Statistical methods for frequency data from complex sampling schemes

Jeffrey Rupert Wilson
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd
Part of the Statistics and Probability Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/7807

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
Wilson, Jeffrey Rupert

STATISTICAL METHODS FOR FREQUENCY DATA FROM COMPLEX SAMPLING SCHEMES

Iowa State University Ph.D. 1984

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106
Statistical methods for frequency data
from complex sampling schemes

by

Jeffrey Rupert Wilson

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work
Signature was redacted for privacy.

For the Major Department
Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1984
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. METHODS OF TESTING HYPOTHESES WHEN THE NUMBER OF</td>
<td>5</td>
</tr>
<tr>
<td>OBSERVATIONS IS LARGE</td>
<td></td>
</tr>
<tr>
<td>2.1. Wald Statistic</td>
<td>5</td>
</tr>
<tr>
<td>2.2. Likelihood Ratio Statistic</td>
<td>12</td>
</tr>
<tr>
<td>2.3. Pearson-Fisher Theorem</td>
<td>13</td>
</tr>
<tr>
<td>2.4. Multinomial Sampling</td>
<td>15</td>
</tr>
<tr>
<td>2.5. Stratified Sampling</td>
<td>19</td>
</tr>
<tr>
<td>2.6. Cluster Sampling</td>
<td>21</td>
</tr>
<tr>
<td>2.7. Cluster Models</td>
<td>23</td>
</tr>
<tr>
<td>2.8. Other Models</td>
<td>26</td>
</tr>
<tr>
<td>3. WALD STATISTICS</td>
<td>31</td>
</tr>
<tr>
<td>3.1. Introduction</td>
<td>31</td>
</tr>
<tr>
<td>3.2. Simple Random Sampling</td>
<td>38</td>
</tr>
<tr>
<td>3.3. Stratified Sampling</td>
<td>43</td>
</tr>
<tr>
<td>3.4. Two-Stage Sampling</td>
<td>65</td>
</tr>
<tr>
<td>4. DIRICHLET-MULTINOMIAL MODEL FOR CLUSTER SAMPLING</td>
<td>87</td>
</tr>
<tr>
<td>4.1. Introduction</td>
<td>87</td>
</tr>
<tr>
<td>4.2. Dirichlet-Multinomial Model</td>
<td>87</td>
</tr>
<tr>
<td>4.3. A Test of Proportions</td>
<td>108</td>
</tr>
<tr>
<td>4.4. Some Examples</td>
<td>125</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>-----------------</td>
<td>------</td>
</tr>
<tr>
<td>5. REFERENCES</td>
<td>227</td>
</tr>
<tr>
<td>6. ACKNOWLEDGMENTS</td>
<td>232</td>
</tr>
<tr>
<td>7. APPENDIX A</td>
<td>234</td>
</tr>
<tr>
<td>8. APPENDIX B</td>
<td>243</td>
</tr>
</tbody>
</table>
1

1. INTRODUCTION

1.1. Introduction

During the 1950s, several review papers on the practical use of \( \chi^2 \) goodness-of-fit statistics appeared which provided advice on the accuracy of the large sample chi-squared approximations for these statistics, e.g., Cochran (1952, 1977) and Watson (1957). Other papers dealt directly with the theory and gave an asymptotic distribution of \( \chi^2 \) statistics computed in a variety of different ways (e.g., Watson (1959)). In the intervening years, considerable attention has been focused on the development of methods for the analysis of categorical data primarily through the use of log-linear and logit models (Bishop et al. (1975), Plackett (1974), and Haberman (1974)). The expanding interest in this topic has kindled further efforts on both the theory for the large sample \( \chi^2 \) test and the use of these statistics in practice.

It is usually unrealistic to regard observations from a complex survey as independent, but practitioners often use standard chi-squared tests with survey data. Holt et al. (1980) examined results from two (2) national surveys which suggest that the effects of complex sampling schemes are severe for tests of goodness-of-fit or homogeneity, but less severe for tests of independence.

Methods for analyzing categorical data have been developed extensively assuming multinomial sampling. In particular, there are \( \chi^2 \) tests for problems involving goodness-of-fit and tests of independence and homogeneity in two-dimensional contingency tables. However,
in practice, many studies employ sampling methods that are more complex than the method of simple random sampling. Most of the commonly used survey designs employ stratification or cluster sampling or both, and hence, do not satisfy the assumption of multinomial sampling. The cost and operational facilities often dictate the use of a complex sampling scheme and so it becomes important to take this into consideration when obtaining a statistic. This has been done for non-linear statistics like regression coefficients, correlations, and principal components in multivariate analysis by Kish and Frankel (1974) and Fuller (1975), among others.

Stratification and cluster sampling share the common feature of having the population of interest partitioned into a set of primary sampling units (psu's). The difference between the two (2) schemes is that for the stratified sampling scheme the population consists of a finite number of psu's and each psu is sampled. In cluster sampling, there is typically a large number of psu's and only a portion of these are included in the sample. In stratified sampling, the psu's are called strata. In cluster sampling, they are called clusters or families.

For stratified sampling, the population consists of a finite number of psu's \( S_1, S_2, \ldots, S_J \). From each of these strata, simple random samples are taken of size \( n_1, n_2, \ldots, n_J \), respectively. The relative sizes of the strata, \( W_1, W_2, \ldots, W_J \), are known. What makes this a stratified sample is the fact that the \( J \) groups comprise the entire population and a sample is taken from each of the strata in the population.
In cluster sampling, only a subset of all psu's in the population is examined. The clusters are selected via a random mechanism. It is this fact that only part of the population was sampled that makes it a cluster sample. Sometimes, there are actually a finite number of clusters but at times it may be assumed that the number is infinite when developing a mathematical model.

Thus, in stratified samples, the desire is to make inferences only about the observed psu's while in cluster samples the wish is to make inferences about the population from which the psu's were drawn.

In this dissertation, statistics are developed to test some of the more common hypotheses when the sampling scheme is complex. The structure of this dissertation is as follows. In Chapter 2, the literature review contains the basic ideas for constructing a Wald Statistic. The properties of such a statistic are noted. The likelihood ratio test statistic is also discussed and its asymptotic distribution is shown to be the same as the asymptotic distribution of the Wald Statistic. The Pearson-Fisher Theorem is discussed in Section 2.3. The approach to complex sampling schemes for the goodness-of-fit problems as examined by Rao and Scott (1981, 1984) is reviewed in Sections 2.4-2.6. In Section 2.7, models for cluster sampling as presented by Brier (1980), Cohen (1976), Altham (1976), and Rao and Scott (1981) are examined.

Chapter 3 deals mainly with the construction of Wald Statistics for several complex sampling schemes. Bounds are obtained for such statistics by the use of matrix theory. Some approximate statistics for these Wald Statistics are also obtained. Stratified and cluster
Sampling schemes are discussed in Sections 3.3 and 3.4.

Chapter 4 uses the Dirichlet-Multinomial as a model for the clustered sampling scheme. Primary units are randomly sampled from each subpopulation and simple random sampling with replacement is used to obtain a vector of frequencies from each of the primary units selected. The true vectors of category proportions may not be the same for all primary units in the same subpopulation. The Dirichlet-Multinomial distribution provides a model for this variation in constructing chi-squared tests. Several examples are analyzed in Section 4.4 using the Dirichlet-Multinomial model.
2. METHODS OF TESTING HYPOTHESES WHEN THE NUMBER OF OBSERVATIONS IS LARGE

2.1. Wald Statistic

Testing hypotheses when the number of observations is large has been the concern of many statisticians for decades. Most of the methods reviewed in this chapter are based on the method developed by Abraham Wald (1943).

Suppose \( X = (x_1, x_2, \ldots, x_n) \) is a random vector of dimension \( r \) involving \( k \) unknown parameters \( \theta = (\theta_1, \theta_2, \ldots, \theta_k) \). Let \( f(x_1, x_2, \ldots, x_r; \theta_1, \theta_2, \ldots, \theta_k) \) be the joint density function for the \( r \) variates. The vector \( \theta \) can be considered as a point in a \( k \)-dimensional Euclidean space. Let \( \Omega \) denote the parameter space and let \( \Omega_\theta \) denote a subset of \( \Omega \). Let \( \theta_0 \) denote a point in \( \Omega_\theta \), and let \( H_w \) denote a simple hypothesis when \( \Omega_\theta \) consists of a single point, otherwise \( H_w \) denotes a composite hypothesis.

Define the maximum likelihood estimates \( \hat{\theta}_n = (\hat{\theta}_1, \ldots, \hat{\theta}_n) \) for the values of \( \theta = (\theta_1, \theta_2, \ldots, \theta_k) \) as the parameter values for which \( \prod_{\alpha=1}^{n} f(x_\alpha, \theta) \) is maximized. Let \( x_n \) denote a sample point in the \( rn \)-dimensional space of \( n \) independent observations on the random vector \( X \). A region \( W_n \) in the \( rn \)-dimensional space is called a critical region for testing \( H_w \) if \( H_w \) is rejected when and only when the observed sample point falls within \( W_n \).

For \( \theta \in \Omega_\theta \) the value of \( P(W_n | \theta) \) is called the power of the critical region \( W_n \) with respect to the alternative hypothesis. Then,
sup \( P(W_n | \theta) \) is the size of the critical region \( W_n \).

Wald discusses the question of an appropriate test for the hypothesis \( H_w \) based on a large number of independent observations on \( \mathcal{X} = (x_1, x_2, \ldots, x_r) \). He makes the following assumptions on the density \( f(\mathcal{X}, \theta) \).

A1. Let \( D_n \) be the set of all sample points for which the maximum likelihood estimate \( \hat{\theta}_n \) exists and the second order partial derivatives are continuous functions of \( \theta \). It is assumed that \( \lim_{n \to \infty} P(D_n | \theta) = 1 \) uniformly in \( \theta \).

A2. The maximum likelihood estimate \( \hat{\theta}_n \) is a uniformly consistent estimate of \( \theta \).

A3. Let \( A = (a_{ij}) \) where \( a_{ij} = -\frac{\partial^2 \log f(\mathcal{X}, \theta)}{\partial \theta_i \partial \theta_j} \) and let \( |A| \) denote the determinant of \( A \). It is assumed that the matrix is positive definite.

A4. It is possible to differentiate under the integral sign in the expression \( \int_{-\infty}^{\infty} f(\mathcal{X}, \theta) \, d\mathcal{X} = 1 \).

A5. There exists a constant \( \eta > 0 \) such that

\[
\mathbb{E}_\theta \left| \frac{\partial}{\partial \theta_i} \log f(\mathcal{X}, \theta) \right|^{2+\eta}
\]

is a bounded function of \( \theta \) for \( i = 1, 2, \ldots, k \).

Using these five (5) assumptions, Wald proved the following theorem.
Theorem 2.1.

As $n \to \infty$ converges uniformly in $\Theta$ to the cumulative multivariate normal distribution with zero mean vector and covariance matrix $\Sigma = A^{-1}$.

He also proved two (2) important lemmas which reduce the general problem of large sample inference to the case where the variates under consideration have a joint normal distribution (Wald (1943, pp. 433-45)).

Definition 2.1.

A critical region $W_n$ is said to have uniformly best average power with respect to the surface $K_c$ and the weight function $w(\theta)$ if for any region $Z_n$ of size equal to that of $W_n$

$$\int_{K_c} P(W_n | \theta) w(\theta) d\mu \geq \int_{K_c} P(Z_n | \theta) w(\theta) d\mu.$$

Definition 2.2.

A critical region $W_n$ for testing $H: \theta = \Theta_0$ is said to have uniformly best constant power on the family of surfaces $\{K_c\}$ if the following two conditions are satisfied: a) $P(W_n | \Theta_1) = P(W_n | \Theta_2)$ for any $(\Theta_1, \Theta_2)$ which lie on the same surface $K_c$, and b) $P(W_n | \Theta) \geq P(Z_n | \Theta)$ for any $Z_n$ which satisfies condition a) and for which $P(Z_n | \Theta_0) = P(W_n | \Theta_0)$.

Definition 2.3.

A critical region $W_n$ is said to be a most stringent test of $H: \theta = \Theta_0$ at the level of significance $\alpha$ if $P(W_n | \Theta_0) = \alpha$ and the
difference

\[ \sup_{\theta} \left[ P(W|\theta) - P(W_n|\theta) \right] \leq \sup_{\theta} \left[ P(W|\theta) - P(Z_n|\theta) \right] \]

for all regions \( Z_n \) for which \( P(Z_n|\theta_0) = \alpha \). The test minimizes the difference between the maximum power attainable and the actual power.

Theorem 2.2.

If \( \chi = (x_1, x_2, \ldots, x_r) \) has a joint normal distribution with unknown mean vector \( \theta_0 = (\theta_1, \theta_2, \ldots, \theta_r) \) and a known covariance matrix \( \Sigma \), then for testing \( H: \theta = \theta_0 \) on the basis of a single observation on the vector \( \chi \), the critical region given by the inequality

\[ (\chi - \theta_0)'A(\chi - \theta_0) = \Sigma \sum_{i,j} a_{ij}(x_i - \theta_{io})(x_j - \theta_{jo}) \geq d \]

where \( \Sigma = A^{-1} \) and \( A = (a_{ij}) \);

i) has uniformly best average power with respect to surfaces defined by

\[ (\theta - \theta_0)'A(\theta - \theta_0) = \Sigma \sum_{i,j} a_{ij}(\theta_i - \theta_{io})(\theta_j - \theta_{jo}) = c \]

and a measure which distributes probability uniformly on a sphere

ii) has uniformly best constant power on the surface

\[ (\theta - \theta_0)'A(\theta - \theta_0) = c \]

iii) is a most stringent test.

Theorem 2.3.

Let \( W_n^* \) be a critical region for testing \( H: \theta = \theta_0 \) defined by
the inequality

\[ n(\hat{\theta}_n - \theta_0)'A(\hat{\theta}_n - \theta_0) \geq d_n \]

where the real number \( d_n \) is chosen so that \( P(W_n^* | \theta_0) = \alpha \). Then, the test \( W_n^* \)

i) has asymptotically best average power with respect to the function \( (\theta - \hat{\theta}_0)'A(\theta - \hat{\theta}_0) = c \),

ii) has asymptotically best constant power on \( (\theta - \hat{\theta}_0)'A(\theta - \hat{\theta}_0) = c \),

iii) is an asymptotically most stringent test.

These results can be extended to the composite hypotheses by assuming, in addition to A1-A5, the assumption

A6. \( H: \xi_1(\theta) = \xi_2(\theta) = \ldots = \xi_r(\theta) = 0 \)

for \( r < k \), and there exist \( k - r \) functions \( \xi_{r+1}(\theta), \ldots, \xi_k(\theta) \) such that the following three (3) conditions are fulfilled:

a) The transformation \( \xi: \theta \rightarrow \xi(\theta) \) is a topological transformation of \( \Theta \) onto itself.

b) The first and second derivative of \( \xi_i(\theta) \) \( i = 1, 2, \ldots, k \) are uniformly continuous and bounded functions of \( \theta \).

c) The matrix formed with entries given by \( \frac{\partial(\xi_1, \xi_2, \ldots, \xi_k)}{\partial(\theta_1, \theta_2, \ldots, \theta_k)} \) is positive definite.

Theorem 2.4.

Let \( W_n^{**} \) be the region defined by the statistic

\[ \chi^2_n = n(\hat{\theta}_n)'B^{-1}(\hat{\theta}_n), \]

where \( B \) is the \( k \times k \) matrix, such that,
B = DED' and D = \left(\frac{\partial \xi}{\partial \theta}\right)_1. Under A6 for testing H: \xi_1(\theta) = 
\xi_2(\theta) = \ldots = \xi_k(\theta) = 0 the test statistic \chi^2_w has

i) asymptotically best average with respect to \((\hat{\theta} - \theta_0)'A(\hat{\theta} - \theta_0) = c,\)

ii) has asymptotically best constant power on the surfaces defined by

\((\hat{\theta} - \theta_0)'A(\hat{\theta} - \theta_0) = c,\) and

iii) is an asymptotically most stringent test.

So far use has been made of i) maximum likelihood estimates for
the unknown parameters, and ii) the assumption that the covariance matrix
is known. However, Stroud (1971) presented an extension of Wald's
asymptotic procedure with some weaker conditions.

Recall that assumption A6 required i) the function \xi to be
uniformly continuous with bounded first and second order partial
derivatives, and ii) required uniform consistency of the maximum
likelihood estimators. Stroud replaces these global conditions with
some local conditions. The use of maximum likelihood estimators of
the parameter is not essential. In fact, Neyman (1949) has shown
that maximum likelihood estimators belong to a class of estimators
called Best Asymptotic Normal estimators (BAN). Neyman shows that the
maximum likelihood estimator is consistent, asymptotically normally
distributed, and has an asymptotic variance which does not exceed the
asymptotic variance of any other consistent estimator with an asym­
totically normal distribution.

Stroud (1971) shows that it is sufficient that

i) the estimators defining the distributions be asymptotically
normal under the sequence of local alternatives,
ii) the estimator of the covariance matrix need only converge stochastically under this sequence to the covariance for the limiting distribution, and

iii) the transformation, $\xi$ is assumed to possess continuous and bounded second partial derivatives locally within a neighborhood of any parameter point.

Stroud's work can be summarized by the following theorem:

Theorem 2.5.

Let \( \{q_{jn}\} \) be a sequence in k-dimensional Euclidean space \( \mathbb{R}^k \) of the form \( q_{jn} = q_{\infty} + n^{-1/2} \delta_n \) where \( \lim_{n \to \infty} \delta_n = 0 \) and \( q_{\infty} \) and \( \delta \) are fixed points. Let \( \{t_n\} \) be a sequence of k-dimensional random vectors, such that, \( n^{1/2}(t_n - q_n) \xrightarrow{D} N(0, \Sigma) \) where \( \Sigma \) is nonsingular. Let \( \{S_n\} \) be a sequence of k x k symmetric random matrices, nonsingular such that \( \lim_{n \to \infty} S_n \xrightarrow{P} \Sigma \) and suppose \( \xi: \mathbb{R}^k \to \mathbb{R}^r \) ( \( r \leq k \) ) is a function, such that, \( \xi(q_{\infty}) = 0 \). Suppose \( \xi \) is bounded and has continuous partial derivatives in a sphere of radius \( \rho \) about \( q_{\infty} \) such that the matrix of partial derivatives \( G = \left( \frac{\partial \xi}{\partial q} \right) \) when evaluated at \( q_{\infty} \) has rank \( r \).

Define the statistic

\[
\chi^2_{\text{wn}} = n[\xi(t_n)]'(G_n S_n G_n')^{-1}[\xi(t_n)]
\]

where \( G_n \) is the value of \( G \) evaluated at \( t_n \). Then as \( n \to \infty \), \( \chi^2_{\text{wn}} \) converges to a chi-square distribution with \( r \) degrees of freedom.

Note in the discussion above it is assumed that \( \Sigma \) is a non-singular matrix; however, in practice \( \Sigma \) may be singular. Moore (1977) showed that Wald's method generalizes to sequences of estimators having
singular covariance matrices by a natural use of generalized inverses. Though this technique may be useful, throughout this dissertation such cases will be handled by reparameterization.

2.2. Likelihood Ratio Statistic

Neyman and Pearson introduced a certain test criteria for purposes of statistical inference called the likelihood ratio. Suppose the density function in the sample space is given by \[ \prod_{\alpha=1}^{n} f(x_{\alpha}, \theta). \] Denote the maximum of this function with respect to \( \theta \) by \( L(x_{n}) \) and let \( L_{w}(x_{n}) \) denote the maximum restricted to \( \theta \) being in the space \( W \).

The likelihood ratio \( \lambda_{n} \) is defined as \( \lambda_{n}(\theta, x_{n}) = \frac{L_{w}(x_{n})}{L(x_{n})} \) where \( 0 < \lambda_{n} < 1 \) and \( x_{n} \) denotes a sample point in the n-dimensional space. The hypothesis \( H_{w} \) is rejected if the value of \( \lambda_{n}(\theta, x_{n}) < \lambda_{n} \) where \( \lambda_{n} \) is a suitably chosen constant and the \( \sup_{\theta} P(\lambda_{n}(\theta, x_{n}) < \lambda_{n}(\theta)|\theta) = \alpha \), the significance level. By use of Taylor's expansion, Wald (1943) showed that

\[ -2 \log \lambda_{n}(\theta) < c \]

where \( c \) is a finite value for all \( n \) and all \( \theta \). He made the following assumption:

A7. The likelihood ratio test is uniformly consistent.

Using this assumption, Wald proved the following theorem:

Theorem 2.6.

For testing \( H: \xi_{1}(\theta) = \xi_{2}(\theta) = \ldots = \xi_{I}(\theta) = 0 \) under assumptions A6 and A7, the likelihood ratio test has
i) asymptotically best average power with respect to the surfaces \( S_c(\theta) \) and weight function \( \eta(\theta) \),

ii) asymptotically best constant power on the surfaces \( S_c(\theta) \), and

iii) is an asymptotically most stringent.

Wilks (1938) derived the distribution of the likelihood ratio for large samples under assumption A6 and showed that it has a limiting central \( \chi^2 \)-square distribution under \( H_0 \) and a limiting noncentral \( \chi^2 \)-square distribution under \( H_a \).

Wald (1943) proved the following theorem which relates the two statistics \( X^2_w \) and \( \lambda_n \).

Theorem 2.7.

For \( H_i: \xi_1(\theta) = \xi_2(\theta) = \ldots = \xi_{I^2}(\theta) \), let

\[
X^2_{wn} = n[\xi(t_n)]' (G_n S_n G_n')^{-1}[\xi(t_n)]
\]

where \( G_n \) is the matrix of partial derivatives evaluated at \( t_n \), then under assumption A6

\[
\lim_{n \to \infty} [P(-2 \log \lambda_n(\theta, x_n^2) < c | \theta) - F(X^2_{wn}(\theta), c)] = 0
\]

uniformly in \( \theta \) and \( c \). \( F(X^2_{wn}(\theta), c) \) denotes the cumulative distribution function for \( X^2_{wn}(\theta) \). The limiting distribution of \(-2 \log \lambda_n(\theta, x_n)\) is the central \( \chi^2 \) (\( \chi^2_{I^2} \)) distribution with \( I \) degrees of freedom under \( H_0 \) and a noncentral \( \chi^2 \)-square under \( H_a \).

2.3. Pearson-Fisher Theorem

Fisher (1932) formulated the theorem that the \( X^2 \) goodness of fit statistic for a multinomial distribution with \( I \) cells and with \( k \) parameters fitted by the method of maximum likelihood is distributed as a central \( \chi^2 \) with \( I-k-l \) degrees of freedom. Karl Pearson (1900) estab-
lished the result for the special case $k = 0$. A somewhat rigorous proof was given by Cramer (1946). One basic assumption of Cramer's proof is that the maximum likelihood estimator is consistent. He proves the theorem for the subclass of maximum likelihood estimators that are consistent. However, Birch's (1964) method of proof consists essentially of showing that the goodness-of-fit statistic can be written as a quadratic form in the observed proportions and distributed as a chi-square random variable when the observed proportions are close to the expected proportions. The Pearson-Fisher Theorem, as stated in the paper by Birch, is as follows:

Theorem 2.8.

Suppose $\xi(\theta) = (\xi_1(\theta), \xi_2(\theta), \ldots, \xi_\ell(\theta))$ is defined for $\theta \in \mathbb{R}$ for $\ell \in \mathbb{R}$ is a subspace of a $k$-dimensional Cartesian Space $\mathbb{R}^k$. For each $\theta$ and $i$, $\xi_i(\theta) \geq 0$ and $\sum_{i=1}^\ell \xi_i(\theta) = 1$. Suppose the following conditions, which are referred to as regularity conditions, are satisfied:

i) $\theta_0$ is an interior point in $\mathbb{R}$.

ii) Given $\epsilon > 0$, there is a $\delta > 0$ such that $|\xi(\theta) - \xi_0(\theta)| > \delta$ whenever $|\theta - \theta_0| > \epsilon$; that is the inverse function $\xi^{-1}$ is continuous at $\theta_0$.

iii) $\xi_i(\theta_0) > 0$ for each $i$.

iv) For each $i$, constants $a_{ij}$ exists, such that, $\xi_i(\theta) = \xi_i(\theta_0) + \left[\xi_i(\theta_0)\right]^{\frac{1}{2}} \sum_j a_{ij} (\theta_j - \theta_0_j) + O(|\theta - \theta_0|)$ as $\theta \to \theta_0$; that is $\xi_i(\theta)$ is totally differentiable at $\theta_0$ with partial derivatives

$$\frac{\partial \xi_i(\theta)}{\partial \theta_j} = a_{ij} \left[\xi_i(\theta_0)\right]^{\frac{1}{2}}.$$

v) The matrix $A = (a_{ij})$ has rank $K$. 
Let \( \{X_i\} \) be a sequence of random variables each taking the value \( i \) with probability \( p_i \) and let \( \hat{p}_i \) be the proportion of \( X \)'s in the first \( n \) trials taking the value \( i \). Let \( \hat{\theta}_n \) be any value of \( \theta \in \Theta \) for which there exists a sequence \( \{\theta_{nm}\} m = 1, 2, \ldots \), such that \( \hat{\theta}_{nm} \in \Theta \) for each \( m \) and

\[
2n \sum_{i=1}^{I} \hat{p}_i [\ln \xi_i(\theta_{nm}) - \ln \hat{p}_i] + \sup_{\theta} 2n \sum_{i=1}^{I} \hat{p}_i [\ln \xi_i(\theta) - \ln \hat{p}_i]
\]

as \( \hat{\theta}_{nm} \to \hat{\theta}_n \). Then, as \( n \to \infty \)

a) \( \sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, A'A) \),

b) \( X^2 = n \sum_{i=1}^{I} \frac{[\hat{p}_i - \xi_i(\hat{\theta}_n)]^2}{\xi_i(\hat{\theta}_n)} \xrightarrow{d} \chi^2_{I-k-1} \), and

c) \( \sqrt{n}(\hat{\theta}_n - \theta_0) \) and \( X^2 \) are independently distributed.

Note in the theorem above \( A'A \) is the information matrix for \( \theta \) at \( \theta_0 \), \( X^2 \) is called the Pearson goodness-of-fit statistic and

\[
G^2 = -2n \sum_{i=1}^{I} \hat{p}_i \ln \xi_i(\theta) - \hat{p}_i \ln \hat{p}_i = -2\ln \lambda_n(\theta, \xi_n)
\]

is called the log likelihood ratio statistic.

### 2.4. Multinomial Sampling

Consider the goodness-of-fit statistic for \( I \) cells and associated probabilities \( \pi_1, \pi_2, \ldots, \pi_I \) (\( \pi_i > 0 \), \( \sum \pi_i = 1 \)). Let \( x_1, x_2, \ldots, x_I \) denote the observed cell frequencies in a sample drawn according to a multinomial sampling design, then

\[
x^2 = \sum_{i=1}^{I} \frac{(x_i - n\pi_{oi})^2}{n\pi_{oi}}\quad \text{(2.4.1)}
\]
where $\sum_{i=1}^{I} x_i = n$ is the total sample size, is the test statistic for testing the hypothesis

$$H_0: \pi_i = \pi_{oi} \quad (i = 1, 2, \ldots, I).$$  \hfill (2.4.2)

This can be written as

$$x^2 = n \sum_{i=1}^{I} \frac{(\hat{\pi}_i - \pi_{oi})^2}{\pi_{oi}}.$$  \hfill (2.4.3)

This can be expressed in a quadratic form as

$$x^2 = n (\hat{\pi} - \pi_o)^T V_0^{-1} (\hat{\pi} - \pi_o)$$  \hfill (2.4.4)

where the last category is deleted from $\hat{\pi}$ and $\pi_o$ and

$$V_0 = n^{-1} (\Delta - \pi_o \pi_o')$$  \hfill (2.4.5)

is the covariance matrix for $\hat{\pi}$ under $H_0$. The last category is deleted so that $V_0$ may be invertible. $x^2$ is a special case of the Wald Statistic and $\Delta$ is a diagonal matrix with entries $\pi_{oi}$.

It was shown in Section 2.2 how to construct Wald's test statistics given the appropriate covariance matrix. However, under different sampling schemes the covariance matrix may be difficult to obtain. Hence, many researchers will assume a multinomial scheme although the actual design may be cluster sampling, stratified sampling, or some more complex sampling scheme.

Rao and Scott (1981, 1984) examined the behavior of the $x^2$ statistic under complex designs by examining the eigenvalues of the product of
the inverse of the covariance matrix under simple random sampling and that for the actual sampling scheme. The following theorem was used:

Theorem 2.9.

Under the hypothesis $H_0: \pi = \pi_0$, $\chi^2 = \sum_{i=1}^{I-1} \lambda_i Z_i^2$ where $Z_1, Z_2, \ldots$, $Z_{I-1}$ are asymptotically independent normal variates with mean zero and unit variance. The eigenvalues of $V_o^{-1}V$, $\lambda_i$, $i = 1, 2, \ldots, I-1$, are such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{I-1}$ where $V$ is the variance of $\pi$ under the actual design and $V_o$ is the variance under multinomial sampling.


This theorem indicates that $\chi^2$ is distributed as a weighted sum of independent central chi-square ($\chi^2$) random variables. The $\lambda_i$'s are referred to as design effects.

The following two corollaries are immediate consequences of Theorem 2.9.

Corollary 2.9.1.

$$\chi^2 / \lambda_1 \leq \sum_{i=1}^{I-1} Z_i^2$$

where

$$\sum_{i=1}^{I-1} Z_i^2 = n (\pi - \pi_0) (\pi - \pi_0)^\top$$

is distributed asymptotically as $\chi^2_{I-1}$ under $H_0$.

Corollary 2.9.2.

$$\chi^2 / \lambda \sim \chi^2_{I-1}$$ for some constant $\lambda$ if and only if $V = \lambda V_o$. 
Holt, Scott, and Ewings (1980) showed that a correction factor for $X^2$ based on the design effect works well for the test of homogeneity. However, for large sample tests of independence an appropriate modifying factor is more difficult to compute.

Ewings (1979) gave extensive results concerning the varying $\lambda_i$'s. He showed the most important factor is the value of $\bar{\lambda}$, the mean of the $\lambda_i$'s. Changes in the distribution of $\lambda_i$'s about $\bar{\lambda}$ produce only a small change in the significance level.

Define $X_m^2 = X^2/\bar{\lambda}$. Holt et al. showed that $X_m^2$ might work well. Though there may be other modifying factors, e.g., the geometric mean of the $\lambda_i$'s, $\bar{\lambda}$ has one very important advantage. Any estimate of the $\lambda_i$'s in general will require an estimate of $V$, and such an estimate may not be readily available. However, $\bar{\lambda} = (I-1)^{-1} \sum_{i=1}^{I} \delta_{ii}/\Pi_{oi}$ where $\delta_{ii}$ is the diagonal element of $V$, $i = 1, 2, \ldots, I$; can be calculated from the cell variances alone. No information is needed on the covariance terms at all.

Some familiar designs are now examined. Theorem 2.9 is used to obtain conservative tests.

Consider the case of simple random sampling without replacement from a finite population. For testing $H_0: \Pi = \Pi_0$, the variance of the estimator $\hat{\Pi}$ is given by

$$V = N^{-1}(N-n)\Pi_0$$

where $n$ is the sample size and $N$ is the population size. Then, the statistic for testing $H_0$ is given by $X_c^2$, where,

$$X_c^2 = N(N-n)^{-1}X^2 \sim \chi^2_{I-1}. \quad (2.4.6)$$
The design effects, $\lambda_i = N^{-1}(N-n) i = 1,2,...,I$ are constant. It follows that $X^2 > x^2$ and as the sample size increases then the correction factor $N(N-n)^{-1}$ becomes larger and $X^2$ provides an increasing conservative test.

2.5. Stratified Sampling

Several methods of specifying sample sizes for stratified sampling are i) Proportional Allocation, ii) X-proportional Allocation, and iii) Neyman Allocation. These are discussed in sampling texts, such as Sukhatme and Sukhatme (1970) and Kish (1965). Here, consideration is given to proportional allocation since it is widely used in practice to obtain self-weighting estimates and the mathematics are more tractable.

Suppose there are $J$ strata, with a sample size $x_{+j}$ drawn with replacement from the $j$th stratum. Let $\alpha_j = x_{+j}/x_+$ denote the proportion of units in stratum $j$, $\pi_{ij}$ the proportion of units in stratum $j$ in category $i$, and $\pi_i$ the probability of a unit is in category $i$ for the entire population. Since the $x_{+j}$ are selected using proportional allocation, $\pi_i = \sum_{j=1}^{J} \alpha_j \pi_{ij}$. Define $\sum_{j=1}^{J} x_{+j} = x_+$ and let $x_{ij}$ denote the number of units in stratum $j$ belonging to category $i$. Let $\hat{\pi}_{ij}$ be an unbiased estimator of $\pi_{ij}$, then $\hat{\pi}_i = \sum_{j=1}^{J} \hat{\pi}_{ij}$ is an unbiased estimator of $\pi_i$, since

$$E(\hat{\pi}_i) = \sum_{j=1}^{J} \alpha_j \pi_{ij}.$$

(2.5.1)

Since samples in different strata are independent, the variance of $\hat{\pi}_i$ is
Denote the vector of estimated probabilities by \( \hat{\pi} = (\hat{\pi}_1, \hat{\pi}_2, \ldots, \hat{\pi}_{I-1}) \) with the last category deleted to obtain invertible covariance matrix. Then,

\[
\text{var}(\hat{\pi}) = \sum_{j=1}^{I-1} \alpha_j (\Delta_{\pi,j} - \pi_j \pi'_j) \tag{2.5.3}
\]

where \( \pi_j = (\pi_{1j}, \pi_{2j}, \ldots, \pi_{I-1j}) \) and \( \Delta_{\pi,j} \) is a diagonal matrix with elements given by \( \pi_j \). Hence, the covariance matrix \( V_s \) can be written as

\[
V_s = x^{-1}(\Delta_{\pi,j} - \sum_{j=1}^{I-1} \alpha_j \pi_j \pi'_j) \tag{2.5.4}
\]

where \( V_o \) is the covariance matrix under multinomial sampling. Then, the eigenvalues for the matrix \( V_o^{-1}V_s \) are all less than or equal to one, that is, \( \lambda_1 \leq 1 \). Hence, by Corollary 2.9.1 it follows that

\[
0 \leq x^2 = \sum_{i=1}^{I-1} \lambda_i z_i^2 \leq \sum_{i=1}^{I-1} z_i^2 \sim \chi^2_{I-1} \tag{2.5.5}
\]

Therefore, using \( x^2 \) as if the observations came from a multinomial sampling scheme results in a statistic that tends to be smaller
than a chi-square random variable, i.e., it provides a conservative
test. Similar conclusions about proportional allocation for the test
of independence were made by Nathan (1975). Fellegi (1980) and Kish
and Frankel (1974) also obtained similar results for the case of
proportional allocation.

2.6. Cluster Sampling

Suppose there are $S$ primary sampling units (psu) and $M$ sec­
ondary sampling units (ssu) in the $j^{th}$ psu. Suppose $S$ psu's are
sampled with replacement, and in each sampled psu, $m$ secondary units are
sampled with replacement. Define $x_{ijk}=1$ if the $k^{th}$ secondary unit
in the $j^{th}$ primary unit is in category $i$ and $x_{ijk}=0$ otherwise,
k = 1,2,...,$m$, j = 1,2,...,$S$, i = 1,2,...,$I$. Let $x_{i++} = \sum \sum x_{ijk}$,
then,

$$E(x_{i++}) = \mathbb{E}_p \sum_{j=1}^S \mathbb{E}_s \sum_{k=1}^m x_{ijk} \quad (2.6.1)$$

where $\mathbb{E}_p$ denotes the expectation over the primary units and $\mathbb{E}_s$
the expectation over the secondary unit.

$$E(x_{i++}) = \mathbb{E}_p \sum_{j=1}^S m \pi_{ij} \quad (2.6.2)$$

where $P(x_{ijk} = 1) = \pi_{ij}$ so

$$E(x_{i++}) = \sum_{j=1}^S N \alpha_j \pi_{ij} \quad (2.6.3)$$

where $\alpha_j$ is the relative size of $j^{th}$ primary unit, and $N = Sm$.

Let $\hat{\pi}_i = N^{-1} \sum \sum \sum x_{ijk}$ be an estimator for $\pi_i$ where
$\pi = (\pi_1, \pi_2, \ldots, \pi_{I-1})$ is the probability vector for the categories for the entire population. The variance for $\pi$ is given by

$$V_{2s} = N^{-2} \sum_{j=1}^{S} \sum_{k=1}^{m} x_{jk} x_{jk}^\top$$

where $x_{jk} = (x_{1jk}, x_{2jk}, \ldots, x_{I-1jk})$.

$$V_{2s} = N^{-2} \left[ \sum_{j=1}^{S} \sum_{k=1}^{m} x_{jk} x_{jk}^\top + \sum_{j=1}^{S} \sum_{k=1}^{m} x_{jk} x_{jk}^\top \right]$$

$$= N^{-1} \left\{ \sum_{j=1}^{S} \sum_{k=1}^{m} \alpha_j^2 \pi_j \pi_j^\top + (m-1) \sum_{j=1}^{S} \alpha_j \pi_j \pi_j^\top + m \sum_{j=1}^{S} \alpha_j \pi_j \pi_j^\top \right\}$$

$$= N^{-1} \left\{ (\Delta_j - \pi_j \pi_j^\top) + (m-1) \sum_{j=1}^{S} \alpha_j (\pi_j - \pi_j^\top) (\pi_j - \pi_j^\top)^\top \right\}$$

then

$$V_{2s} = V_0 + (m-1)N^{-1} A$$

where

$$V_0 = N^{-1} (\pi_j - \pi_j \pi_j^\top),$$

and

$$A = \sum_{j=1}^{S} \alpha_j (\pi_j - \pi_j^\top) (\pi_j - \pi_j^\top)^\top.$$  

So

$$V_0^{-1} V_{2s} = I + (m-1)V_0^{-1} A.$$  

Let $\mu_1$ denote the $i$th eigenvalue of $V_0^{-1} A$, then the $i$th eigenvalue of $V_0^{-1} V_{2s}$, $\lambda_i$ is given by

$$\lambda_i = 1 + (m-1)\mu_i.$$
Consider a vector of constants $c$ with corresponding dimension, then
\[
\frac{c'Ac}{c'Vc} \leq 1 \quad \text{and the largest eigenvalue of } V^{-1}A, \mu_1 \leq 1. \quad \text{Since } m \geq 1,
\]
\[
\sum_{i=1}^{I-1} z_i^2 \leq \sum_{i=1}^{I-1} [1 + (m-1)\mu_{I-1}] z_i^2 \\
\leq \sum_{i=1}^{I-1} \lambda_i z_i^2 \leq x^2. \tag{2.6.10}
\]

The smallest eigenvalue is denoted by $\mu_{I-1}$. But
\[
x^2 = \sum_{i=1}^{I-1} [1 + (m-1)\mu_i] z_i^2 \\
\leq [1 + (m-1)\cdot 1] \sum_{i=1}^{I-1} z_i^2 = m \sum_{i=1}^{I-1} z_i^2, \tag{2.6.11}
\]
hence,
\[
x^2/m \leq \sum_{i=1}^{I-1} z_i^2 / \chi^2_{I-1}. \tag{2.6.12}
\]

A conservative test can be obtained by dividing $x^2$ by the size of the subsample. Cohen (1976) proposed dividing by the size of the largest subsample when the sizes are unequal. However, this is a crude bound and leaves much to be desired.

