to the usual Pearson Chi-square test for independence.

Wilson (1984) constructed tests of the form considered by Wald (1943) without any assumptions on the covariance matrix. One such statistic was obtained for a two-stage sampling scheme in J different strata, j = 1, 2, ..., J. Such a scheme is referred to as a stratified two-stage sampling scheme. Consider testing $H_o: \underset{\sim}{\pi}_j = \underset{\sim}{\pi}_o$ for some unknown vector $\underset{\sim}{\pi}_o$ and j = 1, 2, ..., J. Define

$$\hat{\underset{\sim}{\pi}}_o = \sum_{j=1}^{J} \alpha_j \hat{\underset{\sim}{\pi}}_j$$

as an unbaised estimator for

$$\underset{\sim}{\pi}_o = \sum_{j=1}^{J} \alpha_j \underset{\sim}{\pi}_j .$$

The variance-covariance matrix for $\underset{\sim}{\pi}_j - \underset{\sim}{\pi}_o$ is

$$M_{jj} = \text{var}(\hat{\underset{\sim}{\pi}}_j - \hat{\underset{\sim}{\pi}}_o)$$
$$= R_j - 2\alpha_j^{-1} R_j + \sum_{\ell=1}^{J} \alpha_i^{-2} R_{\hat{\ell}}$$

and the matrix covariances between $(\underset{\sim}{\pi}_j - \hat{\underset{\sim}{\pi}}_o)$ and $(\hat{\underset{\sim}{\pi}}_j - \hat{\underset{\sim}{\pi}}_o)$, is

$$M_{jj'} = \text{cov}(\hat{\underset{\sim}{\pi}}_j - \hat{\underset{\sim}{\pi}}_o, \hat{\underset{\sim}{\pi}}_{j'} - \hat{\underset{\sim}{\pi}}_o)$$
$$= -\alpha_j^{-1} R_j - \alpha_{j'}^{-1} R_{j'} + \sum_{\ell=1}^{J} \alpha_\ell^{-2} R_j ,$$

where

$$R_j = \Sigma_m(j) + \frac{\Sigma n_j^2}{N_j^2} - \frac{1}{N_j} \sum_{\ell=1}^{n_j} \alpha_{j\ell} (\underset{\sim}{p}_{j\ell} - \underset{\sim}{\pi}_j)(\underset{\sim}{p}_{j\ell} - \underset{\sim}{\pi}_j)' ,$$

$\underset{\sim}{p}_{j_\ell}$ is a vector of proportion for the ℓth cluster in the jth subpopulation and

$$\Sigma_m(j) = N_j^{-1}(\Delta_{\underset{\sim}{\pi}_j} - \underset{\sim}{\pi}_j \underset{\sim}{\pi}_j{}').$$

Let Σ be the covariance matrix for

$$(\hat{\underset{\sim}{\pi}} - \hat{\underset{\sim}{\pi}}_o) = (\hat{\underset{\sim}{\pi}}_1 - \hat{\underset{\sim}{\pi}}_{o1}\ \hat{\underset{\sim}{\pi}}_2 - \hat{\underset{\sim}{\pi}}_{o1} \ldots\ldots, \hat{\underset{\sim}{\pi}}_J - \hat{\underset{\sim}{\pi}}_o)$$

then, Σ has $M_{jj'}$ as off diagonal blocks and M_{jj}, as diagonal blocks.

Let Σ_o denote the value of Σ under H_o and R_{o_j} the value of R_j under H_o. Consider J = 2, then under H_o

$$\Sigma_o = \begin{bmatrix} (1-2\alpha_1)R_1 + \sum_{\alpha=1}^{2} \alpha_\ell^2 R_\ell & -\alpha_1 R_1 - \alpha_2 R_2 + \sum_{\ell=1}^{2} \alpha_\ell^2 R_\ell \\ -\alpha_1 R_1 - \alpha_2 R_2 + \sum_{\ell=1}^{2} \alpha_\ell^2 R_\ell & (1-2\alpha_2)R_2 + \sum_{\ell=1}^{2} \alpha_\ell^2 R_\ell \end{bmatrix}$$

where

$$R_j = N_j^{-1}(\Delta_{\underset{\sim}{\pi}_o} - \underset{\sim}{\pi}_o \underset{\sim}{\pi}_o^1) + \left(\frac{\Sigma n_j^2}{N_j^2} - \frac{1}{N_j} \right) \sum_{\ell=1}^{S} \alpha_{j\ell} \{\underset{\sim}{p}_{j\ell} - \underset{\sim}{\pi}_o)(\underset{\sim}{p}_{j\ell} - \underset{\sim}{\pi}_o)'\}$$

$$= N_j^{-1}\{\Delta_{\underset{\sim}{\pi}_o} - \frac{\Sigma n_j^2}{N_j} \underset{\sim}{\pi}_o \underset{\sim}{\pi}_o{}' + (\frac{\Sigma n_j^2}{N_j} - 1) \Sigma \alpha_{j\ell}\, \underset{\sim}{p}_{j\ell}\, \underset{\sim}{p}'_{j\ell}\}$$

When the sample sizes for each cluster within a stratum, n_j, are equal, then

$$R_j = N_j^{-1}[\Delta_{\underset{\sim}{\pi}_o} - \frac{N_j}{S_j} \underset{\sim}{\pi}_o \underset{\sim}{\pi}_o{}' + \frac{1}{S_j} (\frac{N_j}{S_j} - 1) \sum_{\alpha=1}^{S_j} \underset{\sim}{p}_{j\ell}\, \underset{\sim}{p}'_{j\ell}]$$

$$= N_j^{-1}[\Delta_{\underset{\sim}{\pi}_o} - \underset{\sim}{\pi}_o \underset{\sim}{\pi}_o{}'] + \frac{1}{S_j} (\frac{1}{S_j} - \frac{1}{N_j}) \sum_{\alpha=1}^{S_j} \Bigg[$$

$$\{(\underset{\sim}{p}_{j\ell} - \underset{\sim}{\pi}_o)(\underset{\sim}{p}_{j\ell} - \underset{\sim}{\pi}_o)'\}\Big]$$