2.7. Cluster Models

Cohen considered the $x^2$ statistic under cluster sampling with two (2) individuals per cluster. The sample design consists of $g$ independent clusters each of two (2) individuals, a first and second. Let $X_{ij}$ be the number of clusters in which the first individual falls in cell $i$. 
and second in cell \(j\) \((i, j = 1, 2, \ldots, I)\). Define \(X_i = \sum_{j=1}^{I} X_{ij}\) and \\
\(X_{+j} = \sum_{i=1}^{I} X_{ij}\) and \(Y_i = X_i + X_{+j}\). Then \\
\(\sum_{i=1}^{I} Y_i = 2S\) and \\
\(\sum_{i=1}^{I} X_{+j} = S\). \\

To formalize the notion of independence between clusters Cohen supposed \(X_{ij}\) is distributed as a multinomial with parameters \(S\) and \(\pi_{ij}\). Denote such a distribution by \(\mathbb{M}(S, \pi_{ij})\). To reflect clustering, i.e., positive association within clusters, he suggested the model \\
\[P_{ij} = p_i (\alpha \delta_{ij} + (1-\alpha)p_j), \quad (i, j = 1, 2, \ldots, I) \quad (2.7.1)\]

where \(p_i\) is the probability that an individual is in category \(i\), and \(\alpha\) is a constant, such that, \(0 \leq \alpha \leq 1\), and \(\delta_{ij}\) is an indicator variable such that \\
\[\delta_{ij} = 1 \text{ if } i = j \]
\[= 0 \text{ otherwise } . \quad (2.7.2)\]

Hence, \\
\[P_{ij} = p_i (\alpha + p_j - \alpha p_j) \quad i = j \]
\[= (1-\alpha)p_i p_j \quad i \neq j . \quad (2.7.3)\]

There are two (2) special and extreme cases.

i) When \(\alpha = 1\), then \(P_{ij} = p_i\) when \(i = j\) and \(P_{ij} = 0\) if \(i \neq j\). So the status of the second individual is completely determined by the first (complete dependence). Consequently, one is essentially looking at \(S\) observations.
ii) When \( a = 0 \), then \( p_{ij} = p_i^2 \) when \( i = j \) and \( p_{ij} = p_i p_j \) when \( i \neq j \). This is the case of complete independence within the clusters.

In case i)
\[
\chi^2/2 = \sum_{i=1}^{l} (y_i - 2sp_i)^2 / 2sp_i \sim \chi^2_{l-1} \quad (2.7.4)
\]

and in case ii)
\[
\chi^2 = \sum_{i=1}^{l} (y_i - 2sp_i)^2 / 2sp_i \sim \chi^2_{l-1} \quad (2.7.5)
\]

which is the usual statistic under multinomial sampling. Cohen (1976) stated and proved the following theorem which is the highlight of his paper.

Theorem 2.10.

Given the model above in (2.7.1), \( x^2/(1+a) \sim \chi^2_{l-1} \) as \( \delta \to \infty \).

Altham (1976) extended the idea to \( k \) individuals per cluster. She constructed a covariance matrix, \( V \), with (ij)th element given by

\[
V_{ij} = k(p_i - p_i^2) + k(k-1)(p_{ii} - p_i^2) \quad i = j
\]
\[
= k(k-1)p_{ij} \quad i \neq j , \quad (2.7.6)
\]

where \( p_{ii} \) is defined in (2.7.3).

She stated and proved the following theorem.

Theorem 2.11.

Let \( x^2 = \sum_{i=1}^{l} (y_i - ksp_i)/ksp_i \) where \( k \) is the number of individuals per cluster, then
\[ \frac{x^2}{k} \leq \left( \sum_{i=1}^{k} - \frac{1}{k} \right) \leq x^2 \]

where \( \mathbf{y} = (y_1, y_2, \ldots, y_k) \) and \( \mathbf{p} = (p_1, p_2, \ldots, p_k) \).

Brier (1980) made use of the Dirichlet-Multinomial distribution as a means of modeling the dependency within clusters. He assumed that the vectors of probabilities \( \mathbf{p} \) for a cluster had a prior distribution given by the Dirichlet distribution. This distribution allows an arbitrary number of response categories and arbitrary cluster sizes. This model is discussed in greater detail in Section 4.2. One theorem of great importance as stated and proved by Brier is

**Theorem 2.12.**

Let \( X_1, X_2, \ldots, X_S \) be independent and identically distributed as Dirichlet-Multinomial with parameters \( n, \upsilon \) and \( k \) denoted by \( \text{DM}_k(n, \upsilon, k) \). Assume that Birch's regularity conditions as given in Theorem 2.8 hold for the null hypothesis \( H_0: \upsilon = f(\bar{g}) \). Then as \( S \to \infty \)

\[ X^*_2 \xrightarrow{D} \chi^2 \quad \text{and} \quad G^*_2 \xrightarrow{D} \chi^2 \]

where \( X^*_2 \) is the usual Pearson goodness-of-fit statistic that one would compute for a simple random sample. \( G^*_2 \) is the corresponding likelihood ratio goodness-of-fit statistic. The * indicates that the distribution is derived for the Dirichlet sampling model instead of the multinomial model. This model for cluster sampling is further discussed in Chapter 4.

### 2.8. Other Models

Rao and Scott (1981, 1984) gave two models, one for two-stage sampling and another for stratified two-stage sampling.

For the two-stage model, assume that \( S \) primary sampling units (psu)
are sampled from $\mathcal{S}$ psu's and $m_j$ secondary units (ssu) are sampled from $M_j$ ssu's within the $j^{th}$ psu in the sample. There are $x_{jk}$ elements sampled from $k^{th}$ ssu of the $j^{th}$ psu, $k = 1,2,\ldots,m_j$, $j = 1,2,\ldots,n$. Define $\sum_j x_{jk} = n_j$. Define $\sum_k x_{jk} = n$. as the number of sampled elements in the $j^{th}$ psu and let

$$Z_{ijkl} = 1 \text{ if the } l^{th} \text{ element of the } k^{th} \text{ ssu of the } j^{th} \text{ psu is in category } i$$

$$= 0 \text{ otherwise.} \quad (2.8.1)$$

Then,

$$\mathbb{E}(Z_{ijkl}) = \pi_i \quad (2.8.2)$$

where $\pi_i$ is the probability of being in category $i$ and

$$\text{Cov}(Z_{ijkl}, Z_{ij'k'l'}) = b_{ii'}, j=j', k=k', l\neq l'$$

$$= d_{ii'}, j=j', k\neq k', l\neq l'$$

$$= 0 \quad j\neq j'. \quad (2.8.3)$$

Let $\mathbf{x} = (x_{11}, x_{21}, \ldots, x_{I-1,1})$ be the vector of counts and

$$\mathbf{\pi} = (\pi_1, \pi_2, \ldots, \pi_{I-1})$$

the probability vector. The covariance matrix for the estimator $\hat{\pi}$ is given by

$$V_o = V_o + N^{-1} \left[ (\sum_{jk} x_{jk}^2 - N)B + (\sum_j n_j^2 - \sum_{jk} x_{jk}^2)D \right] \quad (2.8.4)$$

where $V_o$ is the covariance matrix under multinomial sampling,

$B = (b_{ii'}), D = (d_{ii'})$ are matrices. It can be shown that $V_oB$ and $V_oD$ are nonnegative definite, and if $\lambda_i(i = 1,2,\ldots,I)$ denote the eigenvalues of $V_o^{-1}V$, then
\[
\hat{\lambda}_1 \leq 1 + \sum_{jk} \left( \frac{x_{jk}^2}{N} - \frac{N}{N} \right) + \left( \frac{\sum_{j} \mu_{jk}^2}{N} - \frac{\sum_{jk} \mu_{jk}^2}{N} \right), \quad (2.8.5)
\]

Therefore, \( \hat{\lambda}_1 \leq \sum_{j} n_j^2/N \) and it follows that

\[
x^2 = \sum_{i=1}^{I-1} \lambda_i z_i^2 \leq \left( \sum_{j} n_j^2/N \right) \sum_{i=1}^{I-1} z_i^2, \quad (2.8.6)
\]

hence,

\[
x^2/\sum_{j} (n_j^2/N) \leq \sum_{i=1}^{I-1} z_i^2 \sim \chi^2_{I-1}, \quad (2.8.7)
\]

so \( x^2 \) is conservative.

In the case of the stratified two-stage sampling, consider sampling \( S_j \) of \( s_j \) psu's in stratum \( j \). Within each stratum from the \( k^{th} \) psu, a sample of \( x_{+jk} \) ssu's is obtained by simple random sampling \( k = 1, 2, \ldots, S_j \); \( j = 1, 2, \ldots, J \). Define

\[
z_{ijk\ell} = 1 \quad \text{if the } \ell^{th} \text{ element of the } k^{th} \text{ ssu in the } j^{th} \text{ stratum is in category } i
\]

\[
= 0 \quad \text{otherwise}, \quad (2.8.8)
\]

and

\[
\text{Cov}(z_{ijk\ell}, z_{ijk\ell'}) = b_{ii} \quad \ell \neq \ell'. \quad (2.8.9)
\]

Let \( \hat{x}_j = (x_{1j}, x_{2j}, \ldots, x_{IJ}) \) denote the vector of category totals for stratum \( j \) and let \( \hat{\pi}_j \) denote the corresponding probability vector. Then, if \( \hat{\pi}_j \) is an unbiased estimator for \( \pi_j \), under the sampling scheme presented here where
\[ \widehat{\mathbf{N}}_j = \mathbf{N}_j^{-1} \mathbf{N}_j \]  

and

\[ \text{var}(\widehat{\mathbf{N}}_j) = \mathbf{V}_{o,j} + \mathbf{N}_j(n_j - 1) \mathbf{B}_j, \]  

where \( x_{+jk} = n_j \) for all \( k = 1, 2, \ldots, S_j \). Furthermore, \( \mathbf{V}_{o,j} \) is the matrix given by

\[ \mathbf{V}_{o,j} = \mathbf{N}_j^{-1} \left( \frac{\mathbf{N}_j}{\mathbf{N}_j} - \mathbf{N}_j \mathbf{N}_j' \right), \]  

and \( \mathbf{B}_j = (b_{ii,j}) \).

Let \( \mathbf{\pi} = (\pi_1, \pi_2, \ldots, \pi_{I-1}) \) be the probability vector for the \( I-1 \) categories. Define \( \mathbf{\pi} = \sum_{j=1}^{J} \alpha_j \mathbf{N}_j \), where \( \alpha_j = N_j/N \), then the variance of the estimator \( \widehat{\mathbf{\pi}} = \sum_{j=1}^{J} \alpha_j \mathbf{N}_j \)

\[ \text{var}(\widehat{\mathbf{\pi}}) = \text{var}(\sum_{j=1}^{J} \alpha_j \mathbf{N}_j) = \sum_{j=1}^{J} \alpha_j^2 \mathbf{N}_j^{-1} \left[ \mathbf{V}_{o,j} + (n_j - 1) \mathbf{B}_j \right]. \]  

Under proportional allocation, \( (\alpha_j = N_j/N) \) the variance of \( \widehat{\mathbf{\pi}} \), is given by

\[ \mathbf{V}_{2st} = \mathbf{N}^{-1} \sum_{j=1}^{J} \alpha_j \left[ \mathbf{V}_{o,j} + (n_j - 1) \mathbf{B}_j \right]. \]  

Consequently,

\[ \mathbf{V}_{o}^{-1} \mathbf{V}_{2st} = \mathbf{V}_{o}^{-1} \left[ \mathbf{N}^{-1} \sum_{j=1}^{J} \alpha_j \left[ \mathbf{V}_{o,j} + (n_j - 1) \mathbf{B}_j \right] \right], \]  

where \( \mathbf{V}_{o} = \mathbf{N}^{-1} \left( \frac{\mathbf{N}}{\mathbf{N}} - \mathbf{N} \mathbf{N}' \right) \).
and since \( V_{oj} - B_{oj} \) is nonnegative definite,

\[
V_{2st} = N^{-1} \sum_{j=1}^{J} \alpha_j [V_{oj} - B_j + n_j B_j]
\]

\[
\leq N^{-1} \sum_{j=1}^{J} \alpha_j [(n_j - 1)V_{oj} + (n_j - 1)B_j]
\]

\[
\leq N^{-1} \sum_{j=1}^{J} \alpha_j (n_j - 1)V_{oj}
\]

\[
\leq N^{-1} \sum_{j=1}^{J} \alpha_j n_j V_{oj} \leq \max_{j=1}^{J} \sum_{j=1}^{J} \alpha_j V_{oj} . \tag{2.8.16}
\]

But

\[
\sum_{j=1}^{J} \alpha_j V_{oj} = V_o - \Sigma_{j} (n_j - \nu_j)\overline{(n_j - \nu_j)} \tag{2.8.17}
\]

therefore,

\[
V_{2st} \leq \max_{j=1}^{J} V_{oj} \tag{2.8.18}
\]

and

\[
V_o^{-1} V_{2st} \leq \max_{j=1}^{J} I = aI \tag{2.8.19}
\]

where \( \max_{j=1}^{J} = a \), then

\[
x^2 = \sum_{i=1}^{I-1} \lambda_i z_i^2 \leq \sum_{i=1}^{I-1} a z_i^2
\]

\[
= a \sum_{i=1}^{I-1} z_i^2 \tag{2.8.20}
\]

hence,

\[
x^2/a \leq \sum_{i=1}^{I-1} z_i^2 \sim \chi^2_{I-1} . \tag{2.8.21}
\]
3. WALD STATISTICS

3.1. Introduction

Methods for the analysis of categorical data have been developed extensively under the assumption of multinomial sampling; in particular, there are chi-square ($\chi^2$) tests for the goodness-of-fit hypothesis and hypotheses of independence and homogeneity in two-dimensional contingency tables. Recent extensions of these methods to multidimensional contingency tables using log linear models (Fienberg (1977), Bishop et al. (1975), and Haberman (1974)) have attracted considerable attention due to their close similarity to analysis of variance models in providing systematic tests of various hypotheses.

However, in many situations, the counts in the table arise from a clustered sampling scheme, a stratified sampling scheme, or a combination of both, and the multinomial distribution will not be the correct distribution for the observed counts. In some situations, the goodness-of-fit statistics based on multinomial sampling may provide conservative tests.

Rao and Scott (1981) showed that the usual statistics based on the multinomial model for the test of independence and the test of homogeneity lead to conservative tests when simple corrections are model for cluster sampling. For the special case of stratified simple random sampling under proportional allocation, the usual statistics provide a conservative test with no correction.
By the use of simple models for clustering, Altham (1976), Cohen (1976), and Brier (1980) derived appropriate $\chi^2$ tests for goodness-of-fit in the case of single-stage cluster sampling (with equal subsample sizes) and two-stage cluster sampling (with equal subsample sizes). They have shown that a simple correction leads to an asymptotically chi-square test statistic for those cluster sampling schemes.

In this chapter, Wald Statistics for the goodness-of-fit problem, the test of homogeneity, and the test of independence, are constructed for different sampling schemes. These test statistics are compared to the results obtained for the multinomial model. Some comparisons are made through the use of the properties of covariance matrices and matrix inversion theorems.

In Section 3.2, the parameter values are defined, the general problem is outlined, and results for multinomial sampling are reviewed. In Section 3.3, a stratified sampling scheme is examined, Wald Statistics are developed for certain hypotheses, and the techniques for stratified sampling used by SUPERCARP are reviewed. In Section 3.4, a two-stage sampling scheme is examined and test statistics are obtained.

Suppose that a classification with $I$ categories can be made in each of $J$ subpopulations, and a unit selected at random from the $j^{th}$ subpopulation is in the $i^{th}$ category with probability $\pi_{ij}$, for $i = 1,2,...,I$ and $j = 1,2,...,J$. A sample is taken from each subpopulation and the units are classified. Denote by $x_{ij}$ the observed frequencies for the $i^{th}$ category in the $j^{th}$ subpopulation, and $X_{ij}$
as the random variable taking on the values $x_{ij}$. The frequencies can be presented as a contingency table of order $I \times J$, where rows correspond to categories and the subpopulations correspond to columns. Define

$$x_{i+} = \sum_{j=1}^{J} x_{ij},$$

$$x_{+j} = \sum_{i=1}^{I} x_{ij},$$

and

$$N = \sum_{i=1}^{I} \sum_{j=1}^{J} x_{ij}.$$

The following matrix results will be useful in the development of the material throughout this chapter. In this chapter, and throughout this dissertation, matrix and vector inequalities refer to elementwise comparisons. For example, $\mathbf{t} \succ 0$ indicates that each element of the vector $\mathbf{t}$ is positive.

**Theorem 3.1.**

Let $A$ be a nonsingular matrix and $u$ and $v$ be two column vectors, then

$$\left(A+uv^T\right)^{-1} = A^{-1} - \frac{(A^{-1}u)(v^TA^{-1})}{1+v^TA^{-1}u}.$$
Proof:

\[
\begin{bmatrix}
A^{-1} - \frac{(A^{-1}u)(y', A^{-1})}{1 + y' A^{-1} u}
\end{bmatrix}
(A + uv') = I - \frac{(A^{-1}u)y'}{1 + y' A^{-1} u} + A^{-1} uv' - \frac{A^{-1}uv'A^{-1}uv'}{1 + y' A^{-1} u}
\]

\[
= I - \frac{A^{-1}uv' + A^{-1}uv'A^{-1}uv'}{1 + y' A^{-1} u} + A^{-1}uv'
\]

\[
= I
\]

where \( I \) is the identity matrix.

Corollary 3.1.1.

Let \( W = \Delta + \alpha \mathbf{t} \mathbf{t}' \) where \( \Delta \) is a nonsingular diagonal matrix with elements \( t_i \) for \( i = 1, 2, \ldots, I \), and \( t = (t_1, t_2, \ldots, t_I) \) is a column vector, and \( \alpha \) is a scalar, such that,

\[
\alpha \neq - \left[ \sum_{i=1}^{I} t_i \right]^{-1}.
\]

Then, the inverse of \( W \) is

\[
W^{-1} = \Delta^{-1} + \gamma \mathbf{J},
\]

where

\[
\gamma = - \alpha \left[ I + \alpha \sum_{i=1}^{I} t_i \right]^{-1},
\]
and \( J \) is a square matrix of order \( I \) with all elements equal to 1.

Proof: Show that \( WW^{-1} = I \). This was done by Graybill (1969).

Theorem 3.2.

Consider the nonsingular matrices \( A \) and \( B \), then if \( I + BA \) is nonsingular and well-defined

\[
M^{-1} = (I+AB)^{-1} = I - A(I+BA)^{-1}B.
\]

Proof:

\[
(I+AB)M^{-1} = I+AB - (I+AB)A(I+BA)^{-1}B
\]

\[
= I+AB - A(I+BA)(I+BA)^{-1}B
\]

\[
= I.
\]

Corollary 3.2.1.

Suppose \( A, B \) and \( A+B \) are nonsingular, then

\[
(A+B)^{-1} = A^{-1} - A^{-1}(B^{-1}+A^{-1})^{-1}A^{-1}.
\]

Proof:

\[
(A+B)^{-1} = [A(I+A^{-1}B)]^{-1}
\]

\[
= (I+A^{-1}B)^{-1}A^{-1},
\]

and by use of Theorem 3.2,

\[
(A+B)^{-1} = [I-A^{-1}(I+BA^{-1})^{-1}B]A^{-1},
\]

\[
= [I-A^{-1}(B^{-1}+A^{-1})^{-1}]A^{-1},
\]

\[
= A^{-1} - A^{-1}(B^{-1}+A^{-1})^{-1}A^{-1}.
\]
Definition 3.1.

An $n \times n$ matrix $A$ is defined to be positive semi-definite, if and only if

i) $A = A'$

ii) $y' Ay \geq 0$ for each and every vector $y$ in $\mathbb{R}^n$ and the equality holds for at least one vector $\tilde{y}$ such that $\tilde{y} \neq 0$.

Definition 3.2.

An $n \times n$ matrix $A$ is defined to be positive definite, if and only if

i) $A = A'$

ii) $y' Ay > 0$ for each and every vector $y$ in $\mathbb{R}^n$ such that $y \neq 0$.

Definition 3.3.

A matrix is defined to be nonnegative definite, if and only if, it is either positive definite or positive semi-definite.

Theorem 3.3.

Let $A$, $B$ and $C$ be symmetric $k \times k$ matrices.

i) If $A$ and $B$ are nonsingular and $A-B$ is positive definite, then $B^{-1} - A^{-1}$ is positive definite.

ii) If $A$ and $B$ are nonsingular, such that, $x'A\tilde{x} > x'B\tilde{x}$ for each and every vector $\tilde{x} \neq 0$, then $x'A^{-1}x < x'B^{-1}x$ for each and every vector $\tilde{x} \neq 0$.

iii) Let $A$ be positive definite. For any $k \times 1$ vectors $\tilde{x}$ and $\tilde{y}$, the following inequality holds:
\[(x'y)^2 \leq (x'Ax)(y'A^{-1}y),\]

and the equality holds, if and only if, there is a scalar \(s\) such that,

\[Ax = sy.\]

iv) If \(A\) is a positive definite \(k \times k\) matrix and \(B\) is a non-negative \(k \times k\) matrix, then

\[\lambda_1 \leq \frac{x'Bx}{x'Ax} \leq \lambda_k\]

for each and every vector \(x \neq 0\) where \(\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_k\) are the roots of

\[|B-\lambda A| = 0.\]


Theorem 3.4.

If \(A\) is a symmetric \(k \times k\) matrix and if

\[a_{ii} > \sum_{j \neq i} |a_{ij}| \quad \text{for all} \quad i = 1, 2, \ldots, k,
\]

where \(A = (a_{ij})\), then \(A\) is a positive definite matrix.


Theorem 3.5.

Let \(\{A_1, A_2, \ldots, A_k\}\) be a collection of \(n \times n\) positive definite matrices and let \(\{\alpha_1, \alpha_2, \ldots, \alpha_k\}\) be a set of positive scalars. The matrix
is also positive definite.


Theorem 3.6,

Let \( A \) be a \( k \times k \) positive definite matrix. If \( a_{ij} < 0 \) for all \( i \neq j \), then every element in \( A^{-1} \) is positive.


### 3.2. Simple Random Sampling

In general, a set of \( J \) subpopulations is considered for which the \( j^{th} \) subpopulation contains \( N_j \) primary units. Each member, or secondary unit, of the \( k^{th} \) primary unit in the \( j^{th} \) subpopulation can be classified into one of \( I \) distinct categories. Let

\[
P_{ijk} = (p_{1jk}, p_{2jk}, \ldots, p_{Ijk})'
\]

be the vector of true proportions for the \( k^{th} \) primary unit of the \( j^{th} \) subpopulation; that is,

\[
p_{ijk} \text{ is the probability that the observation is in category } i \text{ given the } k^{th} \text{ primary unit in the } j^{th} \text{ subpopulation.}
\]

The population parameters of interest are the vectors of proportions for each subpopulation

\[
\pi_j = (\pi_{1j}, \pi_{2j}, \ldots, \pi_{Ij})'
\]

\[
N_j = \sum_{k=1}^{N} a_{jk} \pi_{jk}.
\]
and the vector of proportions for the combined population

\[ \pi_j = \sum_{j=1}^{J} a_{jk} \rho_j, \]  

(3.2.4)

where \( a_{jk} \) is the relative size of the \( k^{th} \) primary unit in the \( j^{th} \) subpopulation, with

\[ \sum_{k=1}^{N_j} a_{jk} = 1. \]  

(3.2.5)

The scalar \( \alpha_j \) is the relative size of the \( j^{th} \) subpopulation with

\[ \sum_{j=1}^{J} \alpha_j = 1. \]  

(3.2.6)

The vector \( \pi_j \) is a weighted linear combination of the vectors \( p_{jk} \) \( (k = 1, 2, \ldots, N_j) \), with weights proportional to size.

When each primary unit is of equal size, it is simply an average of the vectors \( p_{jk} \), \( k = 1, 2, \ldots, N_j \). Before discussing more complicated sampling schemes, it is appropriate to obtain results for simple random sampling from a single population.

Suppose the \( J \) subpopulations are considered as one single population and a sample of \( N \) secondary units is obtained by simple random sampling with replacement from this combined population. The primary unit designation is ignored. Let \( \chi_s = (x_1, x_2, \ldots, x_i) \) be the vector of observed frequencies. Then \( \chi_s \) has a multinomial distribution given by

\[ \Pr(\chi_s = \chi) = N! \prod_{i=1}^{I} \frac{x_i!}{\pi_i!}, \]

(3.2.7)
where \( \prod_{i=1}^{I} \) denotes the product over the possible values of \( i \). The assumption is made without loss of generality that \( \pi_i > 0 \) for every \( i \). This is needed to obtain nonsingular covariance matrices in the development of test statistic. Theorem 2.8 in Section 2.4, showed that the Pearson Statistic,

\[
x_G^2 = N \sum_{i=1}^{I} \frac{(\hat{\pi}_{im} - \pi_{oi})^2}{\pi_{oi}},
\]

for testing the hypothesis

\[
H_0: \pi_i = \pi_{oi}, \text{ for all } i = 1, 2, \ldots, I,
\]

is distributed asymptotically as a chi-square random variable with \( I-1 \) degrees of freedom, where \( \hat{\pi}_{im} \) is the unbiased estimator of \( \pi_i \) given by

\[
\hat{\pi}_{im} = \frac{x_i}{N}.
\]

The covariance matrix of \( \hat{\pi}_{im} \) is

\[
\text{var}(\hat{\pi}_{im}) = N^{-1}(\Delta_{\pi} - \pi \pi'),
\]

Since \( \sum_{i=1}^{I} \pi_i = 1 \), the covariance matrix is singular. Henceforth, the probability vectors will have the \( i^{th} \) category deleted to obtain non-singular covariance matrices. Therefore, in (3.2.4) the probability vector, \( \pi_{\sim} \) is defined as

\[
\pi_{\sim} = (\pi_1, \pi_2, \ldots, \pi_{I-1})',
\]

and \( \Delta_{\pi_{\sim}} \) is a diagonal matrix with elements given by \( \pi_{\sim} \) in (3.2.12).
Let

\[ \Sigma_m = N^{-1}(\hat{\pi}_m - \pi \pi') \quad (3.2.13) \]

denote the \((I-1) \times (I-1)\) covariance matrix for \(\hat{\pi}\). The test statistic \(X^2_G\) can be written in the form of a Wald Statistic given by

\[ X^2_G = N(\hat{\pi}_m - \pi_0)'(\Delta - \pi_0 \pi_0')^{-1}(\hat{\pi}_m - \pi_0) \quad (3.2.14) \]

where

\[ \hat{\pi}_m = (\hat{\pi}_1m, \hat{\pi}_2m, ..., \hat{\pi}_{I-1}m)' \quad (3.2.15) \]

and

\[ \pi_0 = (\pi_{10}, \pi_{20}, ..., \pi_{I-10})' \]

Suppose instead there are the \(J\) subpopulations and the interest was in the hypothesis

\[ H_0 : \pi_{ij} = \pi_{i0} \quad j = 1,2,...,J \quad (3.2.16) \]

that is, the test of homogeneity for parallel samples where the values \(\pi_{i0}, i = 1,2,...,I-1\), are completely specified. For the \(j^{th}\) subpopulation a simple random sample of \(N_j\) units is taken, ignoring the possible designation of primary units. For the \(j^{th}\) subpopulation, a vector of observed frequencies

\[ \hat{x}_j = (x_{1j}, x_{2j}, ..., x_{Ij}) \quad (3.2.17) \]

is obtained where \(x_{ij}\) is the sum of observations in the \(j^{th}\) subpopulation found in category \(i\). An estimator from the \(j^{th}\) subsample
for $\pi_{ij}$ is given by

$$\pi_{ij,m} = \frac{x_{ij}}{N_j} \quad (3.2.18)$$

and a test statistic for testing the hypothesis in (3.2.16) is given by

$$X_{H}^2 = \sum_{j=1}^{J} \sum_{i=1}^{N_j} \left( \pi_{ij,m}^{\wedge} - \pi_{ij}^{\wedge} \right)^2 / \pi_{ij}^{\wedge} \cdot \quad (3.2.19)$$

This statistic may be written as

$$X_{H}^2 = \sum_{j=1}^{J} \sum_{i=1}^{N_j} \left( \pi_{ij,m}^{\wedge} - \pi_{ij}^{\wedge} \right) \Sigma^{-1}(j)(\pi_{ij,m}^{\wedge} - \pi_{ij}^{\wedge}) , \quad (3.2.20)$$

where $\Sigma_m(j)$ is the covariance matrix for $\pi_{ij,m}^{\wedge}$ given by

$$\Sigma_m(j) = x^{-1}(j)(\Lambda_{\pi_{ij,m}^{\wedge}} - \pi_{ij}^{\wedge}) \quad (3.2.21)$$

and

$$\pi_{ij,m}^{\wedge} = (\pi_{1j,m}^{\wedge}, \pi_{2j,m}^{\wedge}, \ldots, \pi_{I-1,j,m}^{\wedge})' . \quad (3.2.22)$$

The statistic $X_{H}^2$ is asymptotically distributed as a chi-square random variable with degrees of freedom $J(I-1)$.

Suppose in (3.2.16) the vector $\pi_{i}^{\wedge}$ was unknown and had to be estimated by, say $\pi_{i}^{\wedge}$, where

$$\pi_{i}^{\wedge} = N^{-1}(\sum_{j=1}^{J} X_j) , \quad (3.2.23)$$

then, the hypothesis
is usually referred to as the hypothesis of independence. The Pearson test statistic can be written as

\[ x^2_i = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(x_{ij} - \hat{\pi}_{ij} N_j / N)^2}{\hat{\pi}_{ij} N_j / N} \]

\[ = \sum_{j=1}^{J} N_j \sum_{i=1}^{I} \frac{(\hat{\pi}_{ij} - \hat{\pi}_{10})^2}{\hat{\pi}_{10}} \]

\[ = \sum_{j=1}^{J} N_j (\frac{\hat{\pi}_{ij} - \hat{\pi}_{10}}{\hat{\pi}_{10}})^2, \quad (3.2.24) \]

and is distributed asymptotically as a chi-square random variable with \((I-1)(J-1)\) degrees of freedom.

3.3. Stratified Sampling

An established fact in simple random sampling is that the variance of the estimate, say of the mean, depends, apart from the sample size, on the variability of the members of the population. If the population is very heterogeneous and considerations of cost limit the size of the sample, it may be impossible to get a sufficiently precise estimate by taking a simple random sample from the entire population. Many populations encountered in practice and, in particular, the example given at the end of this section are quite heterogeneous. In surveys based on income levels, for example, it can be that some income levels occur
frequently, while others, (usually large incomes) occur infrequently. Any estimate made from a direct random sample taken from the totality of all income levels would be subject to exceedingly large sampling fluctuations. A more precise estimate can usually be obtained by dividing the population into parts or strata on the basis of income levels. Then, an independent simple random sample is taken from each stratum. Each of the estimates obtained from the $j^{th}$ stratum is combined to obtain a more precise estimate of the over-all mean. This is the basic idea involved in stratified sampling.

In this section, consideration is given to such a sampling scheme. A Wald test statistic is constructed for each of the following hypotheses. The goodness-of-fit problem, $H_0: \Pi = \Pi_o$, the case of homogeneity among the samples from the different strata, $H_0: \Pi_j = \Pi_o$ ($\Pi_o$ is a known vector) and a test of independence $H_0: \Pi_j = \Pi_o$ ($\Pi_o$ is an unknown vector). The techniques used in the SUPERCARP computer package for testing the hypotheses for the goodness-of-fit and the test of homogeneity are reviewed. Data obtained from Sewell and Shah (1968) based on students from a Wisconsin school are analyzed using the results obtained in this section.

Consider a population divided into $J$ parts (strata). A sample of $N_j$ units is randomly drawn with replacement from the $j^{th}$ stratum, $j = 1, 2, \ldots, J$. Within each stratum, each of the observed units is classified into one of $I$ mutually exclusive categories. For the $j^{th}$ stratum, the corresponding vector of observed counts is denoted by
where $x_{ij}$ is the number of units found in the $i^{th}$ category of the $j^{th}$ stratum. Let

$$\Pi_j = (\pi_{1j}, \pi_{2j}, \ldots, \pi_{lj})'$$

be the true vector of proportions for the $j^{th}$ stratum. Let

$$\pi = (\pi_1, \pi_2, \ldots, \pi_L)'$$

be the true vector of proportions for the entire population. Then,

$$\pi = \frac{1}{\sum_j \alpha_j} \sum_{j=1}^{J} \alpha_j \Pi_j$$

(3.3.1)

where $\alpha_j$ is proportional to the relative size of the $j^{th}$ stratum and

$$\sum_{j=1}^{J} \alpha_j = 1.$$

Let the estimator $\hat{\Pi}_j$ be given by

$$\hat{\Pi}_j = N_j^{-1} \bar{x}_j$$

(3.3.2)

then, $E(\hat{\Pi}_j) = \Pi_j$ since $\bar{x}_j$ has a multinomial distribution with sample size $N_j$ and probability vector $\Pi_j$. The variance of this unbiased estimator of $\Pi_j$ is

$$\text{var}(\hat{\Pi}_j) = N_j^{-1}(\Delta_{\hat{\Pi}_j} - \Pi_j \Pi_j')$$

(3.3.3)
Let the estimator \( \hat{\pi} \) be

\[
\hat{\pi} = \sum_{j=1}^{J} \alpha_j \hat{\pi}_j
\]  
(3.3.4)

then,

\[
E(\hat{\pi}) = \sum_{j=1}^{J} \alpha_j \hat{\pi}_j .
\]  
(3.3.5)

So for fixed \( N_j \), \( \hat{\pi} \) is an unbiased estimator of the vector \( \pi \).

The covariance matrix is

\[
\text{var}(\pi) = \sum_{j=1}^{J} \alpha_j^2 \pi_j \pi_j - \alpha_j \pi_j \pi_j'
\]  
(3.3.6)

which reduces to

\[
\text{var}(\pi) = N^{-1} \pi \pi' - \sum_{j=1}^{J} \alpha_j \pi_j \pi_j',
\]  
(3.3.7)

when \( \alpha_j \) is chosen so that

\[
\alpha_j = N_j / N .
\]  
(3.3.8)

The covariance matrix for \( \hat{\pi} \) can be expressed as

\[
\Sigma_s = N^{-1} (\Delta_\pi - S)
\]  
(3.3.9)

where
\[ \Delta_\pi = \text{diag}(\pi_1, \pi_2, \ldots, \pi_{I-1}) \]

and

\[ S = \sum_{j=1}^{J} \alpha_j \pi_j \pi_j' \quad (3.3.10) \]

for positive scalars \( \alpha_j, j = 1,2,\ldots,J \); and

\[ \pi_j = (\pi_{1j}, \pi_{2j}, \ldots, \pi_{I-1j})' \]

Since all proportions are assumed to be positive, the matrix \( \Sigma_s \) is positive definite. Also, \( \Delta_\pi \) is a positive definite matrix. Furthermore, the elements of the symmetric matrix \( S \) are all positive.

The covariance matrix, \( \Sigma_s \) can be expressed as

\[ \Sigma_s = N^{-1} [\Delta_\pi - \pi \pi'] - \sum_{j=1}^{J} \alpha_j (\pi_j - \pi)(\pi_j - \pi)' \]

\[ = N^{-1}(\Delta_\pi - \pi \pi') - R \quad (3.3.11) \]

where

\[ R = \sum_{j=1}^{J} \alpha_j (\pi_j - \pi)(\pi_j - \pi)' \quad (3.3.12) \]

Let

\[ V = N^{-1}(\Delta_\pi - \pi \pi') \quad (3.3.13) \]
then,

$$\Sigma_s = V^{-R} \quad (3.3.14)$$

and by use of Corollary 3.2.1,

$$\Sigma_s^{-1} = (I - V^{-1}R)^{-1}V^{-1}$$

$$= V^{-1} + V^{-1}(R^{-1} - V^{-1})^{-1}V^{-1} \quad (3.3.15)$$

where $R$ is assumed to be nonsingular, which implies the number of strata is greater than the number of categories, and $(I - V^{-1}R)$ is assumed to be nonsingular. By use of Theorem 3.3, $R^{-1} - V^{-1}$ is positive definite when $\Sigma_s$ is positive definite. Then,

$$x'V^{-1}(R^{-1} - V^{-1})^{-1}V^{-1}x > 0 \quad (3.3.16)$$

for any vector $x \neq 0$, hence

$$x'\Sigma_s^{-1}x \geq x'V^{-1}x \quad (3.3.17)$$

This result will be used to show that the Pearson chi-squared statistic for the multinomial model provides a lower bound and hence, a conservative test for the stratified sampling scheme.

Note that the factor $R$ is a measure of the among strata variation in the probability vector. $R$ is a matrix of zeros when the vectors of proportions are the same in all strata. In that case, there is no stratum effect and the statistic reduces to the usual
Pearson statistic. The matrix $R$ is nonnegative definite and this fact will be used to show that the Pearson statistic is an upper bound.

To test the hypothesis $H_0: \pi = \pi_0$ where $\pi_0$ is known and $\pi$ denotes the true probability vector, a Wald test statistic, $X^2_{WGS}$ can be constructed. One such statistic is

$$X^2_{WGS} = N \left( \pi - \pi_0 \right)' \Sigma_s^{-1} \left( \pi - \pi_0 \right) \quad (3.3.18)$$

where

$$\Sigma_s = \left( \Delta^\pi - \sum_{j=1}^J \alpha_j \pi_j \pi_j' \right)/N, \quad (3.3.19)$$

$\pi = \Sigma x \pi_j$ and $\pi_j$ is a consistent estimator of $\pi_j$.

By expression (3.3.17)

$$X^2_{WGS} = N \left( \pi - \pi_0 \right)' \left( \Delta^\pi - \sum_{j=1}^J \alpha_j \pi_j \pi_j' \right)^{-1} \left( \pi - \pi_0 \right) \geq N \left( \pi - \pi_0 \right)' \left( \Delta^\pi - \sum_{j=1}^J \alpha_j \pi_j \pi_j' \right)^{-1} \left( \pi - \pi_0 \right) \quad (3.3.20)$$

Suppose the hypothesized vector $\pi_0$ is used to obtain the diagonal matrix $\Delta_{\pi_0}$ and the covariance matrix

$$\Sigma_{s0} = \left( \Delta_{\pi_0} - \sum_{j=1}^J \alpha_j \pi_j \pi_j' \right) \quad (3.3.21)$$
then a test statistic $X^2_{WGS}$ using $\hat{\Sigma}_o$ is given by

$$X^2_{WGS} = N \left( \frac{\hat{\Delta}_o}{\hat{\pi}_o} \right)' \left( \sum_{j=1}^{J} \frac{\hat{\Delta}_j^2}{\hat{\pi}_j^2} \right)^{-1} \left( \frac{\Delta_o}{\pi_o} \right)$$

(3.3.22)

which is approximately equal to $X^2_{WGS}$ when $H_o$ is true.

An upper bound for $X^2_{WGS}$ can be achieved through the covariance matrix as a sum of $J$ covariance matrices. So if $\Sigma_s$ is expressed as

$$\Sigma_s = N^{-1} \sum_{j=1}^{J} \alpha_j \left( \frac{\Delta_j}{\pi_j} - \frac{\pi_j^2}{\pi_j^2} \right)$$

(3.3.23)

then, by the following theorem, an upper bound can be constructed.

Theorem 3.7.

Suppose $B_j = \alpha_j \left( \frac{\Delta_j}{\pi_j} - \frac{\pi_j^2}{\pi_j^2} \right)$ is a positive definite matrix for $j = 1, 2, \ldots, J$; then,

$$\sum_{j=1}^{J} \chi^2 \left( \Sigma B_j \right)^{-1} \chi \leq J \sum_{j=1}^{J} \chi^2 B_j^{-1} \chi$$

where $\chi \neq 0$ is any vector.
Proof:

Since $B_j$ is assumed to be positive definite, by Theorem 3.5, \[ \sum_{j=1}^{J} B_j \] is positive definite. Consider $J = 2$ then, by Corollary 3.2.1

\[ X' (B_1 + B_2)^{-1} X = X' B_1^{-1} X - X' B_1^{-1} (B_2 + B_1)^{-1} B_1^{-1} X. \]

Since $B_1^{-1} + B_2^{-1}$ is also positive definite, then

\[ X' (B_1 + B_2)^{-1} X < X' B_1^{-1} X. \]

Similarly,

\[ X' (B_2 + B_1)^{-1} X < X' B_2^{-1} X. \]

hence,

\[ X' (B_1 + B_2)^{-1} X < 2^{-1} X' (B_1^{-1} + B_2^{-1}) X. \]

Suppose for $j = 1, 2, \ldots, J-1$ the relationship holds

\[ X' \left( \sum_{j=1}^{J} B_j \right)^{-1} X \leq (J-1)^{-1} \sum_{j=1}^{J-1} B_j^{-1} X. \]

Furthermore,

\[ \sum_{j=1}^{J} B_j^{-1} X = \sum_{j=1}^{J} (B_j + B_j')^{-1} X. \]
\[
\chi^2 \leq \chi^2'(\sum B_j)^{-1} x
\]

and

\[
\chi^2'(\sum B_j)^{-1} x \leq \chi^2 B_j^{-1} x.
\]

Consequently,

\[
\chi^2'(\sum B_j)^{-1} x \leq \sum_{j=1}^{J-1} \chi^2 B_j^{-1} x.
\]

From Theorem 3.7 follows the result that

\[
\chi^2_{WGS} \leq J^{-1} N \sum_{j=1}^{J} \alpha_j^{-1} (\hat{\Pi}_j - \Pi_0)'(\hat{\Delta}_A - \Pi_j)^{-1}(\hat{\Pi}_j - \Pi_0).
\]

\[
= J^{-1} N \sum_{j=1}^{J} \alpha_j^{-1} (\hat{\Delta}_A - \Pi_j)^{-1} \left( \frac{1}{\hat{\Pi}_j} \right) \left( \hat{\Pi}_j - \Pi_0 \right)
\]

\[
= J^{-1} \sum_{j=1}^{J} \alpha_j^{-2} \sum_{i=1}^{N_j} \frac{(\hat{\Pi}_{ij} - \Pi_{i0})^2}{\hat{\Pi}_{ij}}.
\]  

(3.3.24)

Consider the statistic

\[
\chi^2_k = J^{-2} \sum_{j=1}^{J} (\hat{\Pi}_j - \Pi_0)' \hat{\Delta}_j^{-1} (\hat{\Pi}_j - \Pi_0)
\]

where

\[
\hat{\Delta}_j = N_j^{-1} (\Delta_A - \hat{\Pi}_j \hat{\Pi}_j').
\]
An attempt is now made to show that \( x_k^2 \) is an upper bound for \( x_{WGS}^2 \) for certain types of \( B_j \) matrices. Let \( B_j = \alpha_j (\Lambda_j - \Pi_j \Pi_j') \) then,

\[
\begin{align*}
\chi'(\sum B_j)^{-1} \chi & - J^{-2} \chi' \sum B_j^{-1} \chi \\
& = \chi'[(\sum B_j)^{-1} \{I - J^{-2} (\sum B_j)(\sum B_j^{-1})\}] \chi \\
& = \chi'[(\sum B_j)^{-1} \{I - J^{-2} (\sum B_j^{-1})\}] \chi \\
& = \chi'[(\sum B_j)^{-1} \{I - J^{-1} I - J^{-2} \sum B_j \sum B_j^{-1} - J^{-2} \sum B_j \sum B_j^{-1}\}] \chi \\
& = \chi'[(\sum B_j)^{-1} \{J^{-2} \sum (I - B_k^{-1} B_j^{-1})\}] \chi \tag{3.3.26}
\end{align*}
\]

\( \leq 0 \),

if \( (I - B_k^{-1} B_j^{-1}) \) is negative definite. Note when \( B_j = B_k \)
then, \( (I - B_k^{-1} B_j^{-1}) = -I \) which is negative definite. If \( B_t \),
\( t = 1, 2, \ldots, J \); are diagonal matrices then a diagonal element of

\( (I - B_k^{-1} B_j^{-1}) \) is \( 1 - \frac{b_j}{b_k} - \frac{b_k}{b_j} < 0 \), hence, \( (I - B_k^{-1} B_j^{-1}) \)
is negative definite.

By the Multivariate Central Limit Theorem, Cramer (1946) \[ \sqrt{N} \left( \hat{\pi} - \pi_0 \right) \] converges to a normal distribution with mean vector \( \mathbf{0} \) and covariance matrix \( \Sigma_s \). Then, the asymptotic distribution of \( \chi^2_{WGS} \) is a chi-square with \((I-1)\) degrees of freedom.

Consider testing the homogeneity hypothesis, \( H_0 : \pi_j = \pi_0 \), where \( \pi_0 \) is a known vector for this stratified sampling scheme, then the Wald Statistic

\[
\chi^2_{WHS} = \sum_{j=1}^{J} (\hat{\pi}_j - \pi_0)^\top \Sigma_j^{-1} (\hat{\pi}_j - \pi_0) \quad (3.3.27)
\]

is equivalent to the Pearson Statistic

\[
\chi^2_H = \sum_{j=1}^{J} N_j \sum_{i=1}^{I} (\hat{\pi}_{ij} - \pi_{0i})^2 / \pi_{0i} \quad (3.3.28)
\]

where

\[
\Sigma_j = \left( \Delta_j N_j \right)^{-1} N_j \left( \pi_j - \pi_0 \right) \left( \pi_j - \pi_0 \right)^\top / N_j \quad (3.3.29)
\]

Consider testing the hypothesis of independence, \( H_0 : \pi_j = \pi_0 \), where \( \pi_0 \) is an unknown vector for this stratified sampling design, then a test statistic can be obtained which is based on estimators for the vectors \( \pi_j - \pi_0 \) and a consistent estimator of the corresponding covariance matrix. It will be shown that this Wald Statistic
\[ x^2_{\text{WIS}} = \sum_{J} \left( \frac{\hat{\pi}(J)}{\hat{\pi}_0} - \frac{\pi(J)}{\pi_0} \right)^2 \frac{\pi_0}{\pi_0} \]  

(3.3.30)

is equivalent to the usual Pearson statistic

\[ x^2 = \sum_{J} \left( \sum_{i=1}^{I} \left( \frac{\pi_{ij} - \hat{\pi}_{ij}}{\hat{\pi}_{i0}} \right)^2 \frac{\pi_{i0}}{\pi_{i0}} \right), \]

(3.3.31)

where

\[ \hat{\pi}(J) \hat{\pi}(J) = \left( \hat{\pi}_1 - \hat{\pi}_0, \hat{\pi}_2 - \hat{\pi}_0, \ldots, \hat{\pi}_{J-1} - \hat{\pi}_0 \right)^T, \]

(3.3.32)

and \( V \) is the covariance matrix for \( \hat{\pi}(J) \hat{\pi}(J) \). Here \( \hat{\pi}_0 \) is an estimator for \( \pi_0 \) where \( \sum \alpha_j = 1 \) and \( \alpha_j > 0 \) for \( j = 1, 2, \ldots, J \).

Note that for any such set of weights

\[ \pi_0 = \sum_{J} \alpha_j \pi_j, \]

(3.3.33)

when the null hypothesis is true. Consider the vector \( \hat{\pi}_j - \hat{\pi}_0 \) as an estimator of \( \pi_j - \pi_0 \), then

\[ E(\hat{\pi}_j - \hat{\pi}_0) = \pi_j - \pi_0, \]

(3.3.34)
Each subvector of the vector $\mathbf{\hat{N}}_{\mathbf{J}}$, say $\mathbf{\hat{N}}_{\mathbf{J}}$, has covariance matrix given by

$$\text{var}(\mathbf{\hat{N}}_{\mathbf{J}}) = V_j - 2\alpha_j V_j + \sum_{i=1}^{J} \Sigma_j^2 V_j$$  \hspace{1cm} (3.3.36)

where $V_j = \text{var}(\mathbf{\hat{N}}_{\mathbf{J}})$ is given by

$$V_j = \mathbf{N}_j^{-1} (\mathbf{\Delta}_{\mathbf{J}} - \mathbf{\Pi}_j \mathbf{\Pi}_j')$$  \hspace{1cm} (3.3.37)

The covariance matrix between any two subvectors is given by

$$\text{Cov}(\mathbf{\hat{N}}_{\mathbf{J}}, \mathbf{\hat{N}}_{\mathbf{J}}') = -\alpha_j V_j - \alpha_j V_j + \sum_{c=1}^{J} \Sigma_j^2 V_j$$  \hspace{1cm} (3.3.38)

Consider $J = 3$, then the covariance matrix $V$, for $\mathbf{\hat{N}}_{\mathbf{J}}$ has dimension $2(I-1)$, since the difference corresponding to the third stratum has been deleted and the last category has been deleted from $\mathbf{\hat{N}}_{\mathbf{J}}$ and $\mathbf{\hat{N}}_{\mathbf{J}}$. The covariance matrix can be written as


\[ V = \begin{bmatrix}
(1-2\alpha_1)V_1 + \sum_{c=1}^{3} \alpha_c^2 V_c & -\alpha_1 V_1 - \alpha_2 V_2 + \sum_{c=1}^{3} \alpha_c^2 V_c \\
-\alpha_1 V_1 - \alpha_2 V_2 + \sum_{c=1}^{3} \alpha_c^2 V_c & (1-2\alpha_2)V_2 + \sum_{c=1}^{3} \alpha_c^2 V_c
\end{bmatrix} \]  

(3.3.39)

Under \( H_0 \), the matrix

\[ V_j = N_j^{-1} \left( \frac{\Delta_j}{\Pi_0} - \Pi_0 \Pi' \right) \]

\[ = N_j^{-1} B_0. \]  

(3.3.40)

Then, \( V_\infty \), which is the value of \( V \) under \( H_0 \), is given by

\[ V_\infty = \begin{bmatrix}
(1-2\alpha_1)N_1^{-1} B_0 + \sum_{c=1}^{3} \alpha_c N_1^{-1} B_0 & -(\alpha_1 N_1^{-1} + \alpha_2 N_2^{-1}) B_0 + \sum_{c=1}^{3} \alpha_c N_1^{-1} B_0 \\
-(\alpha_1 N_1^{-1} + \alpha_2 N_2^{-1}) B_0 + \sum_{c=1}^{3} \alpha_c N_1^{-1} B_0 & (1-2\alpha_2)N_2^{-1} B_0 + \sum_{c=1}^{3} \alpha_c N_1^{-1} B_0
\end{bmatrix} \]

(3.3.41)

Suppose \( \alpha_j \)'s are chosen such that

\[ \alpha_j = N_j / N. \]  

(3.3.42)

then,
\[ V_0 = \begin{pmatrix} (N_1^{-1} - N_2^{-1})B_0 & -N_1^{-1}B_0 \\ -N_1^{-1}B_0 & (N_2^{-1} - N_1^{-1})B_0 \end{pmatrix} \]  \hspace{1cm} (3.3.43) \\

where \[ J = \sum_{j=1}^{J} N_j. \]  \hspace{1cm} (3.3.44) \\

\[ V_0 \] can be further expressed as \[ V_0 = \begin{pmatrix} N_1^{-1} & 0 \\ 0 & N_2^{-1} \end{pmatrix} - N_1^{-1}J \otimes B_0. \]  \hspace{1cm} (3.3.45) \\

where \( J \) is a square matrix with entries all equal to one. Then, by Theorem 3.1 
\[ V^{-1} = \begin{pmatrix} N_1 & 0 \\ 0 & N_2 \end{pmatrix} + N_3^{-1} \begin{pmatrix} N_1^2 & N_1N_2 \\ N_1N_2 & N_2^2 \end{pmatrix} \otimes B_0^{-1} \]
\[ = \begin{pmatrix} \left( N_1 + \frac{N_2N_3^{-1}}{N_1N_3} \right)B_0^{-1} & N_3^{-1}N_1N_2B_0^{-1} \\ N_3^{-1}N_1N_2B_0^{-1} & \left( N_2 + \frac{N_2N_3^{-1}}{N_1N_3} \right)B_0^{-1} \end{pmatrix} \]  \hspace{1cm} (3.3.46)
Hence, a test statistic $\chi^2_{WSI}$ for testing

$$H_0: \pi_j = \pi_o \quad j = 1, 2, 3; \quad (3.3.47)$$

where $\pi_o$ is an unknown vector, is given by

$$\chi^2_{WSI} = (\hat{\pi}(J) - \pi_o)(\hat{\pi}(J) - \pi_o)'V_{\pi o}(\hat{\pi}(J) - \pi_o), \quad (3.3.48)$$

where $\hat{V}_{\pi o}$ is a consistent estimator of $V_{\pi o}$. The covariance matrix $V_{\pi o}$ can be obtained by replacing $\pi_o$ with $\hat{\pi}_o$ in $B_o$. The statistic $\chi^2_{WSI}$ reduces to

$$\chi^2_{WSI} = \sum_{j=1}^{2} N_j (\hat{\pi}_j - \pi_o')(B_o(\hat{\pi}_j - \pi_o) + N_3^{-1}[N_3 (\hat{\pi}_3 - \pi_o)(B_o N_3 (\hat{\pi}_3 - \pi_o))], \quad (3.3.49)$$

since

$$\sum_{j=1}^{2} N_j (\hat{\pi}_j - \pi_o) = N_3 (\hat{\pi}_3 - \pi_o). \quad (3.3.50)$$

So

$$\chi^2_{WSI} = \sum_{j=1}^{J} N_j (\hat{\pi}_j - \pi_o)'B_o^{-1}(\hat{\pi}_j - \pi_o) \quad (3.3.51)$$

where

$$B_o^{-1} = (\Delta_{\pi_o} - \hat{\pi}_o \hat{\pi}_o'\pi_o)^{-1}. \quad (3.3.52)$$
Then, by Theorem 3.1

\[ X^2_{WSI} = \sum_{j=1}^{3} N_j \sum_{i=1}^{3} \left( \frac{\hat{\pi}_{ij} - \pi_{io}}{\pi_{io}} \right)^2. \]  

(3.3.53)

This result can be generalized to any number of strata.

The consistent estimator used for the covariance matrix in \( X^2_{WSI} \) was obtained by substituting \( \hat{\pi}_o \) for \( \pi_o \). This substitution depends on \( \hat{\pi}_j \) being a consistent estimator for \( \pi_j \), and this is the case when \( N_j \rightarrow \infty \). Therefore, under the assumption that \( N_j \rightarrow \infty \) and by use of the Multivariate Central Limit Theorem \( X^2_{WSI} \) converges to a chi-square random variable with \( (I-1)(J-1) \) degrees of freedom.

SUPERCARP, a computer program, was developed in the Survey Section of the Statistical Laboratory at Iowa State University. The program calculates variance estimates for regression estimators and estimated variances for common survey estimators among other statistics. The test of the hypothesis that a set of population proportions is equal to a given set of numbers can be done using SUPERCARP. Let

\[ \pi_o = (\pi_{01}, \pi_{02}, \ldots, \pi_{0I})' \]  

(3.3.54)

be the vector of hypothesized proportions for the \( I \) categories of the population. Let

\[ \hat{\pi} = (\hat{\pi}_{10}, \hat{\pi}_{20}, \ldots, \hat{\pi}_{IO})' \]  

(3.3.55)

be the vector of estimated proportions. Let \( \hat{\Sigma} \) be the \((I-1) \times (I-1)\)
estimated covariance matrix of the first $I-1$ estimated proportions. Then, Hidiroglou, Fuller and Hickman (1980) constructed a test of the hypothesis

$$H_0: \pi = \pi_0,$$

where $\pi_0$ is a known vector, as

$$F = [(I-1)D]^{-1}(D-I+2)(\pi - \pi_0)' \sum^{-1}(\pi - \pi_0),$$

(3.3.56)

where $\pi$ and $\pi_0$ are of dimension $I-1$.

$$D = \sum_{j=1}^{J} (N_j - 1),$$

(3.3.57)

$J$ is the number of strata in the sample and $N_j$ is the number of primary sampling units in the $j^{th}$ stratum. Hidiroglou et al. report that in large samples $F$ is approximately distributed as a central $F$ with $I-1$ and $D-I+2$ degrees of freedom when the null hypothesis is true. The Taylor approximation is used to obtain covariance matrices. An example beyond the multinomial for which the program is appropriate, stated in Hidiroglou et al. (1980), is the testing of the hypotheses about the fraction of earned income received by certain age categories of the population.
Example 3.1.

In this example, some of the results obtained in this section are applied to data taken from Sewell and Shah (1968) and given in Appendix A, Table 7.11. The example is concerned with the status for some Wisconsin high school senior boys and girls. There are four levels of status which are treated as strata. Each stratum has eight categories. The four strata are i) low economic level, ii) lower middle level, iii) middle level, and iv) high level. The eight categories are the cross classification of four levels of IQ and two levels of college plans. The four IQ levels are low, lower middle, upper middle, and high. There are two cases for college plans, plans to go and plans not to go. The strata sizes are

\[ \mathbf{N} = (1150, 1298, 1298, 1245)' \]

The hypothesis of interest is

\[ H_0 : \mathbf{\Pi} = \mathbf{\Pi}_0 , \]

where \( \mathbf{\Pi}_0 \) is known. For the purpose of this analysis, \( \mathbf{\Pi}_0 \) was arbitrarily chosen for two such cases. First consider a hypothesized value

\[ \mathbf{\Pi}_0 = (.0338, .2256, .0526, .2068, .1128, .1429, .1504)' , \]

with the 8th category left off. The estimated vector \( \mathbf{\hat{\Pi}} \) has a value of
\( \hat{\Pi} = (0.0289, 0.2186, 0.0843, 0.1888, 0.1108, 0.1373, 0.1645)' \).

The covariance matrix for the estimated vector is

\[
\hat{\Sigma}_S = \begin{bmatrix}
0.02791 & -0.00555 & -0.00286 & -0.00505 & -0.00363 & -0.00382 & -0.00544 \\
-0.00555 & 0.16150 & -0.01579 & -0.04595 & -0.01920 & -0.03246 & -0.02610 \\
-0.00286 & -0.01579 & 0.07563 & -0.01450 & -0.01090 & -0.01110 & -0.01628 \\
-0.00505 & -0.04595 & -0.01450 & 0.15047 & -0.01822 & -0.02736 & -0.02563 \\
-0.00363 & -0.01920 & -0.01090 & -0.01822 & 0.09572 & -0.01379 & -0.02374 \\
-0.00382 & -0.03246 & -0.01110 & -0.02736 & -0.01379 & 0.11749 & -0.01944 \\
-0.00544 & -0.02610 & -0.01628 & -0.02563 & -0.02374 & -0.01944 & 0.12595 \\
\end{bmatrix}
\]

where \( \hat{\Sigma}_S \) is defined in (3.3.19), and

\[
\Sigma_m = (\hat{\Pi}^\top \hat{\Pi} - \bar{\Pi} \bar{\Pi}^\top) / N
\]

has a numerical value of

\[
\Sigma_m = \begin{bmatrix}
0.03266 & -0.00763 & -0.0178 & -0.00699 & -0.00381 & -0.00483 & -0.00508 \\
-0.00763 & 0.17471 & -0.01187 & -0.04665 & -0.02545 & -0.03224 & -0.03393 \\
-0.00178 & -0.01187 & 0.04983 & -0.01088 & -0.00593 & -0.00752 & -0.00791 \\
-0.00699 & -0.04665 & -0.01088 & 0.16403 & -0.02333 & -0.02955 & -0.03110 \\
-0.00381 & -0.02545 & -0.00593 & -0.02333 & 0.10008 & -0.01612 & -0.01697 \\
-0.00483 & -0.03224 & -0.00752 & -0.02955 & -0.01612 & 0.12248 & -0.02149 \\
-0.00508 & -0.03393 & -0.00791 & -0.03110 & -0.01697 & -0.02149 & 0.12778 \\
\end{bmatrix}
\]

The test statistic \( X^2 \), defined in (3.2.14), is
\[ x^2_G = \sum_{i=1}^{I} \frac{(\hat{\pi}_i - \pi_{i0})^2}{\pi_{i0}} \]

\[ = 86.08 \]

where \( N = 4991 \) is the total sample size.

The Wald test statistic given in (3.3.18) is

\[ x^2_{WGS} = N \left( \hat{\pi} - \pi_{\infty} \right)' \hat{\Sigma}^{-1} \left( \hat{\pi} - \pi_{\infty} \right) \]

\[ = 91.80 \]

where \( \hat{\Sigma} \) is given above using (3.3.23). The strata probability vectors are

\[ \hat{\pi}_1 = (0.01478, 0.35913, 0.03652, 0.2426, 0.04348, 0.1565, 0.0513)' \]

\[ \hat{\pi}_2 = (0.02234, 0.24345, 0.054699, 0.22804, 0.080894, 0.15948, 0.10478)' \]

\[ \hat{\pi}_3 = (0.04237, 0.19799, 0.13868, 0.17720, 0.12712, 0.14792, 0.15717)' \]

and

\[ \hat{\pi}_4 = (0.03454, 0.08434, 0.1028, 0.1100, 0.18715, 0.085141, 0.33896)' \]

To illustrate the computations an alternative hypothesis was tested. This time the hypothesized value for \( \pi_{\infty} \) was selected as

\[ \pi_{\infty} = (0.025, 0.220, 0.080, 0.185, 0.115, 0.140, 0.160, 0.075)' \]
The values for $\frac{N_j}{n_j}$, $N_j$, $N$ for $j = 1,2,3,4$; are the same as before, but

$$x_G^2 = 10.68$$

and

$$x_{WGS}^2 = 11.14.$$  

The statistic $x_{WGS}^2$ is asymptotically distributed as a chi-square random variable with seven degrees of freedom.

In the first case, the hypothesis $H_0: \Pi = \Pi_0$ was rejected when compared to a chi-square with seven degrees of freedom. This was the case no matter whether $x_G^2 = 86.08$ or $x_{WGS}^2 = 91.80$ was considered as the test statistic. Whereas, in the second case, the hypothesis $H_0: \Pi = \Pi_0$ when tested using $x_G^2$ or $x_{WGS}^2$ had test statistic values less than the ninety fifth percentile of a chi-square distribution with seven degrees of freedom.

An interesting result from these test statistic values is the fact that under $H_0$ the numerical value of $x_{WGS}^2$ is greater than the numerical value of $x_G^2$. This is primarily due to the larger values on the diagonal in the covariance matrix for $\frac{N_j}{n_j}$ under multinomial sampling as opposed to stratified sampling.

3.4. Two-Stage Sampling

Thus far, samples of units from the entire population or from strata within the population have been considered. However, these
types of sampling schemes are not always appropriate. A commonly used alternative involves cluster sampling or combinations of stratification and cluster sampling. The following discussion will concentrate on cluster sampling. An example of this type of sampling is given in Brier (1980) pertaining to the manner in which people in Minnesota perceive the quality of their housing and their community's housing. The variables of interest in that survey are the family's opinion of their own home (personal satisfaction) and their opinion of the housing in the community as a whole (community satisfaction). In particular, the question of interest is whether a family's classification according to level of personal satisfaction is independent of its classification by level of community satisfaction.

In each community, 5 homes were randomly selected and the family was questioned in two areas: satisfaction with the housing in the community as a whole and satisfaction with their own home. The groups of 5 homes are the clusters. There are a total of 40 clusters, 20 in the metropolitan area and 20 in the outlying region.

The discussion that follows examined two-stage cluster sampling since one-stage sampling is a special case of two-stage sampling. Subsequently, consideration is given to a two-stage sampling scheme for strata within a population. A two-stage sampling scheme performed in several strata is called a stratified two-stage sampling scheme. The results obtained from the stratified two-stage sampling scheme are used to analyze some data obtained from wild turkeys in Iowa and another for the styles of Greek authors.
Consider a system consisting of $s$ sampling units (psu). Select a random sample of $S$ primary units from the $s$ psu's. The psu's are assumed to be chosen with replacement proportional to size. From each of the sampled psu's obtain a random sample of size $n_k$ with replacement, $k = 1, 2, ..., S$. The sample of size $n_k$ is classified into $I$ different categories. So $n_k$ is the total sample size from the $I$ different categories. Let $x_{ik}$ denote the number of units observed in the $i^{th}$ category when the $k^{th}$ sampled primary unit is observed. Define the vector of category counts for the $k^{th}$ sampled primary unit as

$$\mathbf{x}_k = (x_{1k}, x_{2k}, \ldots, x_{Ik})'$$

Then,

$$n_k = \sum_{i=1}^{I} x_{ik} \quad (3.4.1)$$

is the total sample size for the $k^{th}$ sampled psu and

$$N = \sum_{k=1}^{S} n_k \quad (3.4.2)$$

is the total sample size obtained from all $S$ psu's in the sample.

Define $x_{i+}$ as the total count for the $i^{th}$ category over all $S$ sampled units, then
Now consider sampling from \( J \) strata. Define the vector of proportions for the \( j \)th stratum as \( \pi_j \) such that

\[
\pi_j = \sum_{k=1}^{S} \alpha_{jk} \pi_{jk}
\]

(3.4.4)

where \( \alpha_{jk} \) denotes the relative size of the \( k \)th primary unit and the vector \( \pi_{jk} \) is a probability vector for the \( k \)th primary unit. Then,

\[
\pi_{jk} = (p_{1jk}, p_{2jk}, \ldots, p_{ljk})'.
\]

(3.4.5)

Let \( \hat{\pi}_{jk} \) be the vector of observed proportions for the \( k \)th sampled primary unit. As a consequence of simple random sampling within clusters, \( x_{jk} \) is distributed as a multinomial random vector with sample size \( n_{jk} \). The estimated probability vector is

\[
\hat{\pi}_{jk} = \left( \frac{x_{1jk}}{n_{jk}}, \frac{x_{2jk}}{n_{jk}}, \ldots, \frac{x_{ljk}}{n_{jk}} \right)'.
\]

Define the probability vector \( \hat{\pi}_j \) as an estimator for \( \pi_j \) such that

\[
\hat{\pi}_j = N_j^{-1} \bar{x}_j
\]

(3.4.6)

where \( \bar{x}_j = (x_{1j+}, x_{2j+}, \ldots, x_{lj+})' \).
Then, the first moment of $\frac{\hat{\mu}_j}{\mu_j}$ is given by

$$E(\frac{\hat{\mu}_j}{\mu_j}) = E_p[N_j^{-1} \sum_{k=1}^{S_j} E_s(x_{jk})]$$  (3.4.7)

$$E = E_p[N_j^{-1} \sum_{k=1}^{S_j} N_j p_{jk}]$$

where $E_p$ denotes expectation with respect to the possible samples of primary units and $E_s$ denotes the conditional expectation with respect to the secondary units within a given primary unit. So

$$E(\frac{\hat{\mu}_j}{\mu_j}) = N_j^{-1} \sum_{k=1}^{S_j} \left( n_{jk} \sum_{l=1}^{\pi_j} c_{jl} p_{jl} \right)$$

$$= N_j^{-1} \sum_{k=1}^{S_j} n_{jk} \frac{\hat{\mu}_j}{\mu_j}$$

$$= \frac{\hat{\mu}_j}{\mu_j}. \quad (3.4.8)$$

Denote the variance of this unbiased estimate, $\frac{\hat{\mu}_j}{\mu_j}$, by

$$\Sigma_{2s}(j) = \text{var}(\frac{\hat{\mu}_j}{\mu_j})$$

$$= \text{var}(N_j^{-1} x_j)$$
\[
\begin{align*}
\text{where } V_p & \text{ denotes the variance with respect to the possible samples of primary units. Then,} \\
\Sigma_{2s}(j) &= N_j^{-1} \left[ \sum_{\ell=1}^{S_j} \alpha_{j\ell} (\Delta_{j\ell} - \Pi_{j\ell} \Pi_{j\ell}') \right] + \frac{S_j}{N_j} \left( \sum_{j=1}^{N_j^2} \frac{\hat{s}_{j}}{n_j^2} \right) \\
&= N_j^{-1} (\Delta_{j\ell} - \Pi_{j\ell} \Pi_{j\ell}') + N_j^{-1} (\Pi_{j\ell} \Pi_{j\ell}' - \sum_{\ell=1}^{S_j} \alpha_{j\ell} \Pi_{j\ell} \Pi_{j\ell}') \\
&\quad + \frac{1}{N_j^2} \left( \sum_{j=1}^{N_j^2} \frac{\hat{s}_{j}}{n_j^2} \right) \sum_{\ell=1}^{S_j} \alpha_{j\ell} (\Pi_{j\ell} \Pi_{j\ell}' - \Pi_{j\ell} \Pi_{j\ell}')'.
\end{align*}
\]

When the primary units are of equal sample size then, \( S_j n_{jk} = N_j \) for \( k = 1, 2, \ldots, S_j \); and

\[
\Sigma_{2s}(j) = N_j^{-1} \left[ \sum_{\ell=1}^{S_j} \alpha_{j\ell} (\Delta_{j\ell} - \Pi_{j\ell} \Pi_{j\ell}') \right] + S_j^{-1} \left[ \sum_{\ell=1}^{S_j} \alpha_{j\ell} (\Pi_{j\ell} \Pi_{j\ell}' - \Pi_{j\ell} \Pi_{j\ell}')' \right].
\]
The covariance matrix, $\Sigma_{2s}(j)$ is a combination of the variance of $\tilde{N}_j$ under a multinomial sampling scheme and a measure of the differences between the primary units. When the probability vectors for the primary units within the $j^{th}$ stratum are the same then, $\Sigma_{2s}(j)$ is equal to

$$\Sigma_m(j) = N_j^{-1} \left[ \Delta_j - \tilde{N}_j \tilde{N}_j' \right],$$

(3.4.12)

the variance of $\tilde{N}_j$ under the assumption of a multinomial sampling scheme, where $\Delta_j$ is the diagonal matrix with diagonal elements corresponding to the elements of $\tilde{N}_j$.

Denote the difference between the covariance matrices, $\Sigma_{2s}(j)$ and $\Sigma_m(j)$ by

$$R_j = \Sigma_{2s}(j) - \Sigma_m(j)$$

$$= N_j^{-1} \left( \Delta_j - \tilde{N}_j \tilde{N}_j' \right) + N_j^{-1} \left( \tilde{N}_j \tilde{N}_j' - \sum_{k=1}^{n_j} \alpha_j \tilde{N}_j \tilde{N}_j' \right)$$

(3.4.13)

which is referred to here as the correction factor for two-stage sampling scheme. Hence,
\[ \Sigma_{2s}(j) = \Sigma_m(j) + R_j . \] (3.4.14)

When the primary units are of equal size then,

\[ R_j = (S_j^{-1} - N_j^{-1}) \left[ \sum_{k=1}^{\phi_j} \alpha_{jk} (P_{jk} - \bar{P}_j) (P_{jk} - \bar{P}_j)' \right] . \] (3.4.15)

Consider writing

\[ \Sigma_{2s}(j) = \Sigma_m(j) + t_j B_j \] (3.4.16)

where

\[ t_j = \left\{ \frac{S_j}{(\sum_{k=1}^{\phi_j} N_{jk}^2)/N_j^2 - 1/N_j} \right\} \] (3.4.17)

and

\[ B_j = \sum_{k=1}^{\phi_j} \alpha_{jk} (P_{jk} - \bar{P}_j) (P_{jk} - \bar{P}_j)' , \] (3.4.18)

then by Corollary 3.2.1, the inverse of \( \Sigma_{2s}(j) \) is

\[ \Sigma_{2s}^{-1}(j) = \Sigma_m^{-1}(j) - \Sigma_m^{-1}(j)(R_j^{-1} + \Sigma_m^{-1}(j))^{-1} \Sigma_m^{-1}(j) \] (3.4.19)

where \( R_j \) is assumed to be nonsingular, which implies that the number of primary units exceeds the number of categories. Then, for any vector \( \tilde{x} \neq 0 \)

\[ \tilde{x}' \Sigma_{2s}^{-1}(j) \tilde{x} \leq \tilde{x}' \Sigma_m^{-1}(j) \tilde{x} , \] (3.4.20)
since \( \Sigma_m^{-1}(j) + R_j^{-1} \) is nonnegative definite so that

\[
x^t \Sigma_m^{-1}(j) (R_j^{-1} + \Sigma_m^{-1}(j))^{-1} \Sigma_m^{-1}(j)x \geq 0
\]  
(3.4.21)

Applying Corollary 3.2.1 again to expression (3.4.19) gives

\[
\Sigma_{2s}^{-1}(j) = \Sigma_m^{-1}(j) - \Sigma_m^{-1}(j) R_j \Sigma_m^{-1}(j) + \Sigma_m^{-1}(j) R_j (\Sigma_m^{-1}(j) + R_j)^{-1} R_j \Sigma_m^{-1}(j).
\]  
(3.4.22)

Therefore,

\[
x^t \Sigma_{2s}^{-1}(j)x \geq x^t \Sigma_m^{-1}(j)x - x^t \Sigma_m^{-1}(j) R_j \Sigma_m^{-1}(j)x,
\]  
(3.4.23)

and

\[
x^t \Sigma_m^{-1}(j) R_j \Sigma_m^{-1}(j)x = x^t N_j^2 t_j (\Delta^{-1} + \frac{1}{\pi_{ij}} J) B_j (\Delta^{-1} + \frac{1}{\pi_{ij}} J)x
\]  
(3.4.24)

where

\[
\pi_{ij} = 1 - \sum_{i=1}^{I} \pi_{ij}.
\]  
(3.4.25)

Define

\[
C_j = x^t \Sigma_m^{-1}(j) R_j \Sigma_m^{-1}(j)x
\]  
(3.4.26)
\[ m_j = (\Delta^{-1} + \frac{1}{\pi_j}) \Delta_{j} \]

\[ = (m_{1j}, m_{2j}, \ldots, m_{I-1j})' \]  \hspace{1cm} (3.4.27)

then,

\[ C_j = (\Sigma n_j^2 - N_j)m_Bm^j \]  \hspace{1cm} (3.4.28)

Now, consider testing the homogeneity hypothesis

\[ H_0: \tau_j = \tau_o \quad j = 1, 2, \ldots, J; \]  \hspace{1cm} (3.4.29)

under a two-stage sampling scheme when \( \tau_o \) is a known vector. A Wald Statistic, of the type that has been constructed thus far for testing \( H_0: \tau_j = \tau_o \) is

\[ X^2_{WH2S} = \sum_{j=1}^{J} \left( \frac{\hat{\tau}_j - \tau_o}{\hat{\Sigma}_{2s}(j)} \right)' \hat{\Sigma}^{-1}(j)(\frac{\hat{\tau}_j - \tau_o}{\hat{\Sigma}_{2s}(j)}) \]  \hspace{1cm} (3.4.30)

where \( \hat{\Sigma}_{2s}(j) \) is a consistent estimator for \( \Sigma_{2s}(j) \). One such estimator can be obtained by substituting \( \hat{r}_{jkl} \) for \( r_{jkl} \) and \( \tau_o \) for \( \tau_j \) in (3.4.11), where \( \hat{r}_{jkl} = \frac{1}{n_{jkl}} \hat{x}_{jkl} \). By the Multivariate Central Limit Theorem, as the primary unit size \( n_{jkl} \to \infty \) and \( S_j \to \mathbb{S}_j \), then \( X^2_{WH2S} \) has an asymptotic chi-square distribution with \( I-1 \) degrees of freedom.
An upper bound for $X^2_{\text{WH2S}}$ is the Pearson goodness-of-fit statistic

$$X^2_{\text{H}} = \sum_{j=1}^{J} \sum_{i=1}^{N_j} \frac{(\pi_{ij} - \pi_{io})^2}{\pi_{io}}$$  \hspace{1cm} (3.4.31)$$

for the multinomial model which corresponds to simple random sampling from the entire population. Such a result follows from expression (3.4.21).