A test statistic for H_o: $\underset{\sim}{\pi}_j = \underset{\sim}{\pi}_o$ where $\underset{\sim}{\pi}_o$ is an unknown vector is

$$X^2_{WI2S} = (\hat{\underset{\sim}{\pi}} - \hat{\underset{\sim}{\pi}}_o)'\hat{\Sigma}_o^-(\hat{\underset{\sim}{\pi}} - \hat{\underset{\sim}{\pi}}_o)$$

where $\hat{\Sigma}_o$ is a consistent estimator for Σ_o. One such estimator can be obtained by using $\hat{\underset{\sim}{p}}_{j\ell}$ to estimate $\underset{\sim}{p}_{j\ell}$ and $\hat{\underset{\sim}{\pi}}_o$ to estimate $\underset{\sim}{\pi}_o$. For the example considered here the vector of proportion

$$\hat{\underset{\sim}{\pi}}_o = (0.4400, 0.4629, .09714)',$$

and the estimated variance-covariance matrix for $\hat{\underset{\sim}{\pi}}_1 - \hat{\underset{\sim}{\pi}}_o$ is

$$\mathrm{COV}(\hat{\underset{\sim}{\pi}}_1 - \hat{\underset{\sim}{\pi}}_o) = \begin{bmatrix} 71.45 & -62.25 & -9.20 \\ -62.25 & 64.42 & -2.17 \\ -9.20 & -2.17 & 11.37 \end{bmatrix} \mathrm{X}\ 10^{-4}$$

The estimated variance - covariance matrix for $\hat{\underset{\sim}{\pi}}_2 - \hat{\underset{\sim}{\pi}}_o$ is

$$\mathrm{Cov}(\hat{\underset{\sim}{\pi}}_2 - \hat{\underset{\sim}{\pi}}_o) = \begin{bmatrix} 62.63 & -49.97 & -12.65 \\ -49.97 & 59.50 & -9.53 \\ -12.65 & -9.53 & 22.18 \end{bmatrix} \mathrm{X}\ 10^{-4}$$

and

$$\hat{\Sigma}_o = \begin{bmatrix} 31.63 & -26.48 & -51.56 \\ -26.48 & 29.23 & -2.76 \\ -51.56 & -2.76 & 7.91 \end{bmatrix} \mathrm{X}\ 10^{-4}.$$

The test statistic X^2_{WI2S} has the value 3.2601 with an observed significance level of 0.804078. This test statistic is less than half the value of the usual Pearson statistic of 6.8077. X^2_{WI2S} is less than any of the test statistic values obtained using $\hat{C}_B$, $\hat{C}_w$ or $\hat{C}_{wls}$.

5. DISCUSSION

Other data sets were considered by Wilson (1984). The Dirichlet Multinomial provided a good description of the data, the C_B estimator, which requires the estimation of the fewest parameters, was the most stable. The other estimators for $\hat{C}$ had a non-negligible chance of being outside the range of 1 to n_j when the number of clusters sampled was small. The sample sizes within the clusters had relatively little effect on the estimation of C. For large numbers of sampled clusters, the Wald statistic considered in Section 4 should be quite similar to the test based on the Dirichlet Multinomial model using any of the estimators for C, when the later model is correct. In this situation, the Dirichlet Multinomial model may provide a small increase in power for detecting many alternatives because of the simple form of the covariance matrix. For smaller numbers of clusters, the Dirichlet Multinomial model will provide a more reliable test for the same reason, there is much less variation in the estimated covariance matrix. Of course, the Wald statistic in Section 4 can be applied to a wider class of models. Using the Dirichlet Multinomial model requires additional work to check if the model assumptions are well satisfied.

REFERENCES

Altham, P.M.E. 1976. Discrete variables analysis for individuals grouped into families. Biometrika 63: 263-269.

Anderson, T.W. 1958. An introduction to multivariate statistical analysis. John Wiley and Sons, New York.

Birch, M.W. 1964. A new proof of the Pearson-Fisher Theorem. Ann. of Math. Statist. 35: 817-824.

Brier, S.S. 1980. Analysis of contingency tables under cluster sampling. Biometrika 67: 591-596.

Cohen, J.E. 1976. The Distribution of the chi-square statistic under clustered sampling from contingency tables. J. Amer. Statist. Assoc. 71: 665-670.

Cramer, H. 1946. Mathematical methods of statistics. Princeton University Press, Princeton.

Good, I.J. 1965. Estimation of probabilities. MIT Press, Cambridge, MA.

Graybill, F.A. 1969. Introduction to matrices with applications in statistics. Wadsworth, Belmont California.

Johnston, N.L. and Kotz, S. 1970. Continuous univariate distributions. Houghton Mifflin Co., Boston.

Moore, D.S. 1977. Generalized inverse, Wald's Method, and the construction of chi-squared tests of Fit. J. Amer. Statist. Assoc., 72: 131-137.

Mosemann, J.E. 1962. On the compound multinomial distribution, the multivariate β-distribution, and correlation among proportion. Biometrika 49: 65-82.

Rao, J.N.K. and Scott, A.J. 1981. The analysis of categorical data from complex surveys. J. Amer. Statist. Assoc. 76: 221-230.

Wald, A. 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large.

Wilson, J.R. 1984. Statistical methods for frequency data from complex sampling schemes. Ph.D. Dissertation, Iowa State University, Ames, Iowa 50011.