It can be shown that

$$X^2_{\text{WH2S}} = X^2_{\text{WH2S}} - \hat{C}_j,$$  \hspace{1cm} (3.4.32)$$

where $X^2_{\text{WH2S}}$ is given in (3.4.30), is a lower bound for $X^2_{\text{WH2S}}$ by considering Theorems 3.2 and 3.7. The formula for $\hat{C}_j$ under $H_0$ is

$$\hat{C}_j = \left( \sum_{k=1}^{S_j} \frac{n_{jk}^2}{N_j} \right) \left( \frac{\pi_{j.o} - \pi_{.o}}{\pi_{.o}} \right)^T \left( \frac{1}{\pi_{.o}} + \frac{1}{\pi_{j.o}} \right) \left[ \sum_{l=1}^{S_j} \alpha_{j.l} \left( \hat{P}_{j.l} - \pi_{j.o} \right)^T \left( \hat{P}_{j.l} - \pi_{j.o} \right) \right]$$

$$\left( \frac{1}{\pi_{.o}} + \frac{1}{\pi_{j.o}} \right) \left( \frac{1}{\pi_{j.o}} - \pi_{j.o} \right).$$  \hspace{1cm} (3.4.33)$$

By the results of Theorems 3.1, 3.2, 3.3, and 3.4, $\left( \frac{1}{\pi_{.o}} + \frac{1}{\pi_{j.o}} \right)$ is a positive definite matrix and $\hat{C}_j$ is nonnegative.

To test the goodness-of-fit hypothesis $H_0: \pi = \pi_0$, a Wald statistic is

$$X^2_{\text{WG2S}} = \left( \pi - \pi_0 \right)^T \left( \frac{\Delta^{-1}}{\pi_0} + \frac{1}{\pi_{j.o}} \right) \left( \pi - \pi_0 \right)$$  \hspace{1cm} (3.4.34)$$
where $\hat{\Sigma}_{2s}$ is a consistent estimator for $\Sigma_{2s}$, the covariance matrix for $\hat{\pi}$. This estimator for $\Sigma_{2s}$ is given by

$$\hat{\Sigma}_{2s} = \sum_{j=1}^{J} \alpha_j^2 \hat{\Sigma}_{2s}(j)$$  \hspace{1cm} (3.4.35)

since the $J$ strata are independent, where

$$\alpha_j = \frac{N_j}{N}$$

and

$$\hat{\Sigma}_{2s}(j) = N_j^{-1} \left[ \sum_{\ell=1}^{S_j} \alpha_{j\ell} \left( \hat{\Delta}_j - \hat{\pi}_j \hat{\pi}_j' \right) \right] + \sum_{k=1}^{S_j} \frac{n_{jk}}{N_j} \left[ \sum_{\ell=1}^{S_j} \alpha_{j\ell} \left( \hat{\pi}_{j\ell} \hat{\pi}_{j\ell}' \right) \right]$$

$$= N_j^{-1} [\hat{\Delta}_j - \hat{\pi}_j \hat{\pi}_j'] + \frac{\sum n_{jk}}{N_j} \sum_{\ell=1}^{S_j} \alpha_{j\ell} (\hat{\pi}_{j\ell} \hat{\pi}_{j\ell}') \hspace{1cm} (3.4.36)$$

Similar to the derivation for the upper bound for the statistic $X^2_{\text{WH2S}}$, an upper bound for $X^2_{\text{WG2S}}$ is

$$X^2_G = N \sum_{i=1}^{I} \left( \frac{\pi_i - \pi_{io}}{\bar{\pi}_{io}} \right)^2 \hspace{1cm} (3.4.37)$$

the goodness-of-fit statistic for the multinomial model, which corresponds to simple random sampling from the entire population.
Consider again the two-stage sampling scheme in \( J \) different strata, \( j = 1, 2, \ldots, J \). Such a scheme is referred to as a stratified two-stage sampling scheme. Consider testing \( H_0: \pi_j = \pi_0 \) for some unknown vector \( \pi_0 \) and \( j = 1, 2, \ldots, J \). Define

\[
\hat{\pi}_0 = \frac{1}{J} \sum_{j=1}^{J} \hat{\pi}_j
\]

as an unbiased estimator for

\[
\pi_0 = \frac{1}{J} \sum_{j=1}^{J} \pi_j.
\]

The variance-covariance matrix for \( \hat{\pi}_j - \pi_0 \) is

\[
T_{jj} = \text{var}(\hat{\pi}_j - \pi_0) = \alpha_j^2 A_j - 2\alpha_j A_j + \sum_{\ell=1}^{J} \alpha_{\ell}^2 A_{\ell}
\]

(3.4.38)

and the matrix of covariances between \( \hat{\pi}_j \) and \( \hat{\pi}_j \) is

\[
T_{jj} = \text{cov}(\hat{\pi}_j - \pi_0, \hat{\pi}_j - \pi_0) = \alpha_j^2 A_j - \alpha_j A_j + \sum_{\ell=1}^{J} \alpha_{\ell}^2 A_{\ell},
\]

(3.4.39)

where
\[ A_j = \sum_m(j) + \left( \frac{E_{jk}}{N_j^2} - \frac{1}{N_j} \right) \sum_{\ell} \alpha_{j\ell} (p_{j\ell} - \pi_j)(p_{j\ell} - \pi_j)' \]  

and

\[ \Sigma_m(j) = N_j^{-1} (\Delta_j \pi_j \pi_j') \]  

Let \( \Sigma \) be the covariance matrix for

\[ \left( \hat{\Sigma}^{(J)}_{(J)} \right) = \left( \hat{\Sigma}_{11}^{(J)} - \hat{\Sigma}_{12}^{(J)} \hat{\Sigma}_{22}^{(J)}, \ldots, \hat{\Sigma}_{J-1,J}^{(J)} - \hat{\Sigma}_{J,J}^{(J)} \right)' \]  

then, \( \Sigma \) has \( T_{jj} \) as off diagonal blocks and \( T_{jj} \) as diagonal blocks. The \( J \)th subpopulation is left off of (3.4.42) to obtain a nonsingular covariance matrix. Let \( \Sigma_0 \) denote the value of \( \Sigma \) under \( H_0 \) and \( S_{oj} \) the value of \( S_j \) under \( H_0 \). Consider \( J = 3 \), then under \( H_0 \)

\[ \Sigma_0 = \begin{bmatrix} (1-2\alpha_1)A_{o1} + \sum_{\ell=1}^{3} \alpha_{o2}^2 A_{o1} & -\alpha_{o2} A_{o0} + \sum_{\ell=1}^{3} \alpha_{o2}^2 A_{o0} \\ -\alpha_{o1} A_{o0} - \alpha_{o2} A_{o2} + \sum_{\ell=1}^{3} \alpha_{o2} A_{o0} & (1-2\alpha_2)A_{o2} + \sum_{\ell=1}^{3} \alpha_{o2}^2 A_{o2} \end{bmatrix} \]  

where

\[ A_{oj} = N_j^{-1} (\Delta_j \pi_j \pi_j') + \left( \frac{E_{jk}}{N_j^2} - \frac{1}{N_j} \right) \sum_{\ell} \alpha_{j\ell} (p_{j\ell} - \pi_j)(p_{j\ell} - \pi_j)' \]
When the sample sizes for each cluster within a stratum, \( n_{jk} \), are equal, then

\[
N^{-1} = \frac{\sum_{jk}^{2}}{N_{jk}} \times \frac{\sum_{jk}^{2}}{N_{jk}} + \left( \frac{\sum_{jk}^{2}}{N_{jk}} - 1 \right) \sum_{jk} \mathbf{P}_{jk} \mathbf{P}_{jk}^t .
\]  

(3.4.44)

A test statistic for \( H_{0} : \mathbf{\pi}_{j} = \mathbf{\pi}_{O} \) where \( \mathbf{\pi}_{O} \) is an unknown vector is

\[
X_{M2s}^2 = \left( \frac{\hat{\mathbf{A}}(3) \mathbf{A}(3)}{\mathbf{\pi} - \mathbf{\pi}_{O}} \right)^t \mathbf{\Sigma}_{O}^{-1} \left( \frac{\hat{\mathbf{A}}(3) \mathbf{A}(3)}{\mathbf{\pi} - \mathbf{\pi}_{O}} \right)
\]  

(3.4.46)

where \( \hat{\mathbf{A}}_{O} \) is a consistent estimator for \( \mathbf{\pi}_{O} \). One such estimator can be obtained by using \( \hat{\mathbf{A}}_{jk} \) to estimate \( \mathbf{P}_{jk} \) and \( \hat{\mathbf{\pi}}_{O} \) to estimate \( \mathbf{\pi}_{O} \).
Example 3.2.

This example demonstrates the testing of the hypothesis $H_0: \pi_j = \pi_0$ for an unknown $\pi_0$ under a stratified cluster sampling scheme. Data for this example were taken from Morton (1965) and reproduced in Appendix A, Tables 7.1 and 7.2. The data are analyzed using the statistic $X^2_{W12S}$ given in expression (3.4.46).

The example is considered to be a two-stage cluster design. There are two collection of books. From each collection of books, a random sample of size $n_j (j=1, 2)$ is taken. The values of $n_1$ and $n_2$ are nine and eight, respectively. These two subpopulations represent works from two Greek authors, Herodotus ($j=1$) and Thucydides ($j=2$). From each of the nine books (clusters) randomly selected from the works of Herodotus, a random sample of size $N_1 = 200$ (sentences) was taken and the occurrences of einai in these sentences were categorized into five categories: no einai, one einai, two einai, three einai, four or more einai. The vector of true proportions for these five categories is denoted by $\pi_1$. A random sample of $N_2 = 200$ sentences was taken from each of the eight selected books in the works of Thucydides, and the corresponding vector of true proportions is denoted by $\pi_2$.

The hypothesis of interest is whether the styles of the two authors, as reflected through the use of einai, are the same, that is

$$H_0: \pi_j = \pi_0 \quad (j = 1, 2)$$
where $\hat{\pi}_o$ is unknown. The test statistic derived earlier for this hypothesis and given in (3.4.46) is

$$X^2_{\text{W2IS}} = (\hat{\pi}_1 - \pi_o, -\hat{\pi}_2, \Sigma_{2S} (\hat{\pi}_1 - \pi_o)),$$

where $\hat{\pi}_1$ is the linear combination of

$$\hat{\pi}_1 = (.703889, .237778, .0494444, .0061111, .00277778)'$$

and

$$\hat{\pi}_2 = (.681875, .2525000, .05125, .01, .004375)'$$

given by

$$\hat{\pi}_o = \sum_{j=1}^{J} \frac{N_j \hat{\pi}_j}{\sum_{j=1}^{J} N_j}.$$

The values of $N_j$ ($j=1,2$), the total sample sizes for the populations are $N_1 = 1800$ and $N_2 = 1600$. The matrix $\hat{\Sigma}_{2S}$ is a consistent estimate of the covariance matrix for $(\hat{\pi}_1, \hat{\pi}_2) = (\hat{\pi}_1 - \pi_o)$ and is given by

$$\hat{\Sigma}_{2S} = (1-\alpha_1) \Sigma_{2S}(1) + \sum_{j=1}^{2} \alpha_j^2 \hat{\Sigma}_{2S}(j).$$
where

\[ \hat{\Sigma}_{zs}(j) = \frac{S_j}{n_j} \left[ \sum_{\ell=1}^{J} \alpha_{j\ell} (\hat{\alpha}_{\ell} - \hat{\rho}_{j\ell}) \right] \left[ \sum_{\ell=1}^{J} \alpha_{j\ell} (\hat{\rho}_{j\ell} - \hat{\alpha}_{\ell}) \right] \]

\[ s = \sum_{j=1}^{J} S_j , \]

\[ \hat{\alpha}_{\ell} = \text{diag}(\hat{\pi}_{1j}, \hat{\pi}_{2j}, \ldots, \hat{\pi}_{I-1j}) , \]

\[ \hat{\rho}_{jk} = (\hat{\pi}_{1jk}, \hat{\pi}_{2jk}, \ldots, \hat{\pi}_{I-1jk})' , \]

is the vector of proportions for the kth cluster in the jth subpopulation, \( S_j \) is the number of clusters in the jth subpopulation. Since \( N_1 \), the total number of observations for the works of Herodotus, is 1800 and \( N_2 \), the total number of observations for the works of Thucydides, is 1600, then

\[ \alpha_1 = \frac{N_1}{\sum_{j=1}^{J} N_j} = 0.5294 , \]

\[ \alpha_2 = \frac{N_1}{\sum_{j=1}^{J} N_j} = 0.4706 . \]

and
\[ \hat{\Sigma}_0 = (0.6935, 0.2447, 0.0503, 0.0079, 0.0036)' . \]

The estimated covariance matrix \( \hat{\Sigma}(j) \) for the covariance matrix of \( \hat{\Sigma}_j \) is

\[
\hat{\Sigma}(1) = \begin{bmatrix}
56.3408 & -33.2793 & -17.7559 & -3.4192 & -1.8864 \\
-33.2793 & 24.2636 & 6.70357 & 1.38993 & 0.92214 \\
-17.7559 & 6.70357 & 9.33314 & 1.09559 & 0.62363 \\
-3.4192 & 1.38993 & 1.09559 & 0.87656 & 0.057064 \\
-1.8864 & 0.92214 & 0.62363 & 0.057064 & 0.28356 \\
\end{bmatrix} \times 10^{-5}
\]

for the works of Herodotus, and

\[
\hat{\Sigma}(2) = \begin{bmatrix}
33.4301 & -25.2048 & -5.8483 & -1.1223 & -1.2548 \\
-25.2048 & 23.4754 & 0.3127 & 0.26758 & 1.14912 \\
-5.8483 & 0.3127 & 5.95869 & 0.2000 & -0.6231 \\
-1.1223 & 0.26758 & 0.2000 & 0.69609 & -0.041406 \\
-1.2548 & 1.14912 & -0.6231 & -0.041406 & 0.77014 \\
\end{bmatrix} \times 10^{-5}
\]

for the works of Thucydides. Then, the value of the test statistic, \( X^2_{W12} \) from (3.4.46), for testing \( H_0: \Pi_j = \Pi \) \( j = 1, 2 \), is

\[ X^2_{W12} = 1.37217. \]
When the null hypothesis is true, $\chi^2_{\text{W12S}}$ is asymptotically distributed as a chi-square random variable with four degrees of freedom. The p value for the test of the hypothesis is 0.85, hence, the two authors appear to be similar in their use of einai.
4. DIRICHLET-MULTINOMIAL MODEL FOR CLUSTER SAMPLING

4.1. Introduction

In this chapter, a test of proportions is developed for several subpopulations under a Dirichlet-Multinomial model. This is an extension of a model used by Brier (1980) to obtain some test statistics for testing hypotheses for parameters in log-linear models when the frequencies are obtained from a particular kind of cluster sampling from one population.

In Section 4.2, the use of the Dirichlet-Multinomial distribution for the random vector $\mathbf{X}$ which has a conditional multinomial distribution is discussed. A test for the equality of several probability vectors is developed in Section 4.3, both for the case when the common vector is known and when the vector is unknown. Three examples are considered. One pertains to data taken from Morton (1965) as presented in Chapter 3, Section 3.4. The second is the data given in Appendix A for wild turkeys. The third is a data set taken from Brier (1980) pertaining to household satisfaction. In Section 4.4, the results obtained in Chapter 3, Section 3.4, are compared to the results obtained in Section 4.3.

4.2. Dirichlet-Multinomial Model

Consider a population consisting of several clusters. A sample of $S$ clusters is randomly chosen with replacement and with probability proportional to the cluster size. Suppose $n$ individuals are randomly sampled with replacement from each of the $S$ clusters, then the total sample size is
One example can be described as follows. An author's collection of books is randomly sampled and from each chosen book a number of sentences is randomly chosen. This example is discussed in Section 3.4 and is reanalyzed at the end of Section 4.3.

Brier (1980) considered the case of cluster sampling from a single population. Since the model developed in this chapter is an extension of Brier's one population model, a brief review of that work is now given. Brier (1980) considered a vector of proportions $\mathbf{p}_t$ for each cluster in a single population. The vectors corresponding to the sampled clusters,

$$\mathbf{p}_t = (p_{1t}, p_{2t}, \ldots, p_{It})'$$

for $t = 1, 2, \ldots, S; \ldots; S$; are assumed to be distributed independently and identically with distribution function $F(p)$. The distribution $F(p)$ that Brier (1980) considered is the Dirichlet distribution. This distribution, described by Good (1965) as a conjugate prior distribution for the cell probabilities for multinomial models, has a density function

$$f(\mathbf{p}_t | \pi, \mathbf{K}) = \frac{\Gamma(\mathbf{K})}{\Gamma(\sum_{i=1}^{I} \mathbf{K} \pi_i)} \prod_{i=1}^{I} \frac{\mathbf{K} \pi_i - 1}{\mathbf{K} \pi_i} .$$

The parameters are the vector $\pi$ and $K$. The vector $\pi = (\pi_1, \pi_2, \ldots, \pi_I)'$
is a probability vector with dimension $I$, and $K > 0$ is a scaling parameter. As a prior, the Dirichlet distribution has some nice features. It is mathematically tractable and it provides a large class of distributions on the simplex $\xi_I$.

Within each cluster the vector of category counts,

$$X_t = (X_{1t}, X_{2t}, \ldots, X_{It})', \ t = 1, 2, \ldots, S; \quad (4.2.4)$$

has a multinomial distribution, with parameters $n$ and $p_t$, conditional upon $p_t$. With a Dirichlet distribution on the $p_t$ the unconditional distribution of $X_t$ is given by

$$f(X_t | \pi, K) = \frac{n!}{x_{1t}! x_{2t}! \cdots x_{It}!} \prod_{i=1}^{I} \frac{\Gamma(K)}{\Gamma(n+K)} \frac{\Gamma(X_{it}+Kn_i)}{\Gamma(Kn_i)} \quad (4.2.5)$$

where the probability vector $\pi$ and $K$ are parameters. This unconditional distribution is referred to as the Dirichlet-Multinomial distribution and is denoted by $\text{DM}_I(n, \pi, K)$. The use of the Dirichlet-Multinomial distribution for $X_t$ was discussed by Mosemann (1962). Wilks (1962) obtained the moments for the Dirichlet-Multinomial distribution, by first considering the moments of the Dirichlet distribution. Wilks (1962) found that the moments of the Dirichlet distribution are

$$E(p_i) = \pi_i,$$

$$E(p_i^2) = \pi_i (1+K)^{-1} (1+K\pi_i),$$
and

\[ E(p_i p_{i'}) = (1+K)^{-1} K \pi_i \pi_{i'}, \quad i \neq i'. \]  \hfill (4.2.6)

The moments of the multinomial distribution are

\[ E(x_k | \pi) = n \pi_k \]  \hfill (4.2.7)

and

\[ V(x_k | \pi) = n (\Delta - \pi \pi'), \]  \hfill (4.2.8)

where \( \Delta \) is the diagonal matrix,

\[
\Delta = \begin{pmatrix}
\pi_1 & 0 \\
0 & \pi_2 \\
& \ddots \\
0 & 0 & \pi_I
\end{pmatrix}.
\hfill (4.2.9)

Then, the moments for the Dirichlet-Multinomial distribution are

\[ E(x_k | \pi) = n \pi_k \]  \hfill (4.2.10)

and

\[ V(x_k | \pi) = n C (\Delta - \pi \pi'), \]  \hfill (4.2.11)

where

\[ C = (1+K)^{-1} (n+K). \]  \hfill (4.2.12)

Thus, the covariance matrix is a constant \( C \) times the corresponding
multinomial covariance matrix. This relationship between the covariance matrices for $X_t$ under the multinomial distribution and under the Dirichlet-Multinomial distribution will be used to develop estimators for the factor, $C$. The new estimator for $C$ will be compared to the estimator that Brier (1980) obtained.

In general, one can consider hypotheses about $\eta$ where $\eta$ is the vector of interest under this Dirichlet-Multinomial model. The hypotheses can be denoted by

$$H_0: \eta = f(\theta), \quad \theta \in \Omega \subseteq \mathbb{R}^p.$$  \hspace{1cm} (4.2.13)

It is assumed that the function $f$ satisfies Birch's (1964) regularity conditions as stated in Theorem 2.8.

Define the vector of over-all counts

$$X = \sum_{t=1}^{S} X_t = (X_{1+}, X_{2+}, \ldots, X_{I+})'.$$

(4.2.14)

Assume that for each $X_t$, $t = 1,2,\ldots,S$, the sample size $n$ is the same. Then, Brier (1980) proved that, under the null hypothesis, the statistics $X^2_B$ and $G^2_B$ given by

$$X^2_B = \sum_{i=1}^{I} \frac{(X_{i+} - \Lambda_n)^2}{\Lambda_n},$$ \hspace{1cm} (4.2.15)

$$G^2_B = 2 \sum_{i=1}^{I} X_{i+} \log(\frac{X_{i+}}{\Lambda_n}),$$ \hspace{1cm} (4.2.16)
where $\hat{\pi}_i$ is an estimator for $\pi_i$, have asymptotic distributions that are multiples of central chi-square distributions with $I-p-1$ degrees of freedom, where $p$ is the number of parameters to be estimated. The following theorem proved by Brier (1980) gives the asymptotic distributions for these statistics, ($\chi^2_B$ and $\sigma^2_B$).

**Theorem 4.1.**

Let $X_1, X_2, \ldots, X_S$ be iid $DM^+(n, \pi^0, K)$ and let $\hat{\pi}_m = N^{-1} \sum_{j=1}^{S} X_j$ be the moment estimate of $\pi$. Then,

1) $\sqrt{S}(\hat{\pi}_m - \pi^0) \overset{D}{\longrightarrow} N(0, \Sigma)$

where

$$\Sigma = n^{-1}C(\Delta_{\pi_m}^0 - \pi^0 \pi^0)^r$$

and

$$C = (n+K)(1+K)^{-1},$$

2) $\chi^2_B \overset{D}{\longrightarrow} \chi^2_{I-p-1}$,

and

3) $\sigma^2_B \overset{D}{\longrightarrow} \chi^2_{I-p-1}$,

where $\chi^2_{I-p-1}$ denotes a central chi-square distribution with $I-p-1$ degrees of freedom.

In Theorem 4.1, the multiplier $C$ depends on the parameter $K$, which is generally unknown. Brier (1980) noted that
1 \leq C \leq n \tag{4.2.17}

and this can be used to obtain a conservative test.

If the appropriate percentile of a chi-square distribution with \( I-p-1 \) degrees of freedom is less than \( n^{-1}X_B^2 \) or \( n^{-1}G_B^2 \) then reject \( H_0 \). Whereas, if the appropriate percentile of a chi-square distribution with \( I-p-1 \) degrees of freedom is greater than \( X_B^2 \) or \( G_B^2 \) one would not reject \( H_0 \) at the given level. However, if

\[ X_{I-p-1}^2(\alpha) < X_B^2 < nX_{I-p-1}^2(\alpha) \tag{4.2.18} \]

then, a decision cannot be made.

In general, valid asymptotic tests not depending on unknown parameters can be obtained by using the fact that if a random variable \( X_t \) converges in probability to a constant \( b \) and another random variable \( Y_t \) converges in distribution to the random variable \( Y \), then

\[ X_t Y_t \xrightarrow{D} bY. \tag{4.2.19} \]

That is, the product \( X_t Y_t \) of the random variables converges in distribution to the random variable \( bY \). Such a result has been proven by Cramer (1946). Consequently, if a consistent estimator \( \hat{C} \) is available for \( C \), then

\[ \frac{\hat{C}}{C} X_B^2 \xrightarrow{D} X_{I-p-1}^2. \tag{4.2.20} \]
Similarly,

\[ C \overset{A_{-1,2}}{\longrightarrow} B \overset{2}{\longrightarrow} X_{1-p-1}^2. \] (4.2.21)

A simple consistent estimator of \( C \) can be obtained from the method of moments and the fact that the covariance matrix for the random vector

\[ X_\mathbf{t} = (X_{1t}, X_{2t}, \ldots, X_{rt}) \] (4.2.22)

is

\[ \text{var}(X_\mathbf{t}) = n C (\Delta - \bar{\pi} \pi'). \] (4.2.23)

Then,

\[ V(X_{it}) = n C T_i (1 - \bar{\pi}_i) \] (4.2.24)

and a moment estimator for \( \bar{\pi}_i \) is

\[ \hat{\pi}_i = N^{-1} X_{i^+}. \] (4.2.25)

An alternative estimator of the \( V(X_{it}) \) is

\[ V(X_{it}) = (S-1)^{-1} \sum_{t=1}^{S} (X_{it} - S^{-1} X_{i^+})^2. \] (4.2.26)

Then, a method of moments estimator for \( C \) is given by

\[ \hat{C}_i = V(X_{it}) / n \bar{\pi}_i (1 - \bar{\pi}_i) \] (4.2.27a)
based on the sample variances, or
\[
\hat{C}_i = \frac{\text{Cov}(X_{it},X_{i,t})/n_{i}}{\hat{\pi}_i (1-\hat{\pi}_i)} \tag{4.2.27b}
\]
based on the sample covariances, where \(\hat{\text{Cov}}(X_{it},X_{i,t})\) is the estimate of \(\text{Cov}(X_{it},X_{i,t})\). The corresponding estimate of \(K\) is
\[
\hat{K} = \frac{n-C}{C-1}. \tag{4.2.28}
\]
However, for \(I > 2\) one could obtain many different estimators. There is an estimator corresponding to each variance as well as an estimator corresponding to each sample covariance. While each of these moment estimators for \(C\) is consistent one may wish to have a single estimator based on the information contained in all of the estimators. Brier (1980) suggested a way to pool the \(I^2\) estimators given in (4.2.27). Brier's pooled estimator is
\[
\hat{C}_B = \frac{1}{(S-1)(I-1)} \sum_{t=1}^{S} [n^{-1}(X_{it} - \hat{\pi}_i)^\text{T} \Delta_{A}^{-1}(X_{it} - \hat{\pi}_i)]. \tag{4.2.29}
\]
where
\[
\Delta_{A} = \begin{pmatrix}
\hat{\pi}_1 & \hat{\pi}_2 & \cdots & 0 \\
\hat{\pi}_2 & \hat{\pi}_1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \hat{\pi}_I
\end{pmatrix}. \tag{4.2.30}
\]
The estimator \(\hat{C}_B\) was obtained by noting that
\[
\text{var}(\hat{\chi}_c) = nC(\frac{\Delta}{\Pi} - \frac{\Pi}{\Pi'}) \tag{4.2.31}
\]

\[
= nC\frac{\Delta}{\Pi}(I - \sqrt{\frac{\Pi}{\Pi'}} \sqrt{\frac{\Pi'}{\Pi}}) \tag{4.2.32}
\]

where \(I\) is the identity matrix of dimension \(I\) and \(\Pi\) is assumed to be known so that \(R\) is known, then

\[
\sqrt{\frac{\Pi}{\Pi'}} = (\sqrt{\Pi_1}, \sqrt{\Pi_2}, \ldots, \sqrt{\Pi_I})'. \tag{4.2.33}
\]

The expression \(I - \sqrt{\frac{\Pi}{\Pi'}}\) is a projection operator of rank \(I-1\), so an orthogonal matrix \(Q\) can be obtained such that

\[
Q(I - \sqrt{\frac{\Pi}{\Pi'}})Q' = \begin{pmatrix}
I_{I-1} & 0 \\
0 & \vdots \\
0 & 0 & \ddots & 0
\end{pmatrix}. \tag{4.2.34}
\]

Let

\[
R = Q \Delta \frac{\Pi}{\Pi} \tag{4.2.35}
\]

then,

\[
R[\Delta \frac{\Pi}{\Pi}(I - \sqrt{\frac{\Pi}{\Pi'}})\Delta \frac{\Pi}{\Pi}]R' = \begin{pmatrix}
I_{I-1} & 0 \\
0 & \vdots \\
0 & 0 & \ddots & 0
\end{pmatrix}. \tag{4.2.36}
\]

Let

\[
\hat{\chi}_c' = R(\hat{\chi}_c - \Pi) \tag{4.2.37}
\]

where \(\Pi\) is assumed to be known so that \(R\) is known, then
\[ E(Y^*) = 0 \] (4.2.38)

\[ V(Y^*) = n R \text{var}(X^)_R^* \] (4.2.39)

\[ = n C \begin{pmatrix} I_{I-1} & 0 \\ 0 & \ddots \\ 0 & \cdots & 0 \end{pmatrix}. \] (4.2.40)

Suppose \( S \) independent vectors, \( Y^*_1, Y^*_2, \ldots, Y^*_S \), are observed, then from (4.2.40) an estimator of \( C \) is

\[ C^* = n^{-1}(I-I)^{-1}S^{-1} \sum_{i=1}^{I-1} \sum_{j=1}^{S} Y^2_{ij} \] (4.2.41)

where

\[ Y^*_j = (Y_{1j}, Y_{2j}, \ldots, Y_{Ij})' \] (4.2.42)

Note that

\[ I^{-1} \sum_{i=1}^{I} Y_{ij} = 0 . \] (4.2.43)

If \( \pi \) is unknown then

\[ \hat{X}^* = R(\hat{X} - \hat{\pi}) \] (4.2.44)

where \( \hat{R} \) is an estimator of \( R \) given by

\[ \hat{R} = R \hat{R}^{-\frac{1}{2}} \] (4.2.45)
and $\hat{\Pi}_m$ is a moment estimator for $\Pi$.

An estimator $\hat{C}_B$ for $C$ when $\Pi$ is unknown is

$$\hat{C}_B = n^{-1}(I-1)^{-1}\sum_{i=1}^{I-1} [(S-1)^{-1}\sum_{j=1}^S (Y_{ij}^*)^2]$$

(4.2.46)

and this can be written as

$$\hat{C}_B = n^{-1}(I-1)^{-1}(S-1)^{-1}\sum_{j=1}^S \hat{\phi}_j^* \begin{pmatrix} I_{I-1} & 0 \\ 0 & \hat{\lambda}_j \end{pmatrix} \begin{pmatrix} I_{I-1} & 0 \\ 0 & \hat{\lambda}_j \end{pmatrix}$$

(4.2.47)

But

$$\hat{\phi}_j^* \begin{pmatrix} I_{I-1} & 0 \\ 0 & \hat{\lambda}_j \end{pmatrix} \hat{\phi}_j^* = (X_j - \hat{\Pi}_m)^\top R (X_j - \hat{\Pi}_m)$$

(4.2.48)

where

$$\hat{\lambda}_j = (1, 1, \ldots, 1)'$$

(4.2.49)

and

$$(X_j - \hat{\Pi}_m)'\hat{\lambda}_j = 0$$

(4.2.50)

since
so

\[
\hat{C}_B = n^{-1}(I-1)^{-1}(S-1)^{-1} \sum_{j=1}^{S} [\hat{x}_{ij} - \hat{\mu}_j] \Delta^{-1}_{\hat{\mu}_j} (\hat{x}_{ij} - \hat{\mu}_j). \tag{4.2.52}
\]

\(\hat{C}_B\) is a multiple of the Pearson chi-square statistic for testing equality of the \(S\) probability vectors in the \(I \times S\) table formed by classifying units by cluster and by category. \(\hat{C}_B\) measures the differences in the clusters. The multiplier is the reciprocal of the number of degrees of freedom for testing the hypothesis of equality of the \(S\) probability vectors in the \(I \times S\) table. The estimator \(\hat{C}_B\) is consistent as \(S \to \infty\) and is well-defined. It is possible for \(\hat{C}_B\) to be smaller than 1 or greater than \(n\) which are not allowable values. In such cases \(\hat{C}_B\) can be truncated at 1 and \(n\).

A maximum likelihood estimator for \(K\) can be obtained by considering the likelihood function for the \(S\) vectors. Suppose \(X_1, X_2, \ldots, X_S\) are iid \(DM_1(n, \mu, K)\) then, the likelihood function can be written as

\[
L(\pi, K; X_1, X_2, \ldots, X_S) = \prod_{i=1}^{S} \left\{ \binom{n}{x_{1t}, x_{2t}, \ldots, x_{it}} \frac{\Gamma(K)}{\Gamma(n-K)} \frac{\prod_{i=1}^{n} \Gamma(x_{it} + K \mu_i)}{\Gamma(n-K) \prod_{i=1}^{K} \mu_i} \right\}. \tag{4.2.53}
\]

The maximum likelihood estimators for \(\pi\) and \(K\) can be only obtained numerically using an iterative method such as the Newton Raphson method.
Brier (1980) has shown that when the \( \overline{y} \) vector is known, then the likelihood ratio has at most one local maximum. However, when the \( \overline{y} \) vector is unknown it is not clear that the likelihood ratio has a unique maximum. Although simulation carried out by Brier found nothing to contradict the conjecture that the likelihood ratio will have at most one local maximum, there is no closed form solution for the likelihood equations, hence, it is not possible to write down a formula for the maximum likelihood estimators for \( \overline{y} \) and \( K \).

Obtaining the maximum likelihood estimator for \( K \) can be tedious. However, this method of estimation for \( K \) does not require equal cluster sizes.

Under the Dirichlet-Multinomial model, it might be expected that the different \( \chi^2 \) moment estimators that Brier (1980) obtained to estimate \( C \) would be similar. A graph of \( \hat{V}(X_{it}) \) versus \( \hat{\eta}_i (1-\hat{\eta}_i) \) and \( \text{Cov}(X_{it}, X_{i't}) \) versus \( -\hat{\eta}_i \hat{\eta}_i \), should be somewhat helpful. If the values for \( \hat{V}(X_{it}) \) and \( \text{Cov}(X_{it}, X_{i't}) \) are plotted on the y-axis and the values of \( \hat{\eta}_i (1-\hat{\eta}_i) \) and \( -\hat{\eta}_i \hat{\eta}_i \), on the x-axis then, the points should be well approximated by a straight line.

Another method of estimating \( C \) is now considered. Recall that the variance of \( \hat{\eta}_i \) under the Dirichlet-Multinomial model is

\[
\text{var}_{DM}(\hat{\eta}_i) = S^{-1} n^{-1} C(\Delta_{\overline{y}} - \overline{\eta}_{i}\overline{\eta}_{i}')
\]  

and under the multinomial model the covariance matrix is

\[
\text{var}_{M}(\hat{\eta}_i) = S^{-1} n^{-1} (\Delta_{\overline{y}} - \overline{\eta}_{i}\overline{\eta}_{i}')
\]

A simple moment estimator for \( \text{var}_{DM}(\hat{\eta}_i) \) is
\[
\text{var}_{\text{DM}}(\hat{\pi}) = (S-1)^{-1} \sum_{t=1}^{S} \left( \hat{\pi}_{1t} - \bar{\pi}_{1} \right) \left( \hat{\pi}_{1t} - \bar{\pi}_{1} \right)'. 
\] (4.2.56)

where
\[
\bar{\pi}_{1} = S^{-1} \sum_{t=1}^{S} \hat{\pi}_{1t}. 
\] (4.2.57)

The sample covariance matrix in (4.2.56) can be expressed in a vector form by writing

\[
\text{vech}[\text{var}(\hat{\pi})] = (S-1)^{-1} \begin{bmatrix} \sum_{t=1}^{S} (\hat{\pi}_{1t} - \bar{\pi}_{1})^2 \\ \sum_{t=1}^{S} (\hat{\pi}_{1t} - \bar{\pi}_{1})(\hat{\pi}_{2t} - \bar{\pi}_{2}) \\ \vdots \\ \sum_{t=1}^{S} (\hat{\pi}_{I-1t} - \bar{\pi}_{I-1})^2 \end{bmatrix} 
\] (4.2.58)

\[
= \begin{bmatrix} 
\frac{1}{S-1} \sum_{t=1}^{S} (\hat{\pi}_{1t} - \bar{\pi}_{1})^2 \\
\frac{1}{S-1} \sum_{t=1}^{S} (\hat{\pi}_{1t} - \bar{\pi}_{1})(\hat{\pi}_{2t} - \bar{\pi}_{2}) \\
\vdots \\
\frac{1}{S-1} \sum_{t=1}^{S} (\hat{\pi}_{I-1t} - \bar{\pi}_{I-1})^2 
\end{bmatrix}. 
\] (4.2.59)

\text{vech}[\text{var}(\hat{\pi})] \text{ can be further expressed as}

\[
\text{vech}[\text{var}(\hat{\pi})] = \tilde{\nu}, 
\] (4.2.60)
where

$$
\hat{\Sigma} = \begin{bmatrix}
\frac{1}{S-1} \sum_{t=1}^{S} V_{11t} \\
\vdots \\
\frac{1}{S-1} \sum_{t=1}^{S} V_{(I-1)(I-1)t}
\end{bmatrix}
$$

(4.2.61)

and

$$
V_{ii't} = (\hat{\pi}_{it} - \bar{\pi}_{i}) (\hat{\pi}_{i't} - \bar{\pi}_{i})'.
$$

(4.2.62)

The vector of expected values for \( \text{vech}[\text{var}(\eta)] \) is

$$
\mathbb{E}[\bar{\eta}] = n^{-1} \begin{bmatrix}
\pi_1 - \bar{\pi}_1 \\
-\pi_1 \pi_2 \\
\vdots \\
\pi_{I-1} - \bar{\pi}_{I-1}
\end{bmatrix}
$$

(4.2.63)

A generalized least squares estimator for \( C \) is given by

$$
\hat{C}_{\text{wls}} = (\hat{\Sigma}_{\text{wv}})^{-1} \hat{\Sigma}_{\text{wv}}
$$

(4.2.64)

where \( \hat{\Sigma}_{\text{wv}} \) is a consistent estimator of the covariance matrix for \( \bar{\eta} \)

when \( H_0 \) is true, and

$$
\hat{\eta} = (\hat{\pi}_1 - \bar{\pi}_1, \hat{\pi}_1 \hat{\pi}_2, \ldots, \hat{\pi}_{I-1} - \bar{\pi}_{I-1})'.
$$

(4.2.65)

The estimator, \( \hat{C}_{\text{wls}} \) for \( C \) assumes that \( \hat{\eta} \) is a fixed vector, in the computation of the generalized regression estimator.
The generalized least squares technique for estimating $C$ also provides an approximate goodness-of-fit test for the model. If the matrix $\Sigma_{vv}$ is known, the generalized residual mean square has a limiting chi-square distribution with degrees of freedom equal to the number of elements in $\tilde{\Sigma}$ less one.

Three possible estimators for $\Sigma_{vv}$ are now given. The first is a nonparametric estimator associated with (4.2.61). Because $\tilde{\Sigma}$ is expressed as a mean, an estimator of $\text{var}\{\tilde{\Sigma}\}$ is

$$\hat{\Sigma}_{vv} = \text{var}\{\tilde{\Sigma}\}$$

$$= (s-1)^{-2} \sum_{t=1}^{S} (\tilde{\Sigma}_t - \tilde{\Sigma})(\tilde{\Sigma}_t - \tilde{\Sigma})',$$

(4.2.66)

where

$$\tilde{\Sigma}_t = (\tilde{\Sigma}_{11t}, \tilde{\Sigma}_{12t}, \ldots, \tilde{\Sigma}_{I-1,I-1,t})' .$$

(4.2.67)

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

The second estimator of $\Sigma_{vv}$ uses the Dirichlet-Multinomial model and assumes that the cluster sizes are large enough so that the normal distribution can be used to construct the variance of the sample variances. We consider the transformed observations,

$$\tilde{\Sigma}_j = n^{-1}(\tilde{\Sigma}_j - \tilde{\Sigma})R'$$

$$= (\tilde{\Sigma}_j - \tilde{\Sigma})R' ,$$

(4.2.68)
where \( \mathbf{R}' \) is the matrix such that

\[
\mathbf{R}' = \begin{bmatrix}
\mathbf{I}_{I-1} & \mathbf{0} \\
\mathbf{0}' & \mathbf{0}
\end{bmatrix}
\]

and \( \mathbf{\Sigma}_m \) is the multinomial covariance matrix for a sample of size \( n \). Under the model the covariance matrix of the first \( I-1 \) elements of \( \mathbf{Y}_j \) is a multiple of the identity matrix. Let

\[
(S-1)^{-1} \sum_{j=1}^{S} \mathbf{Y}_j \mathbf{Y}_j' = \begin{pmatrix}
\mathbf{V}_{YY} & \mathbf{0} \\
\mathbf{0}' & \mathbf{0}
\end{pmatrix}
\]

Then, under the model

\[
\mathbb{E}\{\mathbf{\hat{V}}_{YY}\} = \mathbf{V}_{YY} = CI.
\]

If \( \mathbf{Y}_j \) is normally distributed,

\[
\text{var}\{\text{vech} \mathbf{\hat{V}}_{YY}\} = (S-1)^{-1} C^2 \mathbf{D}_B,
\]

where

\[
\mathbf{D}_B = \text{diag}(2, 1, 1, \ldots, 2, 1, 2)
\]

and the element of the diagonal matrix \( \mathbf{D}_B \) for an estimated variance is two and the element for an estimated covariance is one. Expression (4.2.70) is a special case of the following theorem taken from Anderson (1958). An estimator of \( \text{var}\{\text{vech} \mathbf{\hat{V}}_{YY}\} \) is obtained by replacing \( C \)
with $\hat{C}_B$ in (4.2.70). This estimator is denoted by $\hat{V}_{\text{vech}} \hat{V}_{YY}$.  

Theorem 4.2.

Let $A(S) = \sum_{j=1}^{S} (z_j - \bar{z})(z_j - \bar{z})'$, where $z_1, z_2, \ldots, z_S$ are independently and identically distributed according to $N(\mu, \Sigma)$. Then, the asymptotic joint distribution of the elements of $B(S) = (S-1)^{-\frac{1}{2}}[A(S)-(S-1)\Sigma]$ is normal with mean $0$ and covariances $E[b_{ij}(S)b_{kl}(S)^\prime] = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}'$, where $b_{ij}(S)$ is the $(i,j)$ element of $B(S)$ and $\sigma_{ij}$ is the $(i,j)$ element of $\Sigma$.

Proof: Showing that $A(S)$ is distributed as $\sum_{j=1}^{S-1} T_j T_j'$ where $T_1, T_2, \ldots, T_{S-1}$ are distributed independently according to $N(0, \Sigma)$ and taking moments of linear combinations of $z_j$, Anderson (1958) proved this theorem.

Because $\Sigma_m$ is unknown it is necessary to replace $\Sigma_m$ with $\hat{\Sigma}_m$ and $R$ with $\hat{R}$, where

$$\hat{R} \hat{\Sigma}_m \hat{R}' = \hat{\Sigma}_m.$$ 

Then, the estimated generalized least squares estimator of $C$ is

$$\hat{C}_B = (\hat{A}'D_B^{-1}\hat{A})^{-1}\hat{A}'D_B^{-1}\hat{H},$$

(4.2.71)

where

$$\hat{A} = \text{vech} I,$$

$$\hat{H} = \text{vech} \hat{V}_{YY}.$$
As the notation indicates, (4.2.71) is another expression for Brier's estimator. Under the normal assumption and with $\Sigma_m$ known, an estimator of the variance of $\hat{C}_B$ is

$$\hat{V}[\hat{C}_B] = (I-1)^{-1} \hat{A}_B^2 (A_B^\prime D_B^{-1} A_B)^{-1}.$$  \hfill (4.2.72)

A lack-of-fit statistic for the model is

$$X^2_B = (I-1) \hat{C}_B^{-2} [H_B^\prime D_B^{-1} H_B - H_B^\prime D_B^{-1} A_B A_B^\prime].$$  \hfill (4.2.73)

If $\Sigma_m$ is known and the model true, the large sample distribution of $\chi^2$ is approximately that of a chi-square random variable with $2^2(I-1)I-1$ degrees of freedom. Here we assume a consistent estimator of $\Sigma_m$ is available and that the subpopulation sample size approaches infinity. Then by the use of the Multivariate Central Limit Theorem $X^2_B$ converges to a chi-square random variable. Alternative estimators of the variance and alternative tests are considered in Section 4.4 and in Appendix B.

The third estimator of $\Sigma_{VV}$ falls between the previous two in the amount of model information used in the construction. Under the model the covariance matrix of $\text{vech} \hat{V}_{YY}$ is a diagonal matrix. An estimator of the variance of the $ij^{th}$ element is

$$D_{Wij} = (S-1)^{-2} \sum_{t=1}^{S} (Y_{it} Y_{jt} - \hat{V}_{YYij})^2,$$  \hfill (4.2.74)
where $\hat{V}_{Yij}$ is the $ij^{th}$ element of $\hat{V}_Y$ defined in (4.2.69). Then, a generalized least squares estimator of $C$ is

$$
\hat{C}_W = (\hat{\Sigma}_W^{-1} \hat{A})^{-1} \hat{\Sigma}_W^{-1} \hat{A}^T,
$$

(4.2.75a)

where

$$
D_W = \text{diag}(D_{W11}, D_{W21}, \ldots, D_{W111}, D_{W111}).
$$

The associated test statistic is

$$
X^2_W = \hat{\Sigma}_W^{-1} \hat{A}^T \hat{\Sigma}_W^{-1} \hat{A}.
$$

(4.2.75b)

If $\Sigma_m$ is known and the model true, the large sample distribution of $X^2_W$ is that of a chi-square random variable with $2^{-1}(1-1)1$ degrees of freedom. Assuming $D_W$ is a consistent estimator of the diagonal covariance matrix and that the subpopulation size is large, the Multivariate Central Limit Theorem can be applied to show that $X^2_W$ has a limiting chi-square distribution.

An alternative way of viewing the model in the least squares context is to note that, under the model

$$
E[\hat{\Pi}_W] = \Pi,
$$

$$
E[\hat{V}_{PP}] = n^{-1} \sum (\hat{\Delta}_{\Pi} - \Pi \Pi')
$$

Thus, one can use both $\hat{\Pi}_W$ and $\hat{V}_{PP}$ to simultaneously estimate the
unknown parameters \( \eta \) and \( C \). The simultaneous estimation requires an estimator of the covariance matrix of \( \left[ \eta', \text{vech } \hat{V}_{pp}' \right]' \). Also, nonlinear estimation methods are required because \( \mathbb{E}[\hat{V}_{pp}] \) is a non-linear function of \( \eta \).

4.3. A Test of Proportions

In the previous section, the description concentrated on one subpopulation to give a basic idea of the Dirichlet-Multinomial model. Extension is now made to the Dirichlet-Multinomial model for \( J > 1 \) subpopulations. Assume that for the \( j \)th subpopulation, \( j = 1, 2, \ldots, J \); the clusters within that subpopulation are randomly selected and a simple random sample of \( n_j \) is taken from each cluster with replacement. Let \( \tilde{X}_{t_j} \) denote the vector of counts for the \( t_j \)th cluster within the \( j \)th subpopulation. Further, assume that

\[
\tilde{X}_{t_j} \sim \text{iid DM}_{1}(n_j, \eta_j, K_j) \tag{4.3.1}
\]

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

The covariance matrix for the vector of sample totals for the \( j \)th subpopulation,

\[
S_j = \sum_{t_j=1}^{s_j} \tilde{X}_{t_j} \quad j = 1, 2, \ldots, J \tag{4.3.2}
\]
\[ \text{var}(X_j) = n_j S_j C_j \left( \bar{\pi}_j - \bar{\pi}_j \bar{\pi}_j' \right) \] (4.3.3)

where

\[ C_j = (1+K_j)^{-1} (n_j+K_j). \] (4.3.4)

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

\[ N_j = n_j S_j, \] (4.3.5)

then,

\[ \text{var}(X_j) = N_j C_j \left( \bar{\pi}_j - \bar{\pi}_j \bar{\pi}_j' \right). \] (4.3.6)

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

\[ \bar{\pi}_j = N_j^{-1} X_j' , \]

\[ = (\bar{\pi}_{1j}, \bar{\pi}_{2j}, \ldots, \bar{\pi}_{l-1j})', \] (4.3.7)

then, \( \bar{\pi}_j \) is an unbiased estimator of \( \pi_j \).

The covariance matrix for \( \bar{\pi}_j \) is

\[ B_j = N_j^{-1} C_j \left( \bar{\pi}_j - \bar{\pi}_j \bar{\pi}_j' \right). \] (4.3.8)
The vector of deviations \((\hat{\theta}_j - \theta_0)\) has a mean vector \(\theta_0\) and covariance matrix \(B_j\). Theorem 4.1 indicates as \(S_j \to \infty\)

\[
\frac{1}{N_j} (\hat{\theta}_j - \theta_0) \xrightarrow{d} N(0, N_j B_j).
\]

(4.3.9)

Suppose the interest is in testing the hypothesis

\[H_0: \theta_j = \theta_0 \quad j = 1, 2, ..., J;\]

(4.3.10)

where \(\theta_0\) is a known vector. Then, \(\hat{\theta}_j - \theta_0\) is an unbiased estimator for the vector \(\hat{\theta}_j - \theta_0\), and the covariance matrix of \(\hat{\theta}_j - \theta_0\) is given by \(B_j\). Under the null hypothesis stated in (4.3.10), a consistent estimator of \(B_j\) is

\[
\hat{B}_j = N_j^{-1} C_j (\hat{\theta}_j - \theta_0 \theta_0') - N_j^{-1} C_j (\hat{\theta}_j - \theta_0 \theta_0'),
\]

(4.3.11)

where \(C_j\) is a consistent estimator for \(C_j\). Then, a Wald test statistic for testing the hypothesis in (4.3.10) is

\[
X^2_{DMH} = (\hat{\theta}_j - \theta_0)' \hat{B}^{-1} (\hat{\theta}_j - \theta_0),
\]

(4.3.12)

where

\[
(\hat{\theta}_j - \theta_0) = (\hat{\theta}_1 - \theta_0, \hat{\theta}_2 - \theta_0, ..., \hat{\theta}_J - \theta_0)',
\]

(4.3.13)

and \(\hat{B}\) is a consistent estimator for the covariance matrix for \((\hat{\theta}_j - \theta_0)\). \(\hat{B}\) is a block diagonal matrix with entries given by
\( \hat{B}_j, j = 1, 2, \ldots, J; \) on the diagonal, where \( \hat{B}_j \) is defined in (4.3.8). The statistic \( x^2_{\text{DMH}} \) can be written as

\[
x^2_{\text{DMH}} = \sum_{j=1}^{J} (\hat{\pi}_j - \hat{\pi}_0)' B_j^{-1} (\hat{\pi}_j - \hat{\pi}_0)
\]

(4.3.14)

\[
x^2_{\text{DMH}} = \sum_{j=1}^{J} C_j^{-1} X_{mj}
\]

(4.3.15)

where

\[
x^2_{mj} = N_j (\hat{\pi}_j - \hat{\pi}_0)' (\hat{\Delta}_j - \hat{\pi}_0)'^{-1} (\hat{\pi}_j - \hat{\pi}_0)
\]

(4.3.16)

denotes a Pearson goodness-of-fit statistic for the \( j^{\text{th}} \) subpopulation and \( \hat{C}_j \) is a consistent estimator for \( C_j, j = 1, 2, \ldots, J. \) Since \( \sqrt{N_j} (\hat{\pi}_j - \hat{\pi}_0) \) has a limiting normal distribution with mean vector \( \hat{\pi}_0 \) and covariance matrix \( B_j \) and the \( \hat{\pi}_j's \) are independent, then, as shown in Cramer (1946), the joint distribution of the \( \hat{\pi}_j's \) will also have a limiting normal distribution with mean vector \( \hat{\pi}_0 \) and covariance matrix \( B. \) Therefore, \( x^2_{\text{DMH}} \) has a limiting chi-square distribution with \( J(J-1) \) degrees of freedom since \( B \) is assumed to be nonsingular.

4.3.1. A test of independence

Suppose in (4.3.12) the vector \( \hat{\pi}_0 \) is unknown and is estimated by a linear combination of the \( J \) unbiased estimators obtained from each of the \( J \) subpopulations. Then, an estimator for \( \hat{\pi}_0 \) is
where the $a_j$'s are positive and sum to one. The estimator $(\hat{\pi}_j - \hat{\pi}_0)$ is an unbiased estimator for $\pi_j - \pi_0$ if the $a_j$'s are fixed. Let $T_{jj}$ denote the covariance matrix for $(\hat{\pi}_j - \hat{\pi}_0)$ and $T_{jj'}$, denote the matrix of covariances between $(\hat{\pi}_j - \hat{\pi}_0)$ and $(\hat{\pi}_{j'} - \hat{\pi}_0)$. Then,

\[
T_{jj} = \text{var}(\hat{\pi}_j - \hat{\pi}_0)
\]

\[
= E[(\hat{\pi}_j - \pi_j)(\hat{\pi}_j - \pi_j)' - (\hat{\pi}_0 - \pi_0)(\hat{\pi}_0 - \pi_0)' - (\hat{\pi}_j - \pi_j)(\hat{\pi}_0 - \pi_0)'](\hat{\pi}_0 - \pi_0)'
\]

\[
= B_j - 2a_j^{-1}B_j + \sum_{k=1}^{J} a_k^{-2}B_k.
\]  

(4.3.18)

and

\[
T_{jj'} = \text{Cov}(\hat{\pi}_j - \hat{\pi}_0)(\hat{\pi}_{j'} - \hat{\pi}_0)
\]

\[
= -a_j^{-1}B_j - a_{j'}^{-1}B_{j'} + \sum_{k=1}^{J} a_k^{-2}B_k.
\]  

(4.3.19)

Then, a test statistic for the hypothesis

\[
H_0: \pi_j = \pi_0, \quad j = 1, 2, \ldots, J,
\]  

(4.3.20)
where $\pi^*_0$ is an unknown vector, can be constructed using a consistent estimator of the covariance matrix for

$$
\hat{\pi}^{(J)} - \pi^*_0 = (\hat{\pi}^{(1)} - \pi^*_0, \hat{\pi}^{(2)} - \pi^*_0, \ldots, \hat{\pi}^{(J-1)} - \pi^*_0). \tag{4.3.21}
$$

The $J^{th}$ subpopulation is left off of (4.3.21) to obtain a nonsingular covariance matrix. Let $M_{H_0}$ denote the nonsingular matrix of dimension $(I-1)(J-1)$ under $H_0$,

$$
M_{H_0} = \text{Cov}(\hat{\pi}^{(J)} - \pi^*_0). \tag{4.3.22}
$$

The matrix $M_{H_0}$ has diagonal blocks $T_{jj}$ and off diagonal blocks $T_{jj'}$.

Consider $J = 3$ then,

$$
T_{jj} = B_j - 2\alpha_j B_j + \sum_{k=1}^{3} \alpha^2_k B_k \tag{4.3.23}
$$

and

$$
T_{jj'} = -\alpha_j B_j - \alpha_j B_j + \sum_{k=1}^{3} \alpha^2_k B_k \tag{4.3.24}
$$

Recall that $B_j = N^{-1} \sigma_j C_j (\pi^*_j - \pi^*_0)$ is the covariance matrix for $\hat{\pi}^*_j$ under $H_0$. Then, for $J = 3$
which reduces to

\[
M_{H_0} = \begin{bmatrix}
3 (1-2\alpha_1)B_1 + \sum_{\ell=1}^{3} \alpha_\ell^2 B_\ell
& -\alpha_1 B_1 - \alpha_2 B_2 + \sum_{\ell=1}^{3} \alpha_\ell^2 B_\ell \\
-\alpha_1 B_1 - \alpha_2 B_2 + \sum_{\ell=1}^{3} \alpha_\ell^2 B_\ell
& (1-2\alpha_2)B_2 + \sum_{\ell=1}^{3} \alpha_\ell^2 B_\ell
\end{bmatrix}
\]

(4.3.25)

where

\[
B_{H_0} = (\Delta_{H_0} - \prod_{H_0}^n).
\]

(4.3.27)

Suppose that the \(\alpha_j\)'s are chosen proportional to the variance of \(\bar{\bar{Y}}_{\alpha_j}\) under \(H_0\) then,

\[
\alpha_j = \frac{N_j C^{-1}_j}{\sum_{\ell=1}^{3} N_j C^{-1}_\ell}.
\]

(4.3.28)

Substituting this value of \(\alpha_j\) in \(M_{H_0}\) in (4.3.26), gives
\[ M_{H_0} = \begin{bmatrix} (N_1^{-1}C_1 - 2g^{-1})B_{H_0} + g^{-1}B_{H_0} & -2g^{-1}B_{H_0} + g^{-1}B_{H_0} \\ -2g^{-1}B_{H_0} + g^{-1}B_{H_0} & (N_2^{-1}C_2 - 2g^{-1})B_{H_0} + g^{-1}B_{H_0} \end{bmatrix} \] 

(4.3.29)

\[ M_{H_0} = \begin{bmatrix} (N_1^{-1}C_1 - g^{-1})B_{H_0} & -g^{-1}B_{H_0} \\ -g^{-1}B_{H_0} & (N_2^{-1}C_2 - g^{-1})B_{H_0} \end{bmatrix}, \]

(4.3.30)

where

\[ g = \sum_{i=1}^{3} \sum_{\ell = 1}^{N_i} C_\ell^{-1}. \] 

(4.3.31)

\[ M_{H_0} \] can also be expressed as

\[ M_{H_0} = \begin{bmatrix} N_1^{-1}C_1 & 0 \\ 0 & N_2^{-1}C_2 \end{bmatrix} - g^{-1}J \otimes B_{H_0}, \] 

(4.3.32)

where \( J \) is a square matrix with entries all equal to one. Then, by Theorem 3.1
\[
M_{H_0}^{-1} = \begin{pmatrix}
N_1C_1^{-1} & 0 \\
0 & N_2C_2^{-1}
\end{pmatrix} + N_3^{-1}C_3 \begin{pmatrix}
N_1C_1^{-2} & N_1N_2C_1^{-1}C_2^{-1} \\
N_1N_2C_1^{-1}C_2^{-1} & N_2C_2^{-2}
\end{pmatrix} \otimes B_{H_0}^{-1}
\]

\[
= \begin{pmatrix}
[N_1C_1^{-1} + (N_3^{-1}C_3)(N_1C_1^{-2})]B_{H_0}^{-1} & (N_3^{-1}C_3)N_1N_2C_1^{-1}C_2^{-1}B_{H_0}^{-1} \\
(N_3^{-1}C_3)N_1N_2C_1^{-1}C_2^{-1}B_{H_0}^{-1} & [N_2C_2^{-1} + (N_3^{-1}C_3)(N_2C_2^{-2})]B_{H_0}^{-1}
\end{pmatrix}
\]

(4.3.33)

So a test statistic \( X_{DMI}^2 \) for testing

\[
H_0: \pi_j = \pi_0, \quad j = 1, 2, 3;
\]

(4.3.34)

where \( \pi_0 \) is unknown is given by

\[
X_{DMI}^2 = \left( \frac{\Lambda^T(\pi) - \Lambda^T(\pi_0)}{\pi - \pi_0} \right)_H, \quad H_{H_0}^{-1} = M_{H_0},
\]

(4.3.35)

when the matrix \( M_{H_0} \) is a consistent estimator of \( M_{H_0} \). One consistent estimator can be obtained by replacing \( \pi_0 \) with \( \pi_0 \) in \( M_{H_0} \). Then, the statistic \( X_{DMI}^2 \) reduces to

\[
X_{DMI}^2 = \sum_{j=1}^{2} N_jC_j^{-1} \left( \frac{\Lambda_j^T(\pi) - \Lambda_j^T(\pi_0)}{\pi_j - \pi_0} \right)_H B_{H_0}^{-1} \left( \frac{\Lambda_j^T(\pi) - \Lambda_j^T(\pi_0)}{\pi_j - \pi_0} \right)_H + N_3^{-1}C_3 \left[ \sum_{j=1}^{2} N_jC_j^{-2} \left( \frac{\Lambda_j^T(\pi) - \Lambda_j^T(\pi_0)}{\pi_j - \pi_0} \right)_H B_{H_0}^{-1} \left( \frac{\Lambda_j^T(\pi) - \Lambda_j^T(\pi_0)}{\pi_j - \pi_0} \right)_H \right]
\]
\[ + N_1 C_1^{-1} \mathbf{I} \mathbf{I}_{\mathbf{H}_0} N_2 C_2^{-1} \mathbf{I} \mathbf{I}_{\mathbf{H}_0} + N_2 C_2^{-1} \mathbf{I} \mathbf{I}_{\mathbf{H}_0} N_1 C_1^{-1} \mathbf{I} \mathbf{I}_{\mathbf{H}_0} \]

\[ = \sum_{j=1}^{2} N_j C_j^{-1} \mathbf{I} \mathbf{I}_{\mathbf{H}_0} + N_3 C_3^{-1} \mathbf{I} \mathbf{I}_{\mathbf{H}_0} \]

\[ = \sum_{j=1}^{2} N_j C_j^{-1} \mathbf{I} \mathbf{I}_{\mathbf{H}_0} + N_3 C_3^{-1} \mathbf{I} \mathbf{I}_{\mathbf{H}_0} \]

\[ = 2 \sum_{j=1}^{2} N_j C_j^{-1} \mathbf{I} \mathbf{I}_{\mathbf{H}_0} + N_3 C_3^{-1} \mathbf{I} \mathbf{I}_{\mathbf{H}_0} \]

\[ (4.3.36) \]

Since

\[ \sum_{j=1}^{J} N_j C_j^{-1} \mathbf{I} \mathbf{I}_{\mathbf{H}_0} = N_j C_j^{-1} \mathbf{I} \mathbf{I}_{\mathbf{H}_0}, \]

\[ (4.3.37) \]

it follows that

\[ x_{DMI}^2 = \sum_{j=1}^{J} N_j C_j^{-1} \mathbf{I} \mathbf{I}_{\mathbf{H}_0} \]

\[ (4.3.38) \]

where

\[ B_{H_0}^{-1} = (\Delta_{\mathbf{H}_0} \mathbf{I} \mathbf{I}_{\mathbf{H}_0})^{-1}. \]

\[ (4.3.39) \]

Then, by Theorem 3.1 the test statistic reduces to

\[ x_{DMI}^2 = \sum_{j=1}^{3} \sum_{i=1}^{N_j C_j^{-1}} (\mathbf{I} \mathbf{I}_{\mathbf{H}_0})^2 \]

\[ (4.3.40) \]
The statistic $X^2_{DMI}$ resembles a Pearson test statistic for independence. However, the estimated vector

$$\hat{\pi}_io = \sum_{j=1}^{J} \frac{N_j C_{i,j}}{\sum_{j=1}^{J} N_j C_{i,j}} \hat{\pi}_{ij}; \quad (4.3.41)$$

is not the usual sample weighted estimator

$$\hat{\pi}_io = \sum_{j=1}^{J} \frac{N_j}{\sum_{j=1}^{J} N_j} \hat{\pi}_{ij}. \quad (4.3.42)$$

By the Multivariate Central Limit Theorem, $\sqrt{N} (\hat{\pi}_o - \pi_o)$ has a limiting normal distribution and since $\sqrt{N} (\hat{\pi}_j - \pi_j)$ is a linear combination, for $C_j$ fixed, of the $\hat{\pi}_j$'s, then by Cramer (1946) $\sqrt{N} (\hat{\pi}_j - \pi_j)$ also has a limiting normal distribution. Under $H_0$, $\sqrt{N} (\hat{\pi}_j - \pi_j)$ has a mean vector 0 and covariance given by $M_j$. But

$$\frac{\hat{\pi}_i^{(J)}}{\pi_i^{(J)}} = [I \otimes (I - \frac{1}{\pi}^T \pi)]^{(J)}$$

is a linear combination of $\hat{\pi}_j^{(J)}$, where $I$ is the identity matrix of dimension $J-1$, $\otimes$ denotes direct product, $\frac{1}{\pi}$ represents the vector of ones of dimension $J-1$ and $\pi = (\pi_1, \pi_2, \ldots, \pi_{J-1})^T$. Therefore, the statistic $X^2_{DMI}$ is distributed asymptotically as a chi-square random variable with $(I-1)(J-1)$ degrees of freedom. If $C_j$ is unknown and a consistent estimator $\hat{C}_j$ is available then the asymptotic distribution of $X^2_{DMI}$ is still chi-square.

Consider the case with $J > 3$ subpopulations, then

$$\hat{\pi}_o = \sum_{j=1}^{J} \alpha_j \hat{\pi}_j, \quad (4.3.43)$$
where

\[ \sum_{j=1}^{J} \alpha_j = 1, \quad (4.3.44) \]

and each \( \alpha_j \) is nonnegative.

The covariance matrix has diagonal and off diagonal blocks as given in (4.3.23) and (4.3.24), respectively. The results obtained for the case \( J = 3 \) can be extended to the case \( J > 3 \). By a similar argument to that presented for \( J = 3 \) one can obtain

\[ X_{DMI}^2 = \sum_{j=1}^{J} N_j \sum_{i=1}^{\Lambda_j} \frac{(\Lambda_i - \Lambda_{io})^2}{\Lambda_{io}^{\Lambda_j}}. \quad (4.3.45) \]

where

\[ \Lambda_{io} = \frac{\sum_{j=1}^{J} N_j \frac{\Lambda_j^{-1}}{\sum_{l=1}^{\Lambda_j} \Lambda_{il}} \Lambda_{ij}}{\sum_{l=1}^{\Lambda_j} \sum_{j=1}^{J} N_j \Lambda_j^{-1} \sum_{l=1}^{\Lambda_j} \Lambda_{il}} \quad (4.3.46) \]

and the \( \alpha_j \) is chosen such that

\[ \alpha_j = \frac{\sum_{j=1}^{J} N_j \frac{\Lambda_j^{-1}}{\sum_{l=1}^{\Lambda_j} \sum_{j=1}^{J} N_j \Lambda_j^{-1} \sum_{l=1}^{\Lambda_j} \Lambda_{il}} \Lambda_{ij}}{\sum_{l=1}^{\Lambda_j} \sum_{j=1}^{J} N_j \Lambda_j^{-1} \sum_{l=1}^{\Lambda_j} \Lambda_{il}}. \quad (4.3.47) \]

The statistic for the case \( J > 3 \) also resembles the usual Pearson statistic for a test of independence. However, it differs from the Pearson
statistic in that the multiplier for \((\hat{\pi}_{ij} - \hat{\pi}_{io})^2 \hat{\pi}_{io} \) is \(N_j \hat{\pi}_{io} \) and not \(N_j \), also \(\hat{\pi}_{io} \) is not the maximum likelihood estimator obtained under multinomial sampling.

Note that if \(\hat{C}_j \equiv 1\) for \(j = 1, 2, \ldots, J\); then \(X^2_{DMI}\) reduces to the usual Pearson statistic,

\[
X^2_{DMI} = \sum_{j=1}^{J} \sum_{i=1}^{I} \frac{(\hat{\pi}_{ij} - \hat{\pi}_{io})^2}{\hat{\pi}_{io}}
\]

where

\[
\hat{\pi}_{io} = \sum_{j=1}^{J} \frac{N_j}{\sum_{j=1}^{J} N_j} \hat{\pi}_{ij} . \tag{4.3.49}
\]

If the \(\hat{C}_j\)'s are the same for each subpopulation, say

\[
\hat{C}_j = \hat{C}, \quad j = 1, 2, \ldots, J; \tag{4.3.50}
\]

then,

\[
X^2_{DMI} = \hat{C}^{-1} \sum_{j=1}^{J} N_j \sum_{i=1}^{I} \frac{(\hat{\pi}_{ij} - \hat{\pi}_{io})^2}{\hat{\pi}_{io}}
\]

where

\[
\hat{\pi}_{io} = \sum_{j=1}^{J} \frac{N_j \hat{\pi}_{ij}}{\sum_{j=1}^{J} N_j} . \tag{4.3.52}
\]
That is, if the $\hat{C}_j$'s are equal to some value $\hat{C}$, the statistic is

$$X_{DMI}^2 = \hat{C} X_{MI}^2$$  \hfill (4.3.53)$$

where $X_{MI}^2$, defined in expression (4.2.15), is the Pearson statistic.

Suppose that the $N_j \hat{C}_j^{-1}$'s are the same, say

$$N_j \hat{C}_j^{-1} = A,$$  \hfill (4.3.54)$$

then,

$$X_{DMI}^2 = A \sum_{j=1}^{J} \sum_{i=1}^{I} \frac{\left(\hat{\pi}_{ij} - \frac{1}{\sum_{j=1}^{J} \hat{\pi}_{ij}}\right)^2}{\sum_{j=1}^{J} \frac{1}{\sum_{i=1}^{I} \hat{\pi}_{ij}}} \left(\frac{\sum_{j=1}^{J} \hat{\pi}_{ij}}{\sum_{j=1}^{J} \sum_{i=1}^{I} \hat{\pi}_{ij}}\right)^2$$  \hfill (4.3.55)$$

$$= AJ^{-1} \sum_{j=1}^{J} \sum_{i=1}^{I} \frac{\left(J \hat{\pi}_{ij} - \sum_{j=1}^{J} \hat{\pi}_{ij}\right)^2}{\sum_{j=1}^{J} \sum_{i=1}^{I} \hat{\pi}_{ij}}.$$  \hfill (4.3.56)$$

The statistic $X_{DMI}^2$ is very appealing since it is relatively simple to compute. The calculation for $X_{DMI}^2$ does not require any matrix inversion as long as the $C_j$ estimators do not require matrix inversion.

Brier's estimator satisfies this criterion. However, $X_{DMI}^2$ assumes that consistent estimators are available for the covariance matrices and that the Dirichlet-Multinomial model is correct.

Thus far, the cluster sizes within the $i$th subpopulation have been assumed to all be equal to $n_j$. However, this is certainly not always
the case. Consider the sampling scheme presented in the beginning of this section for \( J > 1 \) subpopulations. Instead of assuming that there are an equal number of observations for the clusters within the \( j^{th} \) subpopulation, let \( n_{tj} \) denote the size of the sample for the \( t^{th} \) cluster within the \( j^{th} \) subpopulation. The total sample size for the \( j^{th} \) subpopulation is

\[
N_j = \sum_{t=1}^{S_j} n_{tj},
\]

(4.3.57)

where \( S_j \) denotes the number of sampled clusters within the \( j^{th} \) subpopulation.

Suppose the vector of observations \( X_{tj} \) for the \( t^{th} \) cluster in the \( j^{th} \) subpopulation has a Dirichlet-Multinomial distribution denoted by \( \text{DM}_1(n_{tj}, \pi_j, K_j) \).

\[
X_{tj} = \sum_{j=1}^{S_j} X_{tj} \quad j = 1, 2, \ldots, J;
\]

(4.3.58)

then,

\[
E(X_{tj}) = \sum_{j=1}^{S_j} n_{tj} \pi_jj
\]

\[
= N_j \pi_{tj}.
\]

(4.3.59)

Let the covariance matrix for \( X_{tj} \),
be denoted by \( R_j \), where

\[
C_{tj} = (n_{tj} + K_j)(1 + K_j)^{-1}.
\]

(4.3.61)

Note that the \( n_{tj} \)'s are known and that the cluster parameter \( C_{tj} \) is a function of \( n_{tj} \) and the unknown subpopulation parameter \( K_j \). If

\[
\hat{\pi}_j = N_j^{-1} \pi_j
\]

(4.3.62)

is used to estimate \( \pi_j \), then,

\[
V(\hat{\pi}_j) = N_j^{-2} \left( \sum_{t=1}^{S_j} n_{t} C_{tj} \right) (\Delta_{\hat{\pi}_j} - \pi_j \hat{\pi}_j^T)
\]

\[
= N_j^{-2} R_j (\Delta_{\hat{\pi}_j} - \pi_j \hat{\pi}_j^T) \quad \text{where} \quad R_j = \sum_{t=1}^{S_j} n_{t} C_{tj}.
\]

(4.3.63)

The factor \( C_{tj} \) can be estimated for each \( j^{th} \) subpopulation as long as sufficient clusters of the same size are within the \( j^{th} \) subpopulation. For unequal cluster size the maximum likelihood estimator for \( K_j \) is a possibility since such an estimate does not require equal sample size, for the clusters within the subpopulation. To test the hypothesis

\[
H_0: \pi_j = \pi_0 \quad j = 1, 2, ..., J;
\]

(4.3.64)

for an unknown vector \( \pi_0 \), one needs to estimate the vector \( \pi_j - \pi_0 \) and obtain the variance of that estimator. Denote such an estimator for \( \pi_j - \pi_0 \) by
where $\alpha_k > 0$ and $\sum_{k=1}^{J} \alpha_k = 1$. Then, under $H_0$, 

$$E(\hat{\pi}_j - \pi_0) = \frac{1}{2}$$  \hspace{1cm} (4.3.66)$$

and

$$V(\hat{\pi}_j - \pi_0) = \frac{1}{2} R_j - 2\alpha_j R_j + \sum_{k=1}^{J} \alpha_k^2 R_j.$$  \hspace{1cm} (4.3.67)$$

By reasoning similar to that outlined previously for equal cluster sizes within subpopulations, a Wald test statistic for testing the independence hypothesis is

$$x^2_{DMIU} = \sum_{j=1}^{J} \sum_{i=1}^{\pi_j} \left( \frac{\hat{\pi}_{ij} - \pi_0}{\hat{\pi}} \right)^2.$$  \hspace{1cm} (4.3.68)$$

The derivation is similar to obtaining $x^2_{DMI}$. The factor $\hat{\pi}_j^{-1}$ is

$$n_j^{-1} = n_j R_j \left( \sum_{t=1}^{n_{ij}} \hat{\pi}_{ij} \right).$$  \hspace{1cm} (4.3.69)$$

Here, the $\alpha_j$'s must be chosen such that

$$\alpha_j = \frac{n_j^{-1}}{\sum_{j=1}^{J} n_j^{-1}}.$$  \hspace{1cm} (4.3.70)$$
4.4. Some Examples

In this section, three examples are analyzed using the statistics developed in Sections 4.2 and 4.3. Data for the first example were taken from Morton (1965) and are reproduced in Appendix A, Tables 7.1 and 7.2. The data concern the Greek prose for two authors, Herodotus and Thucydides. These data were also analyzed in Chapter 3 using a Wald test constructed in Section 3.4.

The second example concerns the habitat preferences of some wild turkeys in Iowa. The data for the wild turkeys were obtained from animal ecology researchers at Iowa State University. Tables 7.3 to 7.6 in Appendix A contain the data for the wild turkeys.

Data for the third example were taken from Brier (1980). The data concern the satisfaction of households in the metropolitan and non-metropolitan areas in Minnesota. These data are reproduced in Tables 7.7 to 7.12 of Appendix A.

Example 4.1:

There are two subpopulations (J = 2). One subpopulation pertains to the works of the Greek author Herodotus and the other to the works of another Greek author, Thucydides. From the works of Herodotus, 9 clusters (books) were randomly sampled and a random sample of 200 sentences was taken from each book. The occurrence of einai in these chosen sentences were noted and the sentences categorized according to the number of einai found (no einai, one einai, two einai, three einai or greater than three einai). Einai translates as the Greek word for
the verb 'is'. An examination of the data obtained for the works of
Herodotus shows that in 8 of the 9 clusters, 93 or more percent of the
data are in categories one and two. The remaining cluster has 86 per­
cent of the observations in categories one and two.

From the works of Thucydides, a sample of eight clusters (books)
was selected. From each of the sampled clusters, a sample of 200
sentences was taken and categorized according to the number of einai,
(no einai, one einai, two einai, three einai or more than three
einai). Ninety percent or more of the observations in each cluster
can be found in the first two categories. The clusters for the works
of Thucydides are more alike than the clusters for works of Herodotus.

The interest here is to test whether the styles of the two Greek
authors are similar in their use of einai. The hypothesis is

\[ H_0: \pi_j = \pi_0, \quad j = 1, 2, \]  \hspace{1cm} (4.4.1)

where \( \pi_0 \) is an unknown probability vector of dimension 5 and \( \pi_j \) is
the probability vector of the \( j^{th} \) subpopulation. Though the hypothesis
given here is the same as in Example 3.2, the test statistic and the
assumptions made about the covariance matrix for the \( \pi_j \) are
different. Under a Dirichlet-Multinomial distribution, the estimated
common probability vector is

\[ \hat{\pi}_0 = \left[ \sum_{k=1}^{2} N_k \hat{\Lambda}_k \right]^{-1} \sum_{k=1}^{2} \sum_{j=1}^{2} N_k \hat{\Lambda}_k \pi_j. \]  \hspace{1cm} (4.4.2)
Several methods of estimating $C_j$ were given in Section 4.2. These different methods of obtaining an estimator of $C_j$ are now considered. The factor $C_j$ is a measure of the clustering effect present in the data. When $C_j$ has the value 1, there is no clustering effect and the test statistic is equivalent to the Pearson chi-square test of independence. The true value of $C_j$ can take on values between 1 and 200 (the sample size of the clusters). The estimator $\hat{C}_j$ considered requires equal cluster sizes in the $j^{th}$ subpopulation.

One estimator of $C_j$ is the Brier estimator given in expression (4.2.52) as

$$\hat{C}_{jb} = (I-1)^{-1} (S_j-1)^{-1} \sum_{k=1}^{S_j} \sum_{i=1}^{I} n_j (\hat{\pi}_{ijk} - \hat{\pi}_{ij})^2 \hat{\pi}_{ij}, \quad (4.4.3)$$

where $\hat{\pi}_{ijk}$ is the proportion of observations in the $k^{th}$ cluster of the $j^{th}$ subpopulation. The symbol $I$ denotes the number of categories, $S_j$ is the number of sample clusters in the $j^{th}$ subpopulation and $n_j$ is the sample size of the clusters in the $j^{th}$ subpopulation. The estimator $\hat{C}_{jb}$ measures the similarity between clusters within the same subpopulation. If all clusters are nearly alike in the $j^{th}$ subpopulation, then $\hat{C}_{jb}$ would be very small.

For the works of Herodotus the number of categories is $I = 5$, the number of clusters is $S_1 = 9$, the cluster size is $n_1 = 200$, and the estimated probability vector is

$$\hat{\pi}_1 = (.703889, .237778, .049444, .006111, .002778)^t.$$
Hence, $\hat{C}_{1B}$ is 2.0054. For the works of Thucydides the number of categories is $I = 5$, the number of clusters is $S_2 = 8$, the cluster size is $n_2 = 200$, and the estimated probability vector is

$$\hat{\Pi}_2 = (0.681875, 0.252500, 0.051250, 0.010000, 0.004375)' .$$

Hence, $\hat{C}_{2B}$ is 1.17244.

Recall that the data for Thucydides showed greater similarity among clusters than the data for Herodotus. Hence, it is not surprising that the value for $\hat{C}_{2B}$ (Thucydides) is smaller than the value for $\hat{C}_{1B}$ (Herodotus). Using the values for $\hat{C}_{1B}$ and $\hat{C}_{2B}$, the estimated probability vector is

$$\hat{\Pi}_\infty = \left( \sum_{j=1}^{2} N_j \hat{C}_j^{-1} \right)^{-1} \sum_{j=1}^{2} N_j \hat{C}_j^{-1} \hat{\Pi}_j$$

$$= (0.689996, 0.247085, 0.0505838, 0.00856535, 0.00378587)' .$$

(4.4.4)

The vector $\hat{\Pi}_\infty$ is a linear combination of the vectors $\hat{\Pi}_1$ (Herodotus) and $\hat{\Pi}_2$ (Thucydides). The weights for the linear combination are

$$N_1 C_{1B} \left( \sum_{j=1}^{2} N_j C_j^{-1} \right)^{-1} = 0.3689$$

and
The statistic for testing the equality of the vectors of proportions for the two authors is

\[ X^2_{DMI} = 2 \sum_{j=1}^{2} N_j \hat{C}_{jB} \sum_{i=1}^{5} \left( \hat{\pi}_{ij} - \hat{\pi}_{10} \right)^2 \hat{\pi}_{10} \]

\[ = 1.17875. \quad (4.4.5) \]

The observed value of \( X^2_{DMI} \) is less than the 95th percentile of a chi-square distribution with 4 degrees of freedom. The value of \( X^2_{DMI} \) indicates that the styles of the two authors are not significantly different. However, the method used in the computation of \( \hat{C}_{jB}(j=1,2) \) gives no indication of whether the model fits. Only if the Dirichlet-Multinomial is appropriate can \( X^2_{DMI} \), based on the estimator \( \hat{C}_{jB}(j=1,2) \), be recommended.

A second method of estimating \( C_j \), based on the generalized least squares technique, is now illustrated. The advantage of the method based on a generalized least squares estimator is the fact that the method provides a test for the fit of the model. The estimator of the covariance matrix for a simple random sample of 200 observations from a multinomial distribution with probability vector \( \hat{\pi}_{11} \) for the works of Herodotus is
where the diagonal elements are $\pi_{i1}(1-\pi_{i1})(200)^{-1}$ and the off-diagonal elements are $-\pi_{i1}\pi_{j1}(200)^{-1}$. The $4 \times 4$ matrix $\hat{\Sigma}_{ml}$, is nonsingular. The matrix $\hat{\Sigma}_{ml}$ is made into the vector $\hat{\mathbf{w}}_1$ by taking the elements by rows and ignoring the elements below the diagonal. The vector $\hat{\mathbf{w}}_1$ obtained from $\hat{\Sigma}_{ml}$ forms the right side of the regression equation

$$\hat{\mathbf{w}}_1 = C_{1\mathbf{w}} + \xi,$$  \hspace{1cm} (4.4.7)  

where

$$\hat{\mathbf{w}}_1 = (104.2150, -83.6846, -17.4017, -2.1508, 90.6198, -5.8784, -0.7265, 23.4998, -0.15108, 3.03688) \times 10^{-5}.$$  

The estimated covariance matrix constructed using the cluster sampling formulas is

$$\hat{\Sigma}_{DML} = s^{-1} \sum_{k=1}^{9} (\hat{\pi}_{1k} - \hat{\pi}_1)(\hat{\pi}_{1k} - \hat{\pi}_1)' \cdot$$
The upper left $4 \times 4$ submatrix of $\hat{\Sigma}_{\text{DMl}}$ is

$$\hat{V}_{\text{ppl}} = \begin{bmatrix}
455.4866 & -244.0283 & -161.0066 & -32.3611 \\
-244.0283 & 144.4444 & 74.8611 & 14.9653 \\
-161.0066 & 74.8611 & 68.4028 & 11.3194 \\
-32.3611 & 14.9653 & 11.3194 & 5.4861
\end{bmatrix} \times 10^{-5}.$$ (4.4.8)

A graph of the elements of $\hat{V}_{\text{ppl}}$, denoted by $\hat{V}_1$, versus the elements of $\hat{\Sigma}_{\text{ml}}$, denoted by $\hat{\omega}_1$, is given in Figure 4.1. If the Dirichlet-Multinomial model is satisfied, then a factor $C_1$ exists such that

$$\hat{\Sigma}_{\text{DMl}} = C_1 \hat{\omega}_1 \hat{V}_1,$$

where $1 \leq C_1 \leq 200$. An examination of $\hat{V}_1$ and $\hat{\omega}_1$ shows some negative values for $\hat{\omega}_1$ corresponding to positive values for $\hat{V}_1$. A coordinate pair consists of $(\hat{w}_{1t}, \hat{v}_{1t})$, where $\hat{w}_{1t}$ and $\hat{v}_{1t}$ are elements of the vectors $\hat{\omega}_1$ and $\hat{V}_1$, respectively. Since $C_1$ is positive, the model suggests that the coordinate pairs must either be both negative or both positive. However, in Figure 4.1 some coordinates have a negative $\hat{w}_{1t}$ value and a positive $\hat{v}_{1t}$ value.

The degree to which the estimate of $C$ is affected by these points depends on the coordinates of the point $(\hat{w}_{1t}, \hat{v}_{1t})$ and on the covariance matrix of $\hat{V}_{1t}$. The covariance matrix of the estimated covariances defined in (4.2.66) is given in Table 4.1a. The co-
Figure 4.1. A graph of the elements of the estimated covariance matrix under cluster sampling, $\hat{\Sigma}$, versus the elements of the estimated covariance matrix under Multinomial sampling, $\hat{\Sigma}$, for the works of Herodotus.
ordinate pairs lying in the second quadrant of Figure 4.1 with their variances are given in Table 4.2a.

Table 4.1a. Cluster estimated covariance matrix of sample covariances for works of Herodotus

<table>
<thead>
<tr>
<th>Identification</th>
<th>Cluster covariance estimate x 10^7</th>
<th>Identification</th>
<th>Cluster covariance estimate x 10^7</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 11</td>
<td>97.375</td>
<td>14 14</td>
<td>0.68322</td>
</tr>
<tr>
<td>11 12</td>
<td>-50.932</td>
<td>14 22</td>
<td>-1.8830</td>
</tr>
<tr>
<td>11 13</td>
<td>-34.606</td>
<td>14 23</td>
<td>-1.4346</td>
</tr>
<tr>
<td>11 14</td>
<td>-7.6434</td>
<td>14 24</td>
<td>-0.35786</td>
</tr>
<tr>
<td>11 22</td>
<td>26.336</td>
<td>14 33</td>
<td>-0.98469</td>
</tr>
<tr>
<td>11 23</td>
<td>18.46</td>
<td>14 34</td>
<td>-0.24676</td>
</tr>
<tr>
<td>11 24</td>
<td>3.9494</td>
<td>14 44</td>
<td>-0.04759</td>
</tr>
<tr>
<td>11 33</td>
<td>11.865</td>
<td>22 22</td>
<td>7.5441</td>
</tr>
<tr>
<td>11 34</td>
<td>2.7916</td>
<td>22 23</td>
<td>5.0206</td>
</tr>
<tr>
<td>11 44</td>
<td>0.53985</td>
<td>22 24</td>
<td>0.96585</td>
</tr>
<tr>
<td>12 12</td>
<td>26.953</td>
<td>22 33</td>
<td>3.0865</td>
</tr>
<tr>
<td>12 13</td>
<td>17.947</td>
<td>22 34</td>
<td>0.69064</td>
</tr>
<tr>
<td>12 14</td>
<td>3.8427</td>
<td>22 44</td>
<td>0.13442</td>
</tr>
<tr>
<td>12 22</td>
<td>-14.133</td>
<td>23 23</td>
<td>3.5088</td>
</tr>
<tr>
<td>12 23</td>
<td>-9.6844</td>
<td>23 24</td>
<td>0.74148</td>
</tr>
<tr>
<td>12 24</td>
<td>-1.9840</td>
<td>23 33</td>
<td>2.2310</td>
</tr>
<tr>
<td>12 33</td>
<td>-6.0884</td>
<td>23 34</td>
<td>0.52289</td>
</tr>
<tr>
<td>12 34</td>
<td>-1.4010</td>
<td>23 44</td>
<td>0.10224</td>
</tr>
<tr>
<td>12 44</td>
<td>-0.27346</td>
<td>24 24</td>
<td>0.19057</td>
</tr>
<tr>
<td>13 13</td>
<td>12.381</td>
<td>24 33</td>
<td>0.50548</td>
</tr>
<tr>
<td>13 14</td>
<td>2.7862</td>
<td>24 34</td>
<td>0.12608</td>
</tr>
<tr>
<td>13 22</td>
<td>-9.1918</td>
<td>24 44</td>
<td>0.025252</td>
</tr>
<tr>
<td>13 23</td>
<td>-6.5466</td>
<td>33 33</td>
<td>1.5053</td>
</tr>
</tbody>
</table>
Table 4.1a. (continued)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Cluster covariance estimate $\times 10^7$</th>
<th>Identification</th>
<th>Cluster covariance estimate $\times 10^7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>13 24</td>
<td>-1.4359</td>
<td>33 34</td>
<td>0.36491</td>
</tr>
<tr>
<td>13 33</td>
<td>-4.2809</td>
<td>33 44</td>
<td>0.068234</td>
</tr>
<tr>
<td>13 34</td>
<td>-1.0233</td>
<td>34 34</td>
<td>0.092642</td>
</tr>
<tr>
<td>13 44</td>
<td>-0.19570</td>
<td>34 44</td>
<td>0.016516</td>
</tr>
<tr>
<td></td>
<td></td>
<td>44 44</td>
<td>0.0038697</td>
</tr>
</tbody>
</table>

Table 4.1b. Cluster estimated covariance matrix of sample covariances for the works of Thucydides

<table>
<thead>
<tr>
<th>Identification</th>
<th>Cluster covariance estimate $\times 10^7$</th>
<th>Identification</th>
<th>Cluster covariance estimate $\times 10^7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 11</td>
<td>91.31</td>
<td>14 14</td>
<td>0.32</td>
</tr>
<tr>
<td>11 12</td>
<td>-25.55</td>
<td>14 22</td>
<td>7.00</td>
</tr>
<tr>
<td>11 13</td>
<td>-31.04</td>
<td>14 23</td>
<td>-4.48</td>
</tr>
<tr>
<td>11 14</td>
<td>-5.39</td>
<td>14 24</td>
<td>-1.31</td>
</tr>
<tr>
<td>11 22</td>
<td>-118.66</td>
<td>14 33</td>
<td>1.99</td>
</tr>
<tr>
<td>11 23</td>
<td>75.99</td>
<td>14 34</td>
<td>-0.57</td>
</tr>
<tr>
<td>11 24</td>
<td>22.12</td>
<td>14 44</td>
<td>1.48</td>
</tr>
<tr>
<td>11 33</td>
<td>-33.71</td>
<td>22 22</td>
<td>15.42</td>
</tr>
<tr>
<td>11 34</td>
<td>9.63</td>
<td>22 23</td>
<td>-19.75</td>
</tr>
<tr>
<td>11 44</td>
<td>-25.12</td>
<td>22 24</td>
<td>-28.74</td>
</tr>
<tr>
<td>12 12</td>
<td>7.15</td>
<td>22 33</td>
<td>43.80</td>
</tr>
<tr>
<td>12 13</td>
<td>8.69</td>
<td>22 34</td>
<td>-12.52</td>
</tr>
<tr>
<td>12 14</td>
<td>1.51</td>
<td>22 44</td>
<td>32.64</td>
</tr>
</tbody>
</table>
### Table 4.1b. (continued)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Cluster covariance estimate x 10^7</th>
<th>Identification</th>
<th>Cluster covariance estimate x 10^7</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 22</td>
<td>33.27</td>
<td>23 23</td>
<td>63.24</td>
</tr>
<tr>
<td>12 23</td>
<td>-21.27</td>
<td>23 24</td>
<td>18.41</td>
</tr>
<tr>
<td>12 24</td>
<td>-6.19</td>
<td>23 33</td>
<td>-28.05</td>
</tr>
<tr>
<td>12 33</td>
<td>9.43</td>
<td>23 34</td>
<td>8.02</td>
</tr>
<tr>
<td>12 34</td>
<td>-2.70</td>
<td>23 44</td>
<td>-20.90</td>
</tr>
<tr>
<td>12 44</td>
<td>7.03</td>
<td>24 24</td>
<td>5.36</td>
</tr>
<tr>
<td>13 13</td>
<td>10.55</td>
<td>24 33</td>
<td>-8.17</td>
</tr>
<tr>
<td>13 14</td>
<td>1.83</td>
<td>24 34</td>
<td>2.33</td>
</tr>
<tr>
<td>13 22</td>
<td>40.34</td>
<td>24 44</td>
<td>-6.08</td>
</tr>
<tr>
<td>13 23</td>
<td>-25.83</td>
<td>33 33</td>
<td>12.40</td>
</tr>
<tr>
<td>13 24</td>
<td>-7.52</td>
<td>33 34</td>
<td>-3.56</td>
</tr>
<tr>
<td>13 33</td>
<td>11.46</td>
<td>33 44</td>
<td>9.27</td>
</tr>
<tr>
<td>13 34</td>
<td>-3.27</td>
<td>34 34</td>
<td>1.02</td>
</tr>
<tr>
<td>13 44</td>
<td>8.54</td>
<td>34 44</td>
<td>-2.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>44 44</td>
<td>6.91</td>
</tr>
</tbody>
</table>

### Table 4.2a. Second quadrant points of Figure 4.1 and their standard errors

<table>
<thead>
<tr>
<th>Cell</th>
<th>Points (multiplied by 10^5)</th>
<th>Standard error (multiplied by 10^5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,3</td>
<td>(-5.8784, 66.5432)</td>
<td>59.24</td>
</tr>
<tr>
<td>2,4</td>
<td>(-.72654, 13.3025)</td>
<td>13.81</td>
</tr>
<tr>
<td>3,4</td>
<td>(-1.5108, 10.0617)</td>
<td>9.62</td>
</tr>
</tbody>
</table>
For the works of Thucydides similar statistics can be obtained based on the same procedure as used for the works of Herodotus. There are 8 clusters and 5 categories so $\hat{\Sigma}_2$ has dimension 10 and is singular. The matrix for Thucydides, following (4.4.6), is

$$\hat{\Sigma}_{m2} = \begin{bmatrix}
108.461 & -86.087 & -17.473 & -3.409 \\
-86.087 & 94.372 & -6.470 & -1.262 \\
-17.473 & -6.470 & 24.312 & 0.256 \\
-3.409 & -1.262 & 0.256 & 4.950
\end{bmatrix} \times 10^{-5}.$$

The vector $\hat{w}_2$ is

$$\hat{w}_2 = (108.461, -86.0867, -17.473, -3.4094, 94.3719, -6.4703, -1.2625, 24.3117, -0.25625, 4.95) \times 10^{-5}.$$

The estimated vector $\hat{v}_2$ is

$$\hat{v}_2 = (160.586, -116.719, -29.6094, -5.625, 94.375, 9.0625, 3.4375, 23.5937, 1.875, 0.625) \times 10^{-5}.$$

A graph of $\hat{v}_2$ versus $\hat{w}_2$ is given in Figure 4.2. As in the case of the works of Herodotus there are coordinates of contrasting signs in Figure 4.2. The points of contrasting signs which lie in the second quadrant of Figure 4.2 are given in Table 4.2b. The cluster estimated covariance matrix of sample covariances for the works of Thucydides is given in Table 4.1b.
Figure 4.2. A graph of the elements of the estimated covariance matrix under cluster sampling, $\hat{V}$, versus the elements of the estimated covariance matrix under Multinomial sampling, $\hat{W}$, for the works of Thucydides.
Table 4.2b. Second quadrant points of Figure 4.2 and their standard errors

<table>
<thead>
<tr>
<th>Cell</th>
<th>Points (multiplied by $10^5$)</th>
<th>Standard error (multiplied by $10^5$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,3</td>
<td>(-6.4703, 9.0625)</td>
<td>15.28</td>
</tr>
<tr>
<td>2,4</td>
<td>(-1.2625, 3.4375)</td>
<td>2.30</td>
</tr>
<tr>
<td>3,4</td>
<td>(-2.5625, 1.875)</td>
<td>1.62</td>
</tr>
</tbody>
</table>

The Cholesky decomposition of $\hat{\Sigma}_{ml}^{-1}$ is given by

$$\hat{\Sigma}_{ml}^{-1} = \hat{R}_1^\top \hat{R}_1,$$  \hspace{1cm} (4.4.9)

where

$$\hat{R}_1 = \begin{pmatrix} 268.846 & 267.790 & 267.790 & 267.790 \\ 0 & 33.528 & 8.441 & 8.441 \\ 0 & 0 & 65.244 & 3.246 \\ 0 & 0 & 0 & 181.464 \end{pmatrix}.$$ \hspace{1cm} (4.4.10)

Note that $\hat{R}_1$ defines a transformation that is roughly equivalent to the conditional probabilities. It follows that

$$\hat{R}_1 \hat{\Sigma}_{ml}^{-1} \hat{R}_1^\top = I.$$ \hspace{1cm} (4.4.11)

Let $\hat{V}_{YY1}$ denote the upper left $4 \times 4$ matrix of the estimated matrix.
covariance matrix of the linear combination of $\hat{\Pi}$'s defined by

$$\hat{V}_{t1} = R_1 (\hat{\Pi}_{t1} - \hat{\Pi}_{1}) .$$

(4.4.12)

Then,

$$\hat{V}_{y1} = \hat{V}_{pl} \hat{R}_1 .$$

$$= \begin{pmatrix}
1.0537 & -1.1383 & -1.2395 & -0.3489 \\
-1.1383 & -2.2010 & 3.0975 & 1.1679 \\
-1.2396 & 2.0975 & 2.9603 & 1.3725 \\
-0.3489 & 1.1679 & 1.3725 & 1.8065 \\
\end{pmatrix}$$

(4.4.13)

The generalized least squares estimator defined in (4.2.71) is

$$\hat{C}_{B1} = (A' D_B^{-1} A)^{-1} A' D_B^{-1} H_1$$

$$= 2.0054 ,$$

(4.4.14)

where

$$A' = (1, 0, 0, 0, 1, 0, 0, 1, 0, 1),$$

$$D_B = \text{diag}(2, 1, 1, 1, 2, 1, 2, 1, 2),$$

$$H_1 = \text{vech} \hat{V}_{y1} .$$
The estimator \( \hat{c}_{Bl} \) is identical to that previously computed, as it should be (4.4.3).

The estimated covariance matrix for the estimated covariances of the transformed \( \hat{t}_{Bl} \)'s was computed by applying formula (4.2.66) to the \( \hat{t}_{Bl} \)'s. Thus, for example, the cluster estimated variance of the (1,2)-element of \( \hat{V}_{Yy1} \) is

\[
\hat{V}_{[Yy1, 12]} = 8^{-2} \sum_{j=1}^{9} [\hat{Y}_{1jl} \hat{Y}_{2jl} - \hat{V}_{Yy1, 12}]^2
\]

\[
= 1.3063 . \tag{4.4.15}
\]

The matrix of estimated covariances of the estimated covariances is given in Table 4.4. The cluster standard errors of the estimated covariances for the transformed variables are given in the last column of Table 4.3.

Table 4.3. Transformed estimated covariance matrix for proportions of Herodotus

<table>
<thead>
<tr>
<th>Identification</th>
<th>Model estimate of ( V_{Yy1} )</th>
<th>Elements of ( \hat{V}_{Yy1} )</th>
<th>Model standard error</th>
<th>Cluster standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>2.005</td>
<td>1.054</td>
<td>1.406</td>
<td>0.470</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>-1.138</td>
<td>0.994</td>
<td>0.887</td>
</tr>
<tr>
<td>31</td>
<td>0</td>
<td>-1.240</td>
<td>0.994</td>
<td>1.002</td>
</tr>
<tr>
<td>41</td>
<td>0</td>
<td>-0.349</td>
<td>0.994</td>
<td>0.699</td>
</tr>
<tr>
<td>22</td>
<td>2.005</td>
<td>2.201</td>
<td>1.406</td>
<td>1.496</td>
</tr>
</tbody>
</table>
Table 4.3. (continued)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Model estimate of $\hat{v}_{YY1}$</th>
<th>Elements of $\hat{v}_{YY1}$</th>
<th>Model standard error</th>
<th>Cluster standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>0</td>
<td>2.098</td>
<td>0.994</td>
<td>1.665</td>
</tr>
<tr>
<td>42</td>
<td>0</td>
<td>1.168</td>
<td>0.994</td>
<td>1.070</td>
</tr>
<tr>
<td>33</td>
<td>2.005</td>
<td>2.960</td>
<td>1.406</td>
<td>1.794</td>
</tr>
<tr>
<td>43</td>
<td>0</td>
<td>1.372</td>
<td>0.994</td>
<td>1.219</td>
</tr>
<tr>
<td>44</td>
<td>2.005</td>
<td>1.807</td>
<td>1.406</td>
<td>0.687</td>
</tr>
</tbody>
</table>

Table 4.4. Cluster estimated covariance matrix of the transformed covariances for Herodotus

<table>
<thead>
<tr>
<th>Identification</th>
<th>Estimated covariance</th>
<th>Identification</th>
<th>Estimated covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,11</td>
<td>0.2207</td>
<td>14,14</td>
<td>0.4885</td>
</tr>
<tr>
<td>11,12</td>
<td>-0.4093</td>
<td>14,22</td>
<td>-0.9233</td>
</tr>
<tr>
<td>11,13</td>
<td>-0.4617</td>
<td>14,23</td>
<td>-1.0811</td>
</tr>
<tr>
<td>11,14</td>
<td>-0.3148</td>
<td>14,24</td>
<td>-0.7003</td>
</tr>
<tr>
<td>11,22</td>
<td>0.6730</td>
<td>14,33</td>
<td>-1.1917</td>
</tr>
<tr>
<td>11,23</td>
<td>0.7713</td>
<td>14,34</td>
<td>-0.8158</td>
</tr>
<tr>
<td>11,24</td>
<td>0.4849</td>
<td>14,44</td>
<td>-0.3852</td>
</tr>
<tr>
<td>11,33</td>
<td>0.8283</td>
<td>22,22</td>
<td>2.2391</td>
</tr>
<tr>
<td>11,34</td>
<td>0.5437</td>
<td>22,23</td>
<td>2.4595</td>
</tr>
<tr>
<td>11,44</td>
<td>0.2905</td>
<td>22,24</td>
<td>1.3863</td>
</tr>
<tr>
<td>12,12</td>
<td>0.7867</td>
<td>22,33</td>
<td>2.5387</td>
</tr>
<tr>
<td>12,13</td>
<td>0.8601</td>
<td>22,34</td>
<td>1.5925</td>
</tr>
<tr>
<td>12,14</td>
<td>0.5775</td>
<td>22,44</td>
<td>0.8549</td>
</tr>
</tbody>
</table>
Letting $\hat{D}_{W1}$ be the diagonal matrix composed of the diagonal elements of the matrix in Table 4.4, the estimated generalized least squares estimator defined in (4.2.74) is

$$\hat{C}_{W1} = (A'_{W1} D_{W1}^{-1} A_{W1})^{-1} A'_{W1} D_{W1}^{-1}$$

$$= 1.418 , \tag{4.4.16}$$

<table>
<thead>
<tr>
<th>Identification</th>
<th>Estimated covariance</th>
<th>Identification</th>
<th>Estimated covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>12,22</td>
<td>-1.3234</td>
<td>23,23</td>
<td>2.7737</td>
</tr>
<tr>
<td>12,23</td>
<td>-1.4678</td>
<td>23,24</td>
<td>1.6559</td>
</tr>
<tr>
<td>12,24</td>
<td>-0.8609</td>
<td>23,33</td>
<td>2.9301</td>
</tr>
<tr>
<td>12,33</td>
<td>-1.5467</td>
<td>23,34</td>
<td>1.8996</td>
</tr>
<tr>
<td>12,34</td>
<td>-0.9888</td>
<td>23,44</td>
<td>1.0137</td>
</tr>
<tr>
<td>12,44</td>
<td>-0.5171</td>
<td>24,24</td>
<td>1.1452</td>
</tr>
<tr>
<td>13,13</td>
<td>1.0042</td>
<td>24,33</td>
<td>1.8353</td>
</tr>
<tr>
<td>13,14</td>
<td>0.6570</td>
<td>24,34</td>
<td>1.2567</td>
</tr>
<tr>
<td>13,22</td>
<td>-1.4299</td>
<td>24,44</td>
<td>0.6849</td>
</tr>
<tr>
<td>13,23</td>
<td>-1.6322</td>
<td>33,33</td>
<td>3.2191</td>
</tr>
<tr>
<td>13,24</td>
<td>-1.0015</td>
<td>33,34</td>
<td>2.1412</td>
</tr>
<tr>
<td>13,33</td>
<td>-1.7060</td>
<td>33,44</td>
<td>1.1021</td>
</tr>
<tr>
<td>13,34</td>
<td>-1.1174</td>
<td>34,34</td>
<td>1.4870</td>
</tr>
<tr>
<td>13,44</td>
<td>-0.6202</td>
<td>34,44</td>
<td>0.7329</td>
</tr>
<tr>
<td></td>
<td></td>
<td>44,44</td>
<td>0.4721</td>
</tr>
</tbody>
</table>
where $A$ and $B$ are defined following (4.4.14). The estimated variance of $\hat{C}_W$ is

$$\hat{V}(\hat{C}_W) = (A'D_W^{-1}A)^{-1}$$

$$= 0.135 \quad (4.4.17)$$

and the lack-of-fit statistic is

$$\chi^2_W = H_1[D_W^{-1} - A(A'D_W^{-1}A)^{-1} A']H_1$$

$$= 9.40 . \quad (4.4.18)$$

The estimate of $C_\perp$ obtained using the matrix $D_W$ is smaller than that obtained using the matrix $D_B$, because the estimated variances for $\hat{V}_{YY1,1}$ and $\hat{V}_{YY1,4}$ are small.

One can compute the estimated variance of $\hat{C}_B$ in several ways. The method that uses the fewest assumptions is to compute the variance as

$$\hat{V}(\hat{C}_B) = (A'D_B^{-1}A)^{-1} A'D_B^{-1} [\text{vech } \hat{V}_{YY1}]D_B^{-1} A(A'D_B^{-1}A)^{-1} , \quad (4.4.19)$$

where the elements of $\hat{V}_{YY1}$ are given in Table 4.4. For Herodotus we have

$$\hat{V}(\hat{C}_B) = 1.17 .$$
In the computation of this variance the unknown $\Sigma_{ml}$ is treated as known. A more appropriate approximation is given in Appendix B. A second variance estimator can be computed under the assumption that $\text{var}[\text{vech } \hat{V}_{YY}]$ is a diagonal matrix proportional to $D_B$. For Herodotus, the ratio of the diagonal elements of $\hat{V}_{YY}$ to the diagonal elements of $D_B$ is 0.9883. The estimated standard errors for the elements of $\hat{V}_{YY}$ constructed under this model are given in the fourth column of Table 4.3. If we use this form for the estimated variance of the elements of $\hat{V}_{YY}$, we have

$$\hat{V}[\hat{C}_{Bl}] = 0.4942$$

and the standard error of $\hat{C}_{Bl}$ is 0.703. On this basis one would conclude that $C_{Bl}$ is not one. This computation treats $\Sigma_{ml}$ as a known matrix. The estimation of $\Sigma_{ml}$ adds terms of order $I^{-1}$ to the variance and the estimated variance in (4.4.15) is biased. More details are presented in Appendix B.

The 10 x 10 matrix $\text{var}[\text{vech } \hat{V}_{YY}]$ is singular, since the number of clusters is less than 10. Therefore, we construct a lack-of-fit test for the model using the diagonal matrix for the covariance matrix of the estimated covariances. This diagonal matrix allows a non-singular matrix in the computation. If $\Sigma_{ml}$ is known, and if $V[\text{vech } \hat{V}_{YY}]$ is of the specified diagonal form, the quantity
\[ X_{B1}^2 = (H_{A1}^{-1} - C_{B1}^T A_{B1}^{-1} A_{B1} C_{B1}) (0.9883)^{-1} \]

\[ = 11.69 \quad (4.4.20) \]

is approximately distributed as a chi-square random variable with nine degrees of freedom when the null model is true. On the basis of this statistic the model is accepted at the ten percent level. Because \( \Sigma_{m1} \) is estimated for the transformation, the statistic is biased.

Similar statistics can be obtained for the works of Thucydides. The Cholesky decomposition of \( \Lambda_{m2}^{-1} \) is given by

\[ \Lambda_{m2}^{-1} = R_2^T R_2, \]

where

\[
\begin{bmatrix}
214.494 & 213.126 & 213.126 & 213.126 \\
0 & 32.917 & 8.854 & 8.854 \\
0 & 0 & 64.152 & 3.321 \\
0 & 0 & 0 & 142.134
\end{bmatrix}
\]

The estimated covariance matrix \( \hat{V}_{YY2} \) following expression (4.4.13) is

\[
\begin{bmatrix}
2.14958 & -0.74127 & 0.74163 & 0.09569 \\
-0.74127 & 1.27698 & 0.38917 & 0.21976 \\
0.74163 & 0.38917 & 1.11892 & 0.19876 \\
0.09569 & 0.21976 & 0.19876 & 0.1443
\end{bmatrix}
\]
The generalized least squares estimator defined in (4.2.71) and following (4.4.14) is

$$\hat{C}_{B2} = 1.17244.$$ 

The matrix of estimated covariances of $\hat{V}_{yy2}$ is given in Table 4.6. The cluster standard errors of the estimated covariances for the transformed random vectors,

$$\hat{V}_{t2} = R_{t2}(\hat{V}_{t2} - \hat{V}_{2}),$$

are given in the last column of Table 4.5. Letting $\hat{D}_{W2}$ be the diagonal elements of the matrix in Table 4.6, the estimated generalized least squares estimator defined in (4.2.74) is

$$\hat{C}_{W2} = 0.369107.$$ 

The estimated variance of $\hat{C}_{W2}$ is

$$\hat{V}(\hat{C}_{W2}) = 0.0062897$$

and the lack-of-fit statistic is

$$X_{W2}^2 = 50.7521.$$ 

Comparing this to percentiles of the chi-square distribution with nine
Table 4.5. Transformed estimated covariance matrix for proportions for Thucydides

<table>
<thead>
<tr>
<th>Identification</th>
<th>Model estimate of $V_{YY2}$</th>
<th>Elements of $\hat{V}_{YY2}$</th>
<th>Model standard error</th>
<th>Cluster standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>1.175</td>
<td>2.161</td>
<td>0.462</td>
<td>0.813</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>-0.743</td>
<td>0.327</td>
<td>0.484</td>
</tr>
<tr>
<td>31</td>
<td>0</td>
<td>0.744</td>
<td>0.327</td>
<td>0.271</td>
</tr>
<tr>
<td>41</td>
<td>0</td>
<td>0.096</td>
<td>0.327</td>
<td>0.083</td>
</tr>
<tr>
<td>22</td>
<td>1.175</td>
<td>1.277</td>
<td>0.462</td>
<td>0.198</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>0.389</td>
<td>0.327</td>
<td>0.410</td>
</tr>
<tr>
<td>42</td>
<td>0</td>
<td>0.220</td>
<td>0.327</td>
<td>0.135</td>
</tr>
<tr>
<td>33</td>
<td>1.175</td>
<td>1.119</td>
<td>0.462</td>
<td>0.511</td>
</tr>
<tr>
<td>43</td>
<td>0</td>
<td>0.199</td>
<td>0.327</td>
<td>0.160</td>
</tr>
<tr>
<td>44</td>
<td>1.175</td>
<td>0.144</td>
<td>0.462</td>
<td>0.088</td>
</tr>
</tbody>
</table>

Table 4.6. Cluster estimated covariance matrix of the transformed covariances for Thucydides

<table>
<thead>
<tr>
<th>Identification</th>
<th>Estimated covariance</th>
<th>Identification</th>
<th>Estimated covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,11</td>
<td>.66044</td>
<td>14,14</td>
<td>.00691</td>
</tr>
<tr>
<td>11,12</td>
<td>-.28774</td>
<td>14,22</td>
<td>.00247</td>
</tr>
<tr>
<td>11,13</td>
<td>.12305</td>
<td>14,23</td>
<td>.000639</td>
</tr>
<tr>
<td>11,14</td>
<td>-.01576</td>
<td>14,24</td>
<td>.008436</td>
</tr>
<tr>
<td>11,22</td>
<td>-.04060</td>
<td>14,33</td>
<td>-.009998</td>
</tr>
<tr>
<td>11,23</td>
<td>-.19222</td>
<td>14,34</td>
<td>.001022</td>
</tr>
<tr>
<td>11,24</td>
<td>-.04603</td>
<td>14,44</td>
<td>.005178</td>
</tr>
<tr>
<td>11,33</td>
<td>-.15826</td>
<td>22,22</td>
<td>.03909</td>
</tr>
</tbody>
</table>
Table 4.6. (continued)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Estimated covariance</th>
<th>Identification</th>
<th>Estimated covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,34</td>
<td>-.05041</td>
<td>22,23</td>
<td>.029080</td>
</tr>
<tr>
<td>11,44</td>
<td>-.03072</td>
<td>22,24</td>
<td>.002093</td>
</tr>
<tr>
<td>12,12</td>
<td>.23384</td>
<td>22,33</td>
<td>.0086102</td>
</tr>
<tr>
<td>12,13</td>
<td>-.01893</td>
<td>22,34</td>
<td>-.0012914</td>
</tr>
<tr>
<td>12,14</td>
<td>.005879</td>
<td>22,44</td>
<td>.001200</td>
</tr>
<tr>
<td>12,22</td>
<td>.00468</td>
<td>23,23</td>
<td>.16828</td>
</tr>
<tr>
<td>12,23</td>
<td>.13843</td>
<td>23,24</td>
<td>.0019745</td>
</tr>
<tr>
<td>12,24</td>
<td>.03811</td>
<td>23,33</td>
<td>.182903</td>
</tr>
<tr>
<td>12,33</td>
<td>.12989</td>
<td>23,34</td>
<td>.034102</td>
</tr>
<tr>
<td>12,34</td>
<td>.02371</td>
<td>23,44</td>
<td>.0138532</td>
</tr>
<tr>
<td>12,44</td>
<td>.02444</td>
<td>24,24</td>
<td>.01820</td>
</tr>
<tr>
<td>13,13</td>
<td>.07364</td>
<td>24,33</td>
<td>.0130346</td>
</tr>
<tr>
<td>13,14</td>
<td>-.00588</td>
<td>24,34</td>
<td>.015413</td>
</tr>
<tr>
<td>13,22</td>
<td>-.016009</td>
<td>24,44</td>
<td>.011892</td>
</tr>
<tr>
<td>13,23</td>
<td>-.009870</td>
<td>33,33</td>
<td>.26108</td>
</tr>
<tr>
<td>13,24</td>
<td>-.015078</td>
<td>33,34</td>
<td>.043777</td>
</tr>
<tr>
<td>13,33</td>
<td>.039387</td>
<td>33,44</td>
<td>.010308</td>
</tr>
<tr>
<td>13,34</td>
<td>-.015040</td>
<td>34,34</td>
<td>.02555</td>
</tr>
<tr>
<td>13,44</td>
<td>-.009977</td>
<td>34,44</td>
<td>.010756</td>
</tr>
<tr>
<td></td>
<td></td>
<td>44,44</td>
<td>.00781</td>
</tr>
</tbody>
</table>

degrees of freedom suggests the model is not a good fit. The estimate of $C_2$ obtained using the matrix $D_{W2}$ is smaller than that obtained using the matrix $D_B$ in (4.4.14). The estimate of $C_2$ obtained using the matrix $D_{W2}$ is less than one (outside the allowable range for $C_2$).
Recall that the clusters were more alike in the works of Thucydides than in the works of Herodotus.

One possible estimate of the variance of \( \hat{C}_{B2} \), following expression (4.4.19), is

\[
\hat{V}[\hat{C}_{B2}] = 0.034345.
\]

For the works of Thucydides, the ratio of the diagonal elements of \( \hat{V}[\text{vech } \hat{v}_{YY2}] \) to the diagonal elements of \( \hat{D}_B = 0.10677 \). This ratio of diagonal elements for the works of Thucydides differs greatly from one, hence it is not surprising to find that the estimators \( \hat{C}_{B2} \) and \( \hat{C}_{W2} \) differ a great deal. The estimated standard errors for the elements of \( \hat{v}_{YY2} \) constructed under this model are given in the fourth column of Table 4.5. The alternative method which assumes \( V[\text{vech } \hat{v}_{YY2}] \) is a diagonal matrix proportional to \( \hat{D}_B \) yields

\[
\hat{V}[\hat{C}_{B2}] = 0.053385,
\]

and the standard error of \( \hat{C}_{B2} \) is 0.23105. The standard error for \( \hat{C}_{B2} \) suggests that the clustering factor does not differ from one. A lack-of-fit test for the model was calculated. Similar to that in expression (4.4.20), but with 0.10677 replacing 0.9883. The test statistic value obtained, following (4.4.20), is

\[
\chi^2_{B2} = 22.1105
\]

which is approximately distributed as a chi-square random variable.
with nine degrees of freedom when the null model is true. On the basis of this statistic the model was rejected at the one percent level.

Recall that the hypothesis of interest, \( H_0: \pi_j = \pi_0 \) in (4.4.1) can be tested by

\[
X^2_{DMI} = \sum_{j=1}^{2} N_j \left( \sum_{i=1}^{5} \frac{(\hat{\pi}_i - \pi_0)}{\pi_0} \right)^2
\]

as defined in (4.3.45). For each combination of \( C_j \) estimators \( (j = 1 \) Herodotus and \( j = 2 \) Thucydides) the test statistic \( X^2_{DMI} \) was calculated. For the \( \hat{C}_W \) combination (that is \( \hat{C}_{W1} \) and \( \hat{C}_{W2} \) that estimator \( \hat{C}_{W2} \) was truncated. The value of \( \hat{C}_{W2} \) used in the computation of \( X^2_{DMI} \) is one since the value of \( \hat{C}_{W2} \) obtained is less than one.

Table 4.7 gives the set of values obtained for the statistic \( X^2_{DMI} \). However, \( X^2_{DMI} \) is only useful if the model fits and such is not the case for the works of Thucydides.

<table>
<thead>
<tr>
<th>C estimators</th>
<th>Herodotus</th>
<th>Thucydides</th>
<th>( X^2_{DMI} )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{C}_B )</td>
<td>2.0054</td>
<td>1.17244</td>
<td>1.17875</td>
<td>( p &gt; .750 )</td>
</tr>
<tr>
<td>( \hat{C}_W )</td>
<td>1.4181</td>
<td>(0.3691)</td>
<td>1.55195</td>
<td>( p &gt; .750 )</td>
</tr>
<tr>
<td>( \hat{C}_P )</td>
<td>1.0</td>
<td>1.0</td>
<td>1.88373</td>
<td>( p &gt; .750 )</td>
</tr>
</tbody>
</table>
In Table 4.7, the estimator \( \hat{C}_p \) indicates that the statistic \( \chi^2_{DMI} \) is calculated with both \( C_j \) estimators equal to one. When \( \hat{C}_p \) is used to calculate the statistic, \( \chi^2_{DMI} \), the resulting value is equivalent to the Pearson test statistic for independence (under the multinomial assumption). In Table 4.7, the p-values are computed for a chi-square distribution with four degrees of freedom. Since under cluster sampling the \( C_j \)'s cannot be smaller than one, \( \chi^2_{DMI} \) must be smaller than the corresponding Pearson test statistic (computed under the incorrect assumption of simple random sampling).

The major problem in obtaining weighted least estimates for the \( C_j \)'s in this example is the small number of clusters for each author. The most obvious problem is the estimation of the covariance matrix for \( \hat{V}_j \). The small number of clusters was not adequate for computing a simple moment estimator for this covariance matrix. Some additional structure was imposed to limit the number of parameters that had to be estimated to obtain the estimate of this matrix. In particular, some normality assumptions were made. While the Multivariate Central Limit Theorem can be used to justify the normality assumptions for large numbers of clusters, the resulting approximation for the form of the covariance matrix for \( \hat{V}_j \) may not be very good for small numbers of clusters. A further complication is that some categories contained very few observations for all clusters. This resulted in some small values in the estimated covariance matrix for \( \hat{V}_j \). The generalized least square estimators were rather sensitive to differences in these small values appearing in the various estimators used for the covariance matrix for
Another potential problem is that the use of the chi-square approximation for the goodness-of-fit tests is also based on having large numbers of clusters.

Example 4.2:

An experiment was conducted involving wild turkeys and their habitat preferences in a certain area in Iowa. The data, obtained from Animal Ecology Researchers at Iowa State University, are given in Appendix A, Tables 7.3 to 7.6.

During one winter season, samples of four different kinds of wild turkeys were obtained. The captured turkeys were fitted with transmitters and observed repeatedly at random times throughout the season. The turkeys were obtained from four subpopulations: juvenile male, adult male, juvenile female, and adult female.

For the purpose of the experiment, the area in which these turkeys lived was broken up into seven habitat categories. These seven categories are brush, conifer, beans-corn, grass, lowland, oakpole, and oaksaw. The four age/sex categories are considered as subpopulations, and the birds within each subpopulation treated as clusters. For subpopulation 1 (juvenile males) there are 5 turkeys. For subpopulation 2 (adult males) there are 4 turkeys. For subpopulation 3 (juvenile females) there are 10 turkeys. For subpopulation 4 (adult females) there are 21 turkeys. In this example, each turkey is considered as one cluster.
The hypothesis of interest in this case is

$$H_0: \prod_{j=1}^{n_j} = \prod_0 \quad \text{(unknown)} \quad j = 1,2,3,4.$$  \hspace{1cm} (4.4.21)

The statistics used to test this hypothesis assume a Dirichlet-Multinomial model as discussed in Sections 4.2 and 4.3. However, the Dirichlet-Multinomial model, as given in Sections 4.2 and 4.3, requires equal sample sizes for clusters within the same subpopulation. In this turkey data set, each cluster has $n_j = 50 \ (j = 1,2,3,4)$ observations distributed over the 7 categories of interest. Let $N_j$ denote the size of the $j$th subpopulation ($j = 1,2,3,4$), then

$$N_1 = 250, \ N_2 = 200, \ N_3 = 500 \quad \text{and} \quad N_4 = 1050.$$  

Let the total sample size for this example of 40 turkeys be denoted by $N$ then,

$$N = 2000.$$  

Let $\alpha_j$ denote the ratio of the $j$th subpopulation sample size, $N_j$, to the total sample size $N$. Then, the ratios

$$\alpha_j = \frac{N_j}{N} \quad j = 1,2,3,4,$$

are

$$\alpha_1 = .125, \ \alpha_2 = .100, \ \alpha_3 = .250 \quad \text{and} \quad \alpha_4 = .525.$$
The estimated probability vectors for the four subpopulations are

\[ \hat{\pi}_1 = (.020, .112, .308, .152, .016, .312, .08)\],
\[ \hat{\pi}_2 = (.030, .015, .250, .080, .070, .555, 0)\],
\[ \hat{\pi}_3 = (.036, .140, .294, .144, .028, .290, .068)\],
and
\[ \hat{\pi}_4 = (.0389052, .155238, .239048, .103810, .0457143, .349524, .0685714)\].

The test statistic used to test the hypothesis that the habitat preferences are the same for each subpopulation,

\[ H_0: \pi_j = \pi_\infty \quad (j = 1, 2, 3, 4), \]

is

\[ X^2_{DMI} = \sum_{j=1}^{4} N_j C_j^{-1} \sum_{i=1}^{\Lambda-1} (\hat{\pi}_{ij} - \hat{\pi}_{i\infty})^2 \hat{\pi}_{i\infty}, \quad (4.4.22) \]

where the factor \( \hat{\pi}_j \) for the \( j \)th subpopulation is one of the estimators considered in Section 4.2 and

\[ \hat{\pi}_{i\infty} = \sum_{j=1}^{4} N_j \hat{\pi}_j^{-1} \hat{\pi}_{ij} \left[ \sum_{j=1}^{4} N_j \hat{\pi}_j^{-1} \right]^{-1}. \quad (4.4.23) \]

The different methods of obtaining an estimate of \( C_j \) in Section
4.2 are now considered. First consider the estimator given in expres-

\[
\hat{C}_{jB} = (I-1)^{-1}(S_j-1)^{-1} \sum_{k=1}^{S_j} \sum_{i=1}^{I} n_j (\hat{\pi}_{ijk} - \hat{\pi}_{ij})^2 \hat{\pi}_{ij}^{-1}
\]

(4.4.24)

where \( \hat{\pi}_{ijk} \) is the proportion of observations in the \( k \)th cluster of
the \( j \)th subpopulation. The symbol \( I \) denotes the number of habitat
categories, \( S_j \) is the number of turkeys (clusters) in the \( j \)th sub-
population, and \( n_j \) is the number of observations taken on each turkey
in the \( j \)th subpopulation. The numerical values for \( \{\hat{C}_{jB}\} \) for the
subpopulations are

\[
\hat{C}_{1B} = 3.1717, \quad \hat{C}_{2B} = 3.8863, \quad \hat{C}_{3B} = 3.0609, \quad \text{and} \quad \hat{C}_{4B} = 4.2013.
\]

Then,

\[
\hat{\alpha}_o = \frac{1}{4} \sum_{j=1}^{4} N_j \hat{\alpha}_{jB}^{-1} \left[ \sum_{j=1}^{4} N_j \hat{\alpha}_{jB}^{-1} \right]^{-1}
\]

\[
= (0.3408, 0.1311, 0.2666, 0.1206, 0.0384, 0.3456, 0.0636)^
\]

and the value of the test statistic is

\[
\chi^2_{DMI} = 26.9353 .
\]

(4.2.25)

Based on the 95th percentile of the chi-square distribution with 18
degrees of freedom, the p-value for the test lies between .10 and .05.

However, in computing \( \hat{C}_{jB} \) \( (j = 1,2,3,4) \) based on Brier (1980) technique
no assessment was made of whether the Dirichlet-Multinomial model is appropriate for each of the four subpopulations.

In Example 4.1, the clusters in both the works of Herodotus and Thucydides had most of their data in the first two categories with few observations in the other three categories. This combined with small numbers of clusters resulted in generalized least squares estimators which differed somewhat for the various estimation methods. An examination of the turkey data in Appendix A, Tables 7.3 to 7.6, reveals that for adult males most of the observations occur in categories 3 and 6. In the case of juvenile males, the observations were more evenly spread out, but mostly occur in five of the seven categories. The observations for juvenile females are also concentrated in five of the seven categories, while the data for adult females are more evenly spread out over all seven categories. Consequently, it might be expected that the various generalized least squares estimators for $C_j$ would more nearly be consistent for the adult females than the adult males. This expectation considers the low number of adult male turkeys sampled as well as the fact that some categories had few observations. There is a much larger number of adult female turkeys.

The second method of estimating the parameter $C_j$ based on the generalized least squares technique is now illustrated. As discussed in Example 4.1, one advantage of the method based on generalized least squares is that the method provides a test for the fit of the model. The estimator of the covariance matrix for a simple random sample of
50 observations from a multinomial distribution with probability vector $\frac{\hat{\pi}}{\pi_1}$ for the juvenile males is


where the diagonal elements are $\hat{\pi}_{11}(1-\hat{\pi}_{11})(50)^{-1}$ and the off diagonal elements are $-\hat{\pi}_{i1}\hat{\pi}_{j1}(50)^{-1}$. The 6x6 matrix $\hat{\Sigma}_{\text{ml}}$ is nonsingular.

The matrix $\hat{\Sigma}_{\text{ml}}$, is made into the vector $\hat{\mathbf{\omega}}_1$ by taking the elements by rows and ignoring the elements below the diagonal. This vector forms the right side of the regression equation

$$\hat{\mathbf{\omega}}_1 = C_1 \hat{\mathbf{\omega}}_1 + \xi,$$

(4.4.26)

where

\[
\hat{\Sigma}_{\text{DM1}} = 4^{-1} \sum_{k=1}^{5} (\hat{\pi}_{1k} - \hat{\pi}_{11})(\hat{\pi}_{1k} - \hat{\pi}_{11})', \quad (4.4.27)
\]

The upper left 6x6 submatrix of \( \hat{\Sigma}_{\text{DM1}} \) is

\[
\begin{pmatrix}
40 & -70 & -80 & -100 & -10 & -110 \\
-70 & 352 & 188 & 192 & -4.0 & -258 \\
-80 & 188 & 12672 & -1952 & -216 & -12 \\
-100 & 192 & -1952 & 2192 & 236 & 548 \\
-10 & -4.0 & -216 & 236 & 28 & -54 \\
-110 & -258 & -12 & -548 & -54 & 412 \\
\end{pmatrix} \times 10^{-5}
\]

The matrix \( \hat{V}_{\text{PPL}} \) is made into the vector \( \hat{V}_{\text{m}} \) by taking the elements by rows and ignoring the elements below the diagonal. If the Dirichlet-Multinomial model is satisfied, then a factor \( C_1 \) exists such that

\[
\hat{\Sigma}_{\text{DM1}} \sim C_1 \hat{\Sigma}_{\text{m}}, \quad (4.4.28)
\]

where \( 1 \leq C_1 \leq 50 \). An examination of \( \hat{V}_{\text{m}} \) and \( \hat{w}_{\text{m}} \) shows some negative values for \( \hat{w}_{\text{m}} \) corresponding to positive values for \( \hat{V}_{\text{m}} \). A coordinate pair in Figure 4.3 consists of \( (\hat{w}_{lt}, \hat{V}_{lt}) \), where \( \hat{w}_{lt} \) and \( \hat{V}_{lt} \) are elements of the vectors \( \hat{w}_l \) and \( \hat{V}_l \), respectively. Since \( C_1 \) is positive, the model suggests that the points in Figure 4.3 should fall along a straight line with a positive slope passing through the origin. Consequently, the points tend to lie in the first and third quadrants. However, in Figure 4.3, four of the 21 points have contrast-
Figure 4.3. A graph of the elements of the estimated covariance matrix under cluster sampling, \( \hat{\Sigma} \), versus the elements of the estimated covariance matrix under Multinomial sampling, \( \hat{\Sigma}_m \), for juvenile male turkeys.
ing signs. The degree to which such second quadrant points affect
the estimation of $C_1$ depends on the covariance matrix for $\hat{V}_1$.
The covariance matrix of the estimated covariances defined in (4.2.66)
is of dimension 21. The coordinate pairs lying in the second quadrant
of Figure 4.3 with their variances are given in Table 4.8a.

Table 4.8a. Second quadrant points in Figure 4.3 and their standard
errors

<table>
<thead>
<tr>
<th>Cell</th>
<th>Points (multiplied by $10^5$)</th>
<th>Standard error (multiplied by $10^5$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,6</td>
<td>(-12.48, 110.0)</td>
<td>38.329</td>
</tr>
<tr>
<td>2,3</td>
<td>(-68.992, 188.0)</td>
<td>249.014</td>
</tr>
<tr>
<td>2,4</td>
<td>(-34.048, 192.0)</td>
<td>209.361</td>
</tr>
<tr>
<td>4,5</td>
<td>(-4.864, 236.0)</td>
<td>116.850</td>
</tr>
</tbody>
</table>

The second quadrant points have relatively high variances for
$\hat{V}_{1t}$, in comparison to points lying in the first and third quadrants.
However, the second quadrant point for cell (1,6) has relatively small
variance and corresponds to categories with few observations.

Similar statistics can be obtained for the adult male turkeys.
There are 4 clusters and 7 categories. Following the computation of
(4.4.6), the upper left 6x6 submatrix of the covariance matrix for
the sample proportions, incorrectly assuming simple random sampling
and a multinomial model, is
This matrix is singular because no adult male turkeys were observed in the seventh habitat category. A nonsingular matrix $\hat{\Sigma}_{n2}$ is obtained by taking the upper left 5x5 submatrix.

The vector $\hat{w}^2$, formed from $\hat{\Sigma}_{m2}$ is

$$\hat{w}^2 = \begin{pmatrix} 58.2, -0.9, -15.0, -4.8, -4.2, 29.55, -7.5, -2.4, -2.1, 375.0, -40.0, -35.0, -40.0, -11.2, 147.2, -11.2, 130.2 \end{pmatrix} \times 10^{-5}.$$ 

Similarly,

$$\hat{v}^2 = \begin{pmatrix} 120, -20, 480, -266.67, 133.33, 3.67, -14.0, -8.0, -7.333, 204.0, -88.0, 60.0, 117.333, -2.667, 28.0 \end{pmatrix} \times 10^{-5}$$

is obtained from the upper left 5x5 submatrix of $\hat{\Sigma}_{m2}$.

A graph of $\hat{v}^2$ versus $\hat{w}^2$ is given in Figure 4.4. As in the case of juvenile male turkeys, there are coordinates of contrasting signs which lie in the second quadrant of Figure 4.4. Those points are listed in Table 4.8b. Two of the 2nd quadrant points have relatively small variances and correspond to categories 1 and 2 where the observed frequencies are small. This is a primary cause of the differences.
Figure 4.4. A graph of the elements of the estimated covariance matrix under cluster sampling, $\hat{\Sigma}$, versus the elements of the estimated covariance matrix under Multinomial sampling, $\hat{\Psi}$, for the adult male turkeys.
that are later observed for the various generalized least squares estimates for \( C_2 \).

Table 4.8b. Second quadrant points of Figure 4.4 and their standard errors

<table>
<thead>
<tr>
<th>Cell</th>
<th>Points (multiplied by ( 10^5 ))</th>
<th>Standard error (multiplied by ( 10^5 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,3</td>
<td>(-15.0, 480.0)</td>
<td>327.414</td>
</tr>
<tr>
<td>1,5</td>
<td>(-4.2, 133.33)</td>
<td>116.267</td>
</tr>
<tr>
<td>3,5</td>
<td>(-35.0, 600.0)</td>
<td>478.330</td>
</tr>
</tbody>
</table>

Results for the female subpopulations were also obtained. For the juvenile females with 10 clusters the covariance matrix under multinomial sampling (4.4.6) is

\[
\hat{\Sigma}_{m3} = \begin{bmatrix}
-10.08 & 240.8 & -82.32 & -40.32 & -7.84 & -81.2 \\
-21.168 & -82.32 & 415.128 & -84.672 & 246.528 & -170.52 \\
-10.368 & -40.32 & -84.672 & 246.528 & -8.064 & -83.52 \\
-20.88 & -81.2 & -170.52 & -83.52 & -16.24 & 411.8
\end{bmatrix} \times 10^{-5},
\]

resulting in the vector

\[
\hat{W}_3 = (69.408, -10.08, -21.168, -10.368, -2.016, -20.88, 240.8, -82.32, -40.32, -7.84, -81.2, 415.128, -84.672, -16.464, -170.52, 246.528, -8.064, -83.52, 54.432, -16.24, 411.8) \times 10^{-5}.
\]
Similarly, the upper left 6x6 submatrix of $\hat{\Sigma}_{DM3}$ provides

$$\hat{\Sigma}_3 = (96.0, -168.889, -278.222, 272.889, 52.444, -40.0, 942.22,$$
$$351.111, -582.22, -124.44, -13.333, 2533.78, -1339.56,$$
$$-274.667, -593.33, 1393.78, 289.778, -337.778, 64.0,$$
$$-71.111, 926.667) \times 10^{-5}.$$  

A graph of $\hat{\Sigma}_3$ versus $\hat{\Sigma}_3$ is given in Figure 4.5. The standard errors for the second quadrant points of Figure 4.5 are given in Table 4.9a. Six of the 21 points lie in the second quadrant. Three of those points have relatively small variances and they correspond to categories 1, 5, and 7, where the observed frequencies are small.

Table 4.9a. Second quadrant points of Figure 4.5 and their standard errors

<table>
<thead>
<tr>
<th>Cell</th>
<th>Points (multiplied by $10^5$)</th>
<th>Standard error (multiplied by $10^5$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,4</td>
<td>(-10.368, 272.889)</td>
<td>78.377</td>
</tr>
<tr>
<td>1,5</td>
<td>(-2.016, 52.444)</td>
<td>16.022</td>
</tr>
<tr>
<td>2,3</td>
<td>(-82.32, 351.1111)</td>
<td>174.431</td>
</tr>
<tr>
<td>4,5</td>
<td>(-8.064, 303.1111)</td>
<td>125.180</td>
</tr>
</tbody>
</table>

Similarly, for the 21 adult female turkeys the covariance matrix of $\hat{\Sigma}_4$ under multinomial sampling (4.4.6) is
Figure 4.5. A graph of the elements of the estimated covariance matrix under cluster sampling, $\hat{\Sigma}$, versus the elements of the estimated covariance matrix under Multinomial sampling, $\hat{\Psi}$, for juvenile female turkeys.
The upper left 6x6 submatrix of provides

\[
\hat{\Sigma}_{m4} = \begin{pmatrix}
\end{pmatrix} \times 10^{-5},
\]

and

\[
\hat{\mathbf{w}}_4 = (73.288, -11.8277, -18.2132, -7.9093, -3.483, -26.6304,
262.278, -74.2186, -32.2304, -14.1932, -108.519, 363.808,
-49.6308, -21.8558, -167.106, 186.066, -9.4912, -72.5678,
87.249, -31.956, 454.717) \times 10^{-5}.
\]

The upper left 6x6 submatrix of \( \hat{\Sigma}_{DM4} \) provides

\[
\hat{\mathbf{v}}_4 = (203.619, 29.0476, -104.19, -201.238, 133.143, 27.9048,
787.619, -214.476, -88.0952, 48.8571, -441.238, 1061.9,
94.381, -47.4286, -373.048, 818.476, -160.286, -415.81,
704.571, -403.714, 1418.48) \times 10^{-5}.
\]

A graph of \( \hat{\mathbf{v}}_4 \) versus \( \hat{\mathbf{w}}_4 \) is given in Figure 4.6. The points in the second quadrant of Figure 4.6 are given in Table 4.9b. Two of those five points have relatively small variances, and correspond to categories 1 and 5 where the observed frequencies are relatively small. However, the various generalized least squares estimators for \( C_4 \) are not so seriously affected by these points because there is a
Figure 4.6. A graph of the elements of the estimated covariance matrix under cluster sampling, $\hat{\gamma}$, versus the elements of the estimated covariance matrix under Multinomial sampling, $\hat{\nu}$, for adult female turkeys.
moderately large number of adult female turkeys. The various generalized least squares estimators for \( C_4 \) give more nearly consistent values for the adult female turkeys than for the adult male turkeys.

Table 4.9b. Second quadrant points of Figure 4.6 and their standard errors

<table>
<thead>
<tr>
<th>Cell</th>
<th>Points</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(multiplied by ( 10^5 ))</td>
<td>(multiplied by ( 10^5 ))</td>
</tr>
<tr>
<td>1,2</td>
<td>(-11.8277, 29.0476)</td>
<td>81.544</td>
</tr>
<tr>
<td>1,5</td>
<td>(-3.483, 133.143)</td>
<td>78.223</td>
</tr>
<tr>
<td>1,6</td>
<td>(-26.6304, 21.9048)</td>
<td>134.592</td>
</tr>
<tr>
<td>2,5</td>
<td>(-14.1932, 48.8571)</td>
<td>265.735</td>
</tr>
<tr>
<td>3,4</td>
<td>(-49.6308, 94.381)</td>
<td>61.978</td>
</tr>
</tbody>
</table>

Generalized least squares estimates for the \( C_j \)'s can now be computed. The Cholesky decomposition of \( \Lambda_{ml}^{-1} \) (juvenile male turkeys) is given by

\[
\Lambda_{ml}^{-1} = R_1^T R_1, \quad (4.4.29)
\]

where

It follows that

\[ \hat{R}_1 \hat{\Sigma} \hat{R}_1^T = I. \quad (4.4.30) \]

Let \( \hat{V}_{YY1} \) denote the upper left 6x6 matrix of the estimated covariance matrix of the linear combination of \( \hat{\pi}_{t1} \)'s defined by

\[ \hat{V}_{t1} = \hat{R}_1 (\hat{\pi}_{t1} - \hat{\pi}_1). \quad (4.4.31) \]

Then,

\[ \hat{V}_{YY1} = \hat{R}_1 \hat{V}_{PP1} \hat{R}_1^T \quad (4.4.32) \]

\[ = \begin{bmatrix} 0.375 & 0.285546 & 0.162176 & -0.860319 & -0.366202 & 0.153572 \\ 0.285546 & 5.07143 & 1.95735 & 0.443022 & -0.272356 & -1.71243 \\ 0.162176 & 1.95735 & 3.01074 & -3.65078 & -1.06675 & -0.37928 \\ -0.860319 & 0.443022 & -3.65078 & 8.79043 & 2.55703 & 0.822605 \\ -0.366202 & -0.272356 & -1.06675 & 2.55703 & 0.822605 & -0.383454 \\ 0.153572 & -1.71243 & -0.37928 & -1.4696 & -0.383454 & 0.959675 \end{bmatrix} \]
The generalized least squares estimator defined in (4.2.71) is

\[ \hat{C}_{Bl} = (A' \hat{D}_{B}^{-1} A)^{-1} A' \hat{D}_{B}^{-1} \hat{H}_{l} = 3.1717, \quad (4.4.33) \]

where

\[ A' = (1,0,0,0,0,1,0,0,0,1,0,0,1,0,0,1,0,0,0,0,0,0,0,0,0), \]
\[ D_{B} = \text{diag}(2,1,1,1,1,1,2,1,1,1,1,2,1,1,1,2,1,1,2,1,2), \]
\[ H_{l} = \text{vech} \, \hat{V}_{YY_{l}}. \]

The estimator \( \hat{C}_{Bl} \) is identical to that previously computed using expression (4.4.23).

The estimated covariance matrix for the estimated covariances of the transformed \( \hat{Y}_{l} \)'s was computed by applying expression (4.2.66) to the \( \hat{Y}_{l} \)'s. The diagonal elements of the matrix of estimated covariances of the estimated covariances are given in Table 4.10. The cluster standard errors of the estimated covariances for the transformed variables are given in the last column of Table 4.11.

Letting \( \hat{D}_{wl} \) be the diagonal matrix composed of the elements of the matrix in Table 4.10, the estimated generalized least squares estimator defined in (4.2.74) is

\[ \hat{C}_{wl} = (A' \hat{D}_{wl}^{-1} A)^{-1} A' \hat{D}_{wl}^{-1} \hat{H}_{l} = 0.60038, \quad (4.4.34) \]
Table 4.10. Diagonal elements of the estimated covariance matrix of the transformed covariances for juvenile male turkeys

<table>
<thead>
<tr>
<th>Identification</th>
<th>Estimated covariance</th>
<th>Identification</th>
<th>Estimated covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,11</td>
<td>0.0272</td>
<td>33,33</td>
<td>3.8653</td>
</tr>
<tr>
<td>11,22</td>
<td>0.1519</td>
<td>33,44</td>
<td>5.7104</td>
</tr>
<tr>
<td>11,33</td>
<td>0.0830</td>
<td>33,55</td>
<td>0.4732</td>
</tr>
<tr>
<td>11,44</td>
<td>0.8368</td>
<td>33,66</td>
<td>0.2420</td>
</tr>
<tr>
<td>11,55</td>
<td>0.0650</td>
<td>44,44</td>
<td>12.5215</td>
</tr>
<tr>
<td>11,66</td>
<td>0.0687</td>
<td>44,55</td>
<td>1.2766</td>
</tr>
<tr>
<td>22,22</td>
<td>4.2397</td>
<td>44,66</td>
<td>0.9612</td>
</tr>
<tr>
<td>22,33</td>
<td>2.1142</td>
<td>55,55</td>
<td>0.1123</td>
</tr>
<tr>
<td>22,44</td>
<td>4.1270</td>
<td>55,66</td>
<td>0.1234</td>
</tr>
<tr>
<td>22,55</td>
<td>0.3507</td>
<td>66,66</td>
<td>0.0903</td>
</tr>
<tr>
<td>22,66</td>
<td>0.1452</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.11. Transformed estimated covariance matrix for proportions of juvenile male turkeys

<table>
<thead>
<tr>
<th>Identification</th>
<th>Model estimate of ( \hat{\Sigma}_{YY} )</th>
<th>Elements of ( \hat{\Sigma}_{YY} )</th>
<th>Model standard error</th>
<th>Cluster standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>3.172</td>
<td>0.375</td>
<td>1.6686</td>
<td>0.1649</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>0.286</td>
<td>1.1799</td>
<td>0.3898</td>
</tr>
<tr>
<td>31</td>
<td>0</td>
<td>0.162</td>
<td>1.1799</td>
<td>0.2881</td>
</tr>
<tr>
<td>41</td>
<td>0</td>
<td>-0.860</td>
<td>1.1799</td>
<td>0.9148</td>
</tr>
<tr>
<td>51</td>
<td>0</td>
<td>-0.366</td>
<td>1.1799</td>
<td>0.2550</td>
</tr>
<tr>
<td>61</td>
<td>0</td>
<td>0.154</td>
<td>1.1799</td>
<td>0.2621</td>
</tr>
<tr>
<td>22</td>
<td>3.172</td>
<td>5.071</td>
<td>1.6686</td>
<td>2.0591</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>1.957</td>
<td>1.1799</td>
<td>1.4540</td>
</tr>
</tbody>
</table>
where $\hat{A}$ and $\hat{H}$ are defined following (4.4.33). The estimated variance of $\hat{C}_{w1}$ is

$$V(C_{w1}) = (A' D_{w1}^{-1} A)^{-1}$$

$$= .01744$$  (4.4.35)

and the lack-of-fit statistic is

$$x^2_{wl} = H_{w1} A' (A' D_{w1}^{-1} A)^{-1} A' H_{w1}$$

$$= 55.6192 ,$$  (4.4.36)
suggesting that the model is a poor fit. The estimate of $C$ obtained using $\hat{D}_w$ is smaller than that obtained using $D_B$. The value of $C$ obtained using $\hat{D}_w$ is also smaller than one. Recall that the allowable range of values for estimates of $C$ is

$$1 \leq C \leq 50.$$

There are only five juvenile female turkeys (or clusters). This number does not appear to be sufficiently large for $C_w$ to be a reliable estimator.

The variance of $\hat{C}_B$ can be obtained in several ways. One method that uses few assumptions is to compute the variance as

$$V[\hat{C}_B] = (A'D_B^{-1}A)_B^{-1}D_B^{-1}[\hat{v}(\text{vech } \hat{V}_{YY_1})]D_B^{-1}A'(A'D_B^{-1}A)_B^{-1}$$

(4.4.38)

where the elements of $\hat{v}(\text{vech } \hat{V}_{YY_1})$ are given in Table 4.11. For the juvenile female turkeys

$$\hat{V}[\hat{C}_B] = 0.47901.$$ 

The estimated variance of $\hat{C}_B$ is larger than the estimated variance of $\hat{C}_w$. The fourth column of Table 4.11, gives the estimated standard errors for the elements of $\hat{V}_{YY_1}$, if this is used to estimate the variance of the elements of $\hat{V}_{YY_1}$, we have

$$\hat{V}[\hat{C}_B] = 0.4640$$

and the standard error of $\hat{C}_B$ is 0.6812. Based on this standard
error, $C_{B1}$ appears to be different from one. Comparing $\hat{C}_{B1}$ with $\hat{C}_{W1}$ suggests that $\hat{C}_{B1}$ may be more reliable for small samples. The difference in the two estimation procedures is that $D_{W1}$ is estimated in $\hat{C}_{W1}$, and a moderately number of clusters may be needed for $\hat{D}_{W1}$ to be sufficiently accurate. In constructing a lack-of-fit test, the diagonal matrix for the covariance matrix of the estimated covariances is used. Following the procedure for (4.4.20),

$$X^2_{B1} = \left(H^{-1}_{-1} - \hat{C}_{B1}\hat{A}^{-1}_{B}H^{-1}_{-1}\right)(1.3921)^{-1} = 41.8186$$

is approximately distributed as a chi-square random variable with twenty degrees of freedom when the null model is true. On the basis of this statistic the model is rejected at the one percent level.

For the adult male turkeys, similar statistics can be obtained. There are 4 clusters and 7 categories. The last category (Oaksaw) has no observation for each of the 4 clusters. So here we are considering 5x5 covariance matrices instead of 6x6 matrices.

The Cholesky decomposition of $\Lambda^{-1}_{nm2}$ (adult male turkeys) is given by

$$\Lambda^{-1}_{nm2} = \hat{R}^\top \hat{R},$$

where
The 5x5 matrix, $\hat{Y}^2_{Y2}$ following the expression in (4.4.32) is

$$
\hat{R}_2 = \begin{bmatrix}
41.9137 & 2.14942 & 2.14942 & 2.14942 & 2.14942 \\
0 & 58.4705 & 1.46176 & 1.46176 & 1.46176 \\
0 & 0 & 16.8325 & 4.95074 & 4.95074 \\
0 & 0 & 0 & 26.15 & 2.24946 \\
0 & 0 & 0 & 0 & 27.7137 \\
\end{bmatrix}.
$$

The generalized least squares estimator defined in (4.2.71) is

$$
\hat{C}_{B2} = 3.8863
$$

and as for the juvenile male turkeys the value for $\hat{C}_{B2}$ is identical to $\hat{C}_{2B}$ in (4.4.24).

Table 4.12 gives the estimated covariances of the estimated co-variances. The cluster standard errors of the estimated covariances for the transformed variables are given in the last column of Table 4.13.

The value of $\hat{C}_{B2}$ is obtained from (4.2.71)

$$
\hat{C}_{B2} = (A'D^{-1}A)_{B2} - 1_A'D^{-1}A_{B2}^1B
$$

and as for the juvenile male turkeys the value for $\hat{C}_{B2}$ is identical to $\hat{C}_{2B}$ in (4.4.24).
### Table 4.12. Diagonal elements of the estimated covariance matrix of the transformed covariances for adult male turkeys

<table>
<thead>
<tr>
<th>Identification</th>
<th>Estimated covariance</th>
<th>Identification</th>
<th>Estimated covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,11</td>
<td>2.5515</td>
<td>22,55</td>
<td>0.0891</td>
</tr>
<tr>
<td>11,22</td>
<td>0.3340</td>
<td>33,33</td>
<td>0.844</td>
</tr>
<tr>
<td>11,33</td>
<td>5.0843</td>
<td>33,44</td>
<td>5.4165</td>
</tr>
<tr>
<td>11,44</td>
<td>3.4152</td>
<td>33,55</td>
<td>3.0672</td>
</tr>
<tr>
<td>11,55</td>
<td>1.7301</td>
<td>44,44</td>
<td>20.6907</td>
</tr>
<tr>
<td>22,22</td>
<td>0.1372</td>
<td>44,55</td>
<td>2.0974</td>
</tr>
<tr>
<td>22,33</td>
<td>0.6449</td>
<td>55,55</td>
<td>0.8809</td>
</tr>
<tr>
<td>22,44</td>
<td>1.1340</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 4.13. Transformed estimated covariance matrix for proportions of adult male turkeys

<table>
<thead>
<tr>
<th>Identification</th>
<th>Model estimate of $\hat{\Lambda}_{YY2}$</th>
<th>Elements of $\hat{\Lambda}_{YY2}$</th>
<th>Model standard error</th>
<th>Cluster standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>3.8863</td>
<td>2.80434</td>
<td>2.3899</td>
<td>1.5973</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>-0.519049</td>
<td>1.6899</td>
<td>0.5779</td>
</tr>
<tr>
<td>31</td>
<td>0</td>
<td>3.79877</td>
<td>1.6899</td>
<td>2.2548</td>
</tr>
<tr>
<td>41</td>
<td>0</td>
<td>-2.6544</td>
<td>1.6899</td>
<td>1.8480</td>
</tr>
<tr>
<td>51</td>
<td>0</td>
<td>2.01341</td>
<td>1.6899</td>
<td>1.3153</td>
</tr>
<tr>
<td>22</td>
<td>3.8863</td>
<td>0.813675</td>
<td>2.3899</td>
<td>0.3703</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>-1.30764</td>
<td>1.6899</td>
<td>0.8031</td>
</tr>
<tr>
<td>42</td>
<td>0</td>
<td>-1.18967</td>
<td>1.6899</td>
<td>1.0649</td>
</tr>
<tr>
<td>52</td>
<td>0</td>
<td>-0.842626</td>
<td>1.6899</td>
<td>0.2985</td>
</tr>
<tr>
<td>33</td>
<td>3.8863</td>
<td>5.65647</td>
<td>2.3899</td>
<td>3.1376</td>
</tr>
<tr>
<td>43</td>
<td>0</td>
<td>-2.13361</td>
<td>1.6899</td>
<td>2.3273</td>
</tr>
</tbody>
</table>
Table 4.13. (continued)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Model estimate of $\hat{V}_{\text{YY2}}$</th>
<th>Elements of $\hat{V}_{\text{YY2}}$</th>
<th>Model standard error</th>
<th>Cluster standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>53</td>
<td>0</td>
<td>3.14653</td>
<td>1.6899</td>
<td>1.7513</td>
</tr>
<tr>
<td>44</td>
<td>3.8863</td>
<td>3.14653</td>
<td>8.00633</td>
<td>4.5487</td>
</tr>
<tr>
<td>54</td>
<td>0</td>
<td>-0.0187023</td>
<td>1.6899</td>
<td>1.4482</td>
</tr>
<tr>
<td>55</td>
<td>3.8863</td>
<td>2.15054</td>
<td>2.3899</td>
<td>0.9386</td>
</tr>
</tbody>
</table>

where

$$\hat{A}_2 = (1,0,0,0,0,1,0,0,1,0,1,0,1,0,1)$$

$$D_{B2} = \text{diag}(2,1,1,1,2,1,1,1,2,1,1,1,2,1,2)$$

$$H_2 = \text{vech} \hat{V}_{\text{YY2}}.$$  

The diagonal matrix, $\hat{D}_2$, consisting of the elements of Table 4.12, is used in expression (4.2.74). The resulting generalized least squares estimator for adult male turkeys is

$$\hat{C}_2 = 1.1635.$$  

The estimated variance of $\hat{C}_2$ following (4.4.35) is

$$\hat{V}[\hat{C}_2] = 0.1115,$$

and the lack-of-fit statistic from (4.4.36) is
\[ x^2_{w2} = 31.3549. \]

The estimate \( \hat{C}_{B2} (= 4.2013) \) is larger than \( \hat{C}_{w2} (= 1.1635) \). One estimate of the variance of \( \hat{C}_{B2} \) following (4.4.38) is

\[ \hat{V}_{\{\hat{C}_{B2}\}} = 1.4037. \]

Another estimate of the variance of \( \hat{C}_B \) using the elements of \( \hat{V}_{YY2} \) in Table 4.13 is

\[ \hat{V}_{\{\hat{C}_{B2}\}} = 1.1424. \]

The two estimates of the variance of \( \hat{C}_{B2} \) are very similar. The estimates of the variances of \( \hat{C}_{B2} \) both suggest that \( C_{B2} \) is different from one. The associated lack-of-fit statistic (4.4.39) is

\[ x^2_{B2} = 21.3437. \]

When the model is true \( x^2_{B2} \) is approximately distributed as a chi-square random variable with 14 degrees of freedom. On the basis of this statistic the model is accepted at the ten percent level.

Similarly, the cholesky decomposition of the matrices \( \hat{A}^{-1}_{mm3} \) (juvenile female turkeys) and \( \hat{A}^{-1}_{mm4} \) (adult female turkeys) are

\[ \hat{A}^{-1}_{mj} = \hat{P}_{j} \hat{\Sigma}_{j} \hat{P}_{j} \quad (j = 3,4), \]

where
\[
\hat{R}^3 = \begin{bmatrix}
0 & 0 & 19.3646 & 10.5821 & 10.5821 & 10.5821 \\
0 & 0 & 0 & 20.98 & 4.4298 & 4.4298 \\
0 & 0 & 0 & 0 & 43.1164 & 1.7004 \\
0 & 0 & 0 & 0 & 0 & 15.5832
\end{bmatrix}
\]

and

\[
\hat{R}^4 = \begin{bmatrix}
0 & 0 & 0 & 24.1135 & 4.13917 & 4.13917 \\
0 & 0 & 0 & 0 & 34.2991 & 2.41048 \\
0 & 0 & 0 & 0 & 0 & 14.8297
\end{bmatrix}
\]

The 6x6 covariance matrices, \( \hat{V}_{YY3} \) and \( \hat{V}_{YY4} \), for \( \hat{X}_{tj} \) (4.4.31), j = 3,4, respectively, are

\[
\hat{V}_{YY3} = \begin{bmatrix}
0.8547 & 0.1734 & -0.3180 & 0.5910 & 0.1857 & -0.5083 \\
0.1734 & 5.0787 & 0.7040 & -3.7405 & -1.5238 & -0.2557 \\
-0.3180 & 0.7040 & 2.8573 & -2.8237 & -1.1058 & -0.9366 \\
0.5910 & -3.7405 & -2.8237 & 6.2121 & 2.5516 & -0.5137 \\
0.1857 & -1.5238 & -1.1058 & 2.5516 & 1.1123 & -0.2323 \\
-0.5083 & -0.2557 & -0.9366 & -0.5137 & -0.2323 & 2.2503
\end{bmatrix}
\]

and
The diagonal elements of the estimated covariance matrices of \( \hat{c}_{B3} \) and \( \hat{c}_{B4} \) are given in Table 14.4 and Table 14.6, respectively. The cluster standard error of the estimated covariance for the transformed variables \( Y^t_1 \) and \( Y^t_4 \) are given in the last column of Tables 4.15 and 4.17, respectively. The generalized least squares estimators defined in (4.2.74) are

\[
\hat{c}_{W3} = 1.2880
\]

and

\[
\hat{c}_{W4} = 3.4854.
\]
Table 4.14. Diagonal elements of the estimated covariance matrix of the transformed covariances for juvenile female turkeys

<table>
<thead>
<tr>
<th>Identification</th>
<th>Estimated covariance</th>
<th>Identification</th>
<th>Estimated covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,11</td>
<td>0.1542</td>
<td>33,33</td>
<td>3.7636</td>
</tr>
<tr>
<td>11,22</td>
<td>0.4007</td>
<td>33,44</td>
<td>3.4223</td>
</tr>
<tr>
<td>11,33</td>
<td>0.0504</td>
<td>33,55</td>
<td>0.4845</td>
</tr>
<tr>
<td>11,44</td>
<td>0.2478</td>
<td>33,66</td>
<td>0.8538</td>
</tr>
<tr>
<td>11,55</td>
<td>0.0391</td>
<td>44,44</td>
<td>4.9053</td>
</tr>
<tr>
<td>11,66</td>
<td>0.3926</td>
<td>44,55</td>
<td>0.9222</td>
</tr>
<tr>
<td>22,22</td>
<td>2.7411</td>
<td>44,66</td>
<td>1.8690</td>
</tr>
<tr>
<td>22,33</td>
<td>0.9447</td>
<td>55,55</td>
<td>0.1894</td>
</tr>
<tr>
<td>22,44</td>
<td>2.2983</td>
<td>55,66</td>
<td>0.3686</td>
</tr>
<tr>
<td>22,55</td>
<td>0.4832</td>
<td>66,66</td>
<td>1.0298</td>
</tr>
<tr>
<td>22,66</td>
<td>1.1797</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.15. Transformed estimated covariance matrix for proportions of juvenile female turkeys

<table>
<thead>
<tr>
<th>Identification</th>
<th>Model estimate of $\hat{V}_{YY3}$</th>
<th>Elements of $\hat{V}_{Y3}$</th>
<th>Model standard error</th>
<th>Cluster standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>3.0609</td>
<td>0.8547</td>
<td>1.4074</td>
<td>0.39266</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>0.1734</td>
<td>0.9952</td>
<td>0.63300</td>
</tr>
<tr>
<td>31</td>
<td>0</td>
<td>-0.3180</td>
<td>0.9952</td>
<td>0.22440</td>
</tr>
<tr>
<td>41</td>
<td>0</td>
<td>0.5910</td>
<td>0.9952</td>
<td>0.49777</td>
</tr>
<tr>
<td>51</td>
<td>0</td>
<td>0.1857</td>
<td>0.9952</td>
<td>0.19768</td>
</tr>
<tr>
<td>61</td>
<td>0</td>
<td>-0.5083</td>
<td>0.9952</td>
<td>0.62661</td>
</tr>
<tr>
<td>22</td>
<td>3.0609</td>
<td>5.0787</td>
<td>1.4074</td>
<td>1.65562</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>0.7040</td>
<td>0.9952</td>
<td>0.97196</td>
</tr>
</tbody>
</table>
Table 4.15. (continued)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Model estimate of $\hat{V}_{YY3}$</th>
<th>Elements of $\hat{V}_{YY3}$</th>
<th>Model standard error</th>
<th>Cluster standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>0</td>
<td>-3.7405</td>
<td>.9952</td>
<td>1.51602</td>
</tr>
<tr>
<td>52</td>
<td>0</td>
<td>-1.5238</td>
<td>.9952</td>
<td>0.69511</td>
</tr>
<tr>
<td>62</td>
<td>0</td>
<td>-0.2557</td>
<td>.9952</td>
<td>1.08615</td>
</tr>
<tr>
<td>33</td>
<td>3.0609</td>
<td>2.8573</td>
<td>1.4074</td>
<td>1.94000</td>
</tr>
<tr>
<td>43</td>
<td>0</td>
<td>-2.8237</td>
<td>.9952</td>
<td>1.84995</td>
</tr>
<tr>
<td>53</td>
<td>0</td>
<td>-1.1058</td>
<td>.9952</td>
<td>0.69604</td>
</tr>
<tr>
<td>63</td>
<td>0</td>
<td>-0.9366</td>
<td>.9952</td>
<td>0.92399</td>
</tr>
<tr>
<td>44</td>
<td>3.0609</td>
<td>6.2121</td>
<td>1.4074</td>
<td>2.21480</td>
</tr>
<tr>
<td>54</td>
<td>0</td>
<td>2.5516</td>
<td>.9952</td>
<td>0.96033</td>
</tr>
<tr>
<td>64</td>
<td>0</td>
<td>-0.5137</td>
<td>.9952</td>
<td>1.36713</td>
</tr>
<tr>
<td>55</td>
<td>3.0609</td>
<td>1.1123</td>
<td>1.4074</td>
<td>0.43521</td>
</tr>
<tr>
<td>65</td>
<td>0</td>
<td>-0.2323</td>
<td>.9952</td>
<td>0.60709</td>
</tr>
<tr>
<td>66</td>
<td>3.0609</td>
<td>2.2503</td>
<td>1.4074</td>
<td>1.01477</td>
</tr>
</tbody>
</table>

Table 4.16. Diagonal elements of the estimated covariance matrix of the transformed covariance for adult female turkeys

<table>
<thead>
<tr>
<th>Identification</th>
<th>Estimated covariance</th>
<th>Identification</th>
<th>Estimated covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,11</td>
<td>2.0372</td>
<td>33,33</td>
<td>1.3082</td>
</tr>
<tr>
<td>11,22</td>
<td>0.9457</td>
<td>33,44</td>
<td>0.7232</td>
</tr>
<tr>
<td>11,33</td>
<td>0.6351</td>
<td>33,55</td>
<td>2.5646</td>
</tr>
<tr>
<td>11,44</td>
<td>0.7125</td>
<td>33,66</td>
<td>1.0039</td>
</tr>
<tr>
<td>11,55</td>
<td>1.2744</td>
<td>44,44</td>
<td>0.8538</td>
</tr>
<tr>
<td>11,66</td>
<td>0.6156</td>
<td>44,55</td>
<td>1.0000</td>
</tr>
<tr>
<td>22,22</td>
<td>0.9314</td>
<td>44,66</td>
<td>0.6482</td>
</tr>
</tbody>
</table>
Table 4.16. (continued)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Estimated covariance</th>
<th>Identification</th>
<th>Estimated covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>22,33</td>
<td>0.4201</td>
<td>55,55</td>
<td>13.2999</td>
</tr>
<tr>
<td>22,44</td>
<td>0.7466</td>
<td>55,66</td>
<td>1.3914</td>
</tr>
<tr>
<td>22,55</td>
<td>0.6368</td>
<td>66,66</td>
<td>0.5980</td>
</tr>
<tr>
<td>22,66</td>
<td>0.4284</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.17. Transformed estimated covariance matrix for proportions of adult female turkeys

<table>
<thead>
<tr>
<th>Identification</th>
<th>Model estimate of $\hat{V}_{YY4}$</th>
<th>Elements of $\hat{V}_{YY4}$</th>
<th>Model standard error</th>
<th>Cluster standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>4.2013</td>
<td>4.01</td>
<td>1.5581</td>
<td>1.4273</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>1.03818</td>
<td>1.1018</td>
<td>0.9725</td>
</tr>
<tr>
<td>31</td>
<td>0</td>
<td>0.52977</td>
<td>1.1018</td>
<td>0.7970</td>
</tr>
<tr>
<td>41</td>
<td>0</td>
<td>-1.10635</td>
<td>1.1018</td>
<td>0.8441</td>
</tr>
<tr>
<td>51</td>
<td>0</td>
<td>2.20365</td>
<td>1.1018</td>
<td>1.1289</td>
</tr>
<tr>
<td>61</td>
<td>0</td>
<td>-0.56765</td>
<td>1.1018</td>
<td>0.7846</td>
</tr>
<tr>
<td>22</td>
<td>4.2013</td>
<td>3.0244</td>
<td>1.5581</td>
<td>0.9651</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>0.68232</td>
<td>1.1018</td>
<td>0.6482</td>
</tr>
<tr>
<td>42</td>
<td>0</td>
<td>0.38159</td>
<td>1.1018</td>
<td>0.8641</td>
</tr>
<tr>
<td>52</td>
<td>0</td>
<td>0.18354</td>
<td>1.1018</td>
<td>0.7980</td>
</tr>
<tr>
<td>62</td>
<td>0</td>
<td>-1.5289</td>
<td>1.1018</td>
<td>0.6545</td>
</tr>
<tr>
<td>33</td>
<td>4.2013</td>
<td>3.71561</td>
<td>1.5581</td>
<td>1.1438</td>
</tr>
<tr>
<td>43</td>
<td>0</td>
<td>0.87809</td>
<td>1.1018</td>
<td>0.8504</td>
</tr>
<tr>
<td>53</td>
<td>0</td>
<td>-0.25763</td>
<td>1.1018</td>
<td>1.6014</td>
</tr>
<tr>
<td>63</td>
<td>0</td>
<td>-0.40030</td>
<td>1.1018</td>
<td>1.0020</td>
</tr>
<tr>
<td>44</td>
<td>4.2013</td>
<td>3.80026</td>
<td>1.5581</td>
<td>0.9240</td>
</tr>
</tbody>
</table>
Table 4.17. (continued)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Model estimate of $\hat{V}_{yy4}$</th>
<th>Elements of $\hat{V}_{yy4}$</th>
<th>Model standard error</th>
<th>Cluster standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>54</td>
<td>0</td>
<td>-1.19094</td>
<td>1.1018</td>
<td>1.0000</td>
</tr>
<tr>
<td>64</td>
<td>0</td>
<td>-0.92541</td>
<td>1.1018</td>
<td>0.8051</td>
</tr>
<tr>
<td>55</td>
<td>4.2013</td>
<td>7.53828</td>
<td>1.5581</td>
<td>3.6469</td>
</tr>
<tr>
<td>65</td>
<td>0</td>
<td>-2.05505</td>
<td>1.1018</td>
<td>1.1796</td>
</tr>
<tr>
<td>66</td>
<td>4.2013</td>
<td>3.11949</td>
<td>1.5581</td>
<td>0.7733</td>
</tr>
</tbody>
</table>

Note the great contrast between the estimators $\hat{C}_{b3}$ and $\hat{C}_{w3}$ for the juvenile female turkeys. However, for adult female turkeys with 21 clusters, the greatest number for the four subpopulations, $\hat{C}_{b4}$ and $\hat{C}_{w4}$ are quite similar.

The estimated variance of $\hat{C}_{w3}$, according to (4.4.35), is

$$\hat{V}[\hat{C}_{w3}] = 0.0737,$$

and the estimated variance of $\hat{C}_{w4}$ is

$$\hat{V}[\hat{C}_{w4}] = 0.19056.$$

The associated lack-of-fit statistics (4.4.36) are

$$x^2_{w3} = 42.8555 \quad \text{(juvenile female turkeys)}$$

and
\[ x^2_{w4} = 23.4507 \text{ (adult female turkeys)}. \]

One estimate of the variance of \( \hat{C}_{B3} \) is

\[ \hat{V}[\hat{C}_{B3}] = 0.60656, \]

by use of (4.4.38). Similarly, the variance of \( \hat{C}_{B4} \) is

\[ \hat{V}[\hat{C}_{B4}] = 0.65876. \]

Suppose the estimate of the variance of \( \hat{C}_{B3} \) and \( \hat{C}_{B4} \) are computed based on standard errors obtained from \( \hat{V}_{YY3} \) and \( \hat{V}_{YY4} \) which are listed in the fourth column of Tables 4.15 and 4.17, respectively. Then,

\[ \hat{V}[\hat{C}_{B3}] = 0.33013 \]

and

\[ \hat{V}[\hat{C}_{B4}] = 0.40463. \]

The estimated standard errors for \( \hat{C}_{B3} \) and \( \hat{C}_{B4} \), based on the latter variance estimates, are

\[ \text{Std error } \{ \hat{C}_{B3} \} = 0.5746 \]

and

\[ \text{Std error } \{ \hat{C}_{B4} \} = 0.6361. \]
On the basis of these standard errors, one would conclude that neither \( C_{B3} \) nor \( C_{B4} \) is one. The lack-of-fit statistics according to (4.4.39) are

\[
X^2_{B3} = 46.682
\]

and

\[
X^2_{B4} = 20.843.
\]

When the model is true, \( X^2_{Bj} \) (\( j = 3, 4 \)) is approximately distributed as a chi-square random variable with 20 degrees of freedom. On the basis of \( X^2_{B3} \) and \( X^2_{B4} \) the model is rejected for juvenile female turkeys and not rejected for adult female turkeys.

The hypothesis of interest (4.4.20) is

\[
H_0: \pi_j = \pi_0 \quad \text{(unknown)} \quad j = 1, 2, 3, 4.
\]

The statistic used to test \( H_0 \) is given in (4.3.45) as

\[
X^2_{DMI} = \sum_{j=1}^{4} N_j \hat{c}_j \sum_{i=1}^{7} \left( \hat{n}_{ij} - \hat{\pi}_{io} \right)^2 \left( \hat{\pi}_{io} \right).
\]

For each combination of \( C_j \) (\( j = 1, 2, 3, 4 \)) estimators the test statistic can be calculated. However, the lack-of-fit test statistic (4.4.39) suggested that the model may not be reasonable for juvenile subpopulations. In spite of this \( X^2_{DMI} \) was computed and a table of values is presented in Table 4.18. In these calculations \( \hat{c}_{w1} \) was replaced with
a value of one.

Table 4.18. A summary of \( C_j \) estimators and the test statistic value for testing the hypothesis \( H_0: \pi_j = \pi_o \) for juvenile male, adult male, juvenile female, and adult female turkeys

<table>
<thead>
<tr>
<th>C estimators</th>
<th>Juvenile male turkeys</th>
<th>Adult male turkeys</th>
<th>Juvenile female turkeys</th>
<th>Adult female turkeys</th>
<th>( X^2_{DMI} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{C}_B )</td>
<td>3.172</td>
<td>3.886</td>
<td>3.061</td>
<td>4.201</td>
<td>26.935</td>
</tr>
<tr>
<td>( \hat{C}_W )</td>
<td>(0.6004)</td>
<td>1.164</td>
<td>1.288</td>
<td>4.854</td>
<td>79.804</td>
</tr>
<tr>
<td>( \hat{C}_p )</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>99.238</td>
</tr>
</tbody>
</table>

Note in Table 4.18 the value of \( \hat{C}_p \) is one for each of the four subpopulations. The value of \( X^2_{DMI} \) using \( \hat{C}_p \) estimates is equivalent to the Pearson test statistic for independence. For the \( \hat{C}_W \) and \( \hat{C}_B \) estimators, there are rather large differences except for the adult female turkeys. The adult female subpopulation consists of a great number of clusters. The indication here is that the two estimators \( \hat{C}_B \) and \( \hat{C}_W \) are about equal when the number of clusters are large relative to the number of categories. As was the case in Example 4.1, the value of \( X^2_{DMI} \) is smaller for the \( \hat{C}_B \) and \( \hat{C}_W \) estimators than with the \( \hat{C}_p \) estimators. This numerical comparison for \( X^2_{DMI} \) suggests that in the presence of clustering the usual Pearson statistic would tend to reject too often.
Example 4.3.

Brier (1980) considered data pertaining to the manner in which people in 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 (personal satisfaction) and the housing in the community as a whole (community satisfaction). There were 97 families questioned in the metropolitan area and 96 questioned in the outlying area.

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. The groups of five homes are the clusters. There are some clusters with fewer than five homes responding. There are a total of 40 clusters, 20 in the metropolitan Minneapolis-St. Paul area and 20 in the outlying region. The data given in Brier (1980) are reproduced in Appendix A, Tables 7.7 to 7.10.

In this analysis, the interest is in the distribution of the responses for the two areas for the personal satisfaction categories and the community satisfaction categories. The hypotheses are

\[ H_{op} : \pi_j = \pi_0 \text{ (unknown)} \quad j = 1, 2; \]  
\[ (4.4.40) \]

for personal satisfaction categories and

\[ H_{oc} : \pi_j = \pi_0 \text{ (unknown)} \quad j = 1, 2; \]  
\[ (4.4.41) \]
for community satisfaction categories.

The two subpopulations correspond to the nonmetropolitan (nonmetro) area and the metropolitan (metro) area, so \( J = 2 \). Let subpopulation 1 correspond to the nonmetro area and subpopulation 2 correspond to the metro area. Then, for personal satisfaction

\[
\hat{\pi}_1 = (.52083, .40625, .07292)'
\]

and

\[
\hat{\pi}_2 = (.32990, .51546, .15464)'
\]

For community satisfaction, the estimated vectors are

\[
\hat{\pi}_1 = (.25000, .61458, .13542)'
\]

and

\[
\hat{\pi}_2 = (.25773, .60825, .13402)'.
\]

In applying the Dirichlet-Multinomial model, attention is drawn to the fact that there is not the same number of homes in each cluster. To overcome this problem those clusters with less than five homes are allowed to keep the relative proportion of elements in each category but the total is now five. This minor adjustment results in some categories having noninteger counts, but that does not seriously affect the results. The test statistic is
\[ X_{DMI}^2 = \sum_{j=1}^{N_j} C_j^{-1} \sum_{i=1}^{\pi_j \pi_0} (\pi_j \pi_0)^{\pi_j \pi_0} \]  

(4.4.42)

where \( N_j \) is the total sample for the \( j \)th subpopulation, \( \pi_j = (\pi_{1j}, \pi_{2j}, \ldots, \pi_{ij})' \) is the observed vector of proportions for the \( j \)th subpopulation, \( C_j \) is a consistent estimator for the factor \( C_j \) in the covariance matrix for the Dirichlet-Multinomial distribution,

\[ \pi_0 = \sum_{j=1}^{N_j} C_j^{-1} \pi_j \sum_{i=1}^\infty \sum_{j=1}^{\pi_j \pi_0} (\pi_j \pi_0)^{\pi_j \pi_0} \]  

(4.4.43)

and \( I \) is the number of categories (\( I = 3 \)).

The data considered here differ greatly from that in Examples 4.1 and 4.2 in that there is a moderately large number of clusters in relation to the number of categories. A second difference is that the cluster sizes are relatively small, only five homes are sampled in each neighborhood.

As was done in Examples 4.1 and 4.2, several estimators for the clustering factor \( C_j \) are obtained based on the methods suggested in Section 4.2. First consideration is given to the estimator \( C_{jB} \) proposed by Brier (1980).

This estimator,

\[ C_{jB} = (I-1)^{-1}(s_j^{-1})^{-1} \sum_{k=1}^{i} \sum_{l=1}^{s_j} n_{ij} (\pi_{ijk} - \pi_{ij})^2 \pi_{ij} \]  

(4.4.44)
where

\[ \hat{\pi}_{jk} = (\hat{\pi}_{1jk}, \hat{\pi}_{2jk}, \ldots, \hat{\pi}_{Ijk})', \]

is the vector of proportions for the \( k \)th cluster of the \( j \)th sub-population. \( I \) is the number of categories, \( s_j \) is the number of clusters, and \( n_j \) is the size of the clusters in the \( j \)th sub-population. The values for \( \hat{C}_{jB}, j = 1,2 \), for personal satisfaction are

\[ \hat{C}_{1B} = 2.0768 \text{ (nonmetro area) and } \hat{C}_{2B} = 1.6212 \text{ (metro area)}. \]

The estimated vector \( \hat{\pi}_o \) for \( \pi_o \) is

\[ \hat{\pi}_o = (.4136, .4676, .1188)', \]

and the test statistic value based on expression (4.3.45) is

\[ \chi^2_{DMI} = 4.5932. \]

Consider the multinomial model for which \( \hat{C}_j \) is one for \( j = 1,2 \), then the test statistic,

\[ \chi^2_{DMI} = \sum_{j=1}^{2} \sum_{i=1}^{3} (\hat{\pi}_{ij} - \pi_{io})^2 \pi_{io} \]

\[ = \chi^2_{I}, \tag{4.4.45} \]

is the Pearson test statistic for the test of independence. When
$\hat{C}_j$ is one, $j = 1, 2,$

$$\hat{\Pi}_0 = (\cdot 4254, \cdot 4609, \cdot 1137)',$$

and

$$\chi^2 = 8.5137.$$

The observed value of $\chi^2_I$ is larger than the observed value of $\chi^2_{DMI}$. The observed value of $\chi^2_{DMI}$ is less than the 95th percentile of a chi-square distribution with two degrees of freedom, but the observed value of the usual Pearson statistic $\chi^2_I$ is greater than the 95th percentile of a chi-square distribution with two degrees of freedom. The value of $\chi^2_{DMI}$ indicates that there are no significant difference regarding satisfaction levels (personal) for the two areas, whereas the value of $\chi^2_I$ indicates that a significant difference may exist between the two areas, regarding personal satisfaction. The true significance level for $\chi^2_I$ may be much smaller than the nominal level. Before deciding if $\chi^2_{DMI}$ provides a more reliable test, it is wise to assess the fit of the Dirichlet-Multinomial model.

As mentioned in Examples 4.1 and 4.2, the method used in the computation of $\hat{C}_{jB}(j = 1, 2)$ gives no indication of whether the model fits. The alternative methods of constructing estimators for $C_j$ have the advantage that they provide a test for the fit of the model. The alternative methods are based on the generalized least squares technique. The estimator of the covariance matrix for a simple
A random sample of five observations from a multinomial distribution with probability vector \( \hat{\mathbf{\pi}}_1 \) for the nonmetro area is

\[
\hat{\mathbf{\Sigma}}_{ml} = \begin{bmatrix}
500 & -39.0 \\
-39.0 & 475.8
\end{bmatrix} \times 10^{-4},
\]

where the diagonal elements are \( \pi_{il}(1-\pi_{il})(5)^{-1} \) and the off diagonal elements are \( -\pi_{il} \pi_{lj}(5)^{-1} \). The 2x2 matrix is nonsingular. The matrix \( \hat{\mathbf{\Sigma}}_{ml} \) is made into a vector \( \hat{\mathbf{w}}_1 \) by taking the elements by rows and ignoring the elements below the diagonal. The vector \( \hat{\mathbf{w}}_1 \) obtained from \( \hat{\mathbf{\Sigma}}_{ml} \) forms the right side of the regression equation

\[
\hat{\mathbf{y}}_1 = \mathbf{C}_1 \hat{\mathbf{w}}_1 + \xi,
\]

where

\[
\hat{\mathbf{w}}_1 = (500, -39.0, 475.8) \times 10^{-4}.
\]

The estimated covariance matrix constructed using the cluster sampling formula is

\[
\hat{\mathbf{\Sigma}}_{DM1} = 19^{-1} \sum_{k=1}^{20} (\hat{\mathbf{r}}_{1k} - \hat{\mathbf{r}}_{11})(\hat{\mathbf{r}}_{1k} - \hat{\mathbf{r}}_{11})'.
\]

The upper 2x2 submatrix of \( \hat{\mathbf{\Sigma}}_{DM1} \) is

\[
\begin{bmatrix}
989.474 & -726.316 \\
-726.316 & 904.211
\end{bmatrix} \times 10^{-4}.
\]
A graph of the elements of $\hat{V}_{ppl}$, denoted by $\hat{V}_1$, versus the corresponding elements of $\hat{V}_{m1}$, denoted by $\hat{V}_1$, is given in Figure 4.7. If the Dirichlet-Multinomial model is satisfied, then a factor $C_1$ exists such that

$$\hat{\Sigma}_{Dm1} \sim C_1 \hat{\Sigma}_{m1}.$$  
(4.4.48)

where $1 \leq C_1 \leq 5$. An examination of $\hat{V}_1$ and $\hat{V}_1$ shows that there are no points lying in the second quadrant. When the estimates are accurate there should be no points in the second quadrant because all covariances are negative and variances are positive.

The covariance matrix of the estimated covariances defined in (4.2.66) is given in Table 4.19.

<table>
<thead>
<tr>
<th>Identification</th>
<th>Cluster covariance estimate $\times 10^4$</th>
<th>Identification</th>
<th>Cluster covariance estimate $\times 10^4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,11</td>
<td>3.5776</td>
<td>12,12</td>
<td>4.0067</td>
</tr>
<tr>
<td>11,12</td>
<td>-3.2394</td>
<td>12,22</td>
<td>-42.5244</td>
</tr>
<tr>
<td>11,22</td>
<td>2.8452</td>
<td>22,22</td>
<td>5.12238</td>
</tr>
</tbody>
</table>

For the metro area, similar statistics can be obtained based on the same procedure as used for the nonmetro area. The covariance matrix, $\hat{\Sigma}_{m2}$, for the metro area is
Figure 4.7. A graph of the elements of the estimated covariance matrix under cluster sampling, $\hat{\Sigma}_c$, versus the elements of the estimated covariance matrix under Multinomial sampling, $\hat{\Sigma}_m$, for the nonmetro area (personal satisfaction)
\[
\hat{\Sigma}_{m2} = \begin{pmatrix}
435.20 & -320.00 \\
-320.00 & 500.00
\end{pmatrix} \times 10^{-4}.
\]

The vector \( \hat{\mathbf{w}}_2 \) obtained from the covariance matrix \( \hat{\Sigma}_{m2} \) is

\[
\hat{\mathbf{w}}_2 = (435.20, -320.00, 500.00)' \times 10^{-4}.
\]

The estimated vector \( \hat{\mathbf{v}}_2 \), obtained from \( \hat{\Sigma}_{DM2} \), is

\[
\hat{\mathbf{v}}_2 = (985.26, -715.79, 778.95) \times 10^{-4}.
\]

A graph of \( \hat{\mathbf{v}}_2 \) versus \( \hat{\mathbf{w}}_2 \) is given in Figure 4.8. As in the case of the nonmetro area, there are no second quadrant points. The covariance matrix of the estimated covariances defined in (4.2.66) is given in Table 4.20.

Table 4.20. Cluster estimated covariance matrix of sample covariances for the metro area (personal satisfaction)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Cluster covariance estimate ( \times 10^4 )</th>
<th>Identification</th>
<th>Cluster covariance estimate ( \times 10^4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,11</td>
<td>4.84046</td>
<td>12,12</td>
<td>6.08985</td>
</tr>
<tr>
<td>11,12</td>
<td>-5.02414</td>
<td>12,22</td>
<td>-5.56445</td>
</tr>
<tr>
<td>11,22</td>
<td>4.72709</td>
<td>22,22</td>
<td>6.90861</td>
</tr>
</tbody>
</table>

The Cholesky decomposition of \( \hat{\Sigma}_{m1}^{-1} \) is given by
Figure 4.8. A graph of the elements of the estimated covariance matrix under cluster sampling, $\hat{\Sigma}_V$, versus the elements of the estimated covariance matrix under Multinomial sampling, $\hat{\Sigma}_W$, for the metro area (personal satisfaction).
\[ \hat{\Sigma}_{m1}^{-1} = \hat{R}_1 \hat{R}_1', \]  \hspace{1cm} (4.4.49) 

where

\[ \hat{R}_1 = \begin{pmatrix} 7.44678 & 6.10392 \\ 0 & 4.58446 \end{pmatrix}. \]

It follows that

\[ \hat{R}_1 \hat{\Sigma}_{m1}^{-1} \hat{R}_1' = I. \]  \hspace{1cm} (4.4.50)

Let \( \hat{\mathbf{V}}_{Y1} \) denote the upper left 2x2 matrix of the estimated covariance matrix of the linear combination of \( \hat{\mu}' \)'s defined by

\[ \hat{\mathbf{Y}}_{t1} = \hat{R}_1 (\hat{\mathbf{Y}}_{t1} - \hat{\mathbf{Y}}_l). \]  \hspace{1cm} (4.4.51)

Then,

\[ \hat{\mathbf{V}}_{Y1} = \hat{R}_1 \hat{\mathbf{V}}_\text{pp} \hat{R}_1' \]  \hspace{1cm} (under the cluster model)  

\[ = \begin{pmatrix} 2.2531 & 0.05066 \\ 0.05066 & 1.9004 \end{pmatrix}. \]  \hspace{1cm} (4.4.52)

The generalized least squares estimator defined in (4.2.71) is

\[ \hat{C}_{BL} = (A'D_B^{-1}A)^{-1} A'D_B^{-1} B_l \]

\[ = 2.0768 \]  \hspace{1cm} (4.4.53)

where
\( A' = (1,0,1), \)
\( D_B = \text{diag}(2,1,2) \)

(4.4.54)

\( H_1 = \text{vech} \hat{V}_{YY1}. \)

As discussed in Section 4.2 and in Examples 4.1 and 4.2, the value of \( \hat{C}_{BL} \) is identical to that previously computed in (4.4.44).

The estimated covariance matrix for the estimated covariances of the transformed \( \hat{\gamma}_{itl} \)'s was computed by applying formula (4.2.68) to the \( \hat{\gamma}_{itl} \)'s. For example, the cluster estimated variance of the (1,2) element of \( \hat{V}_{YY1} \) is

\[
\hat{U}_{YY,l2}^2 = 19^{-2} \sum_{t=1}^{20} [\hat{\gamma}_{1t} \hat{\gamma}_{2t+1} - \hat{\gamma}_{YYl1}]^2.
\]

(4.4.55)

The matrix of estimated covariances is given in Table 4.21. The cluster standard errors of the estimated covariances for the transformed variables are given in the last column of Table 4.22.

Table 4.21. Cluster estimated covariance matrix of the transformed covariances for nonmetro area (personal satisfaction)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Estimated covariance</th>
<th>Identification</th>
<th>Estimated covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,11</td>
<td>1.1415</td>
<td>12,12</td>
<td>0.2428</td>
</tr>
<tr>
<td>11,12</td>
<td>0.4587</td>
<td>12,22</td>
<td>0.0208</td>
</tr>
<tr>
<td>11,22</td>
<td>0.0288</td>
<td>22,22</td>
<td>0.2945</td>
</tr>
</tbody>
</table>
Table 4.22. Transformed estimated covariance matrix for proportions for the nonmetro area (personal satisfaction)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Identification Model estimate</th>
<th>Elements of $\hat{V}_{YY1}$</th>
<th>Model standard error</th>
<th>Cluster standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,1</td>
<td>2.0768</td>
<td>2.2531</td>
<td>0.81944</td>
<td>1.0685</td>
</tr>
<tr>
<td>1,2</td>
<td>0</td>
<td>0.0507</td>
<td>0.5794</td>
<td>0.4922</td>
</tr>
<tr>
<td>2,2</td>
<td>2.0768</td>
<td>1.9004</td>
<td>0.81944</td>
<td>0.5427</td>
</tr>
</tbody>
</table>

Letting $\hat{D}_{wl}$ be the diagonal matrix composed of the diagonal elements of the matrix in Table 4.21, the estimated generalized least squares estimator defined in (4.2.74) is

$$\hat{C}_{wl} = (\hat{A}'\hat{D}_{wl}^{-1}\hat{A})^{-1}\hat{A}'\hat{D}_{wl}^{-1}\hat{H}_1$$

$$= 1.9727,$$  \hspace{1cm} (4.4.56)

where $\hat{A}$ and $\hat{H}_1$ are defined following (4.4.54). The estimated variance of $\hat{C}_{wl}$ is

$$\hat{V}[\hat{C}_{wl}] = (\hat{A}'\hat{D}_{wl}^{-1}\hat{A})^{-1}$$

$$= 0.2341$$  \hspace{1cm} (4.4.57)

and the lack-of-fit statistic is

$$X^2_{wl} = H_1'[\hat{D}_{wl}^{-1} - \hat{A}(\hat{A}'\hat{D}_{wl}^{-1}\hat{A})^{-1}\hat{A}']H_1^{-1}$$

$$= 0.0972.$$  \hspace{1cm} (4.4.58)
The estimated variance of $\hat{C}_{Bl}$ can be obtained in several ways. The method that uses the fewest assumptions is to compute the variance as

$$\hat{V}[\hat{C}_{Bl}] = (A' D_B^{-1} A)^{-1} A' D_B^{-1} [\hat{V} \{\text{vech} \hat{V}_{YY_1}\}] D_B^{-1} A (A' D_B^{-1} A)^{-1},$$

where the elements of $\hat{V} \{\text{vech} \hat{V}_{YY_1}\}$ are given in Table 4.21. For the nonmetro area, the estimated variance is

$$\hat{V}[\hat{C}_{Bl}] = 0.3734.$$

In the computation of this variance, the unknown matrix $\Sigma_{ml}$ is treated as known. For the nonmetro area, the ratio of the diagonal elements of $\hat{V} \{\text{vech} \hat{V}_{YY_1}\}$ to the diagonal elements of $D_B$ is 0.3357. This provides an alternative variance estimate

$$\hat{V}[\hat{C}_{Bl}] = 0.3357,$$

and the standard error of $\hat{C}_{Bl}$ is 0.5794. Note that the two methods of estimating the variance of $\hat{C}_B$ gave similar values. On the basis of the standard error, it seems that $C_{Bl}$ is not one.

The 3x3 matrix $\hat{V} \{\text{vech} \hat{V}_{YY_1}\}$ is nonsingular. Therefore, we can construct a lack-of-fit test for the model using the diagonal matrix for the covariance matrix of the estimated covariances. If $\Sigma_{ml}$ is known, and if $\hat{V} \{\text{vech} V_{YY_1}\}$ is of the specified diagonal form, then from (4.2.73) the quantity
\[ X^2_{B1} = (H_1'B_1^{-1}H_1 - C_{B1}A_1'D_B^{-1}A_1)^{-1} \]

\[ = 0.10027 \]  \hspace{1cm} (4.4.60)

is approximately distributed as a chi-square random variable with two degrees of freedom when the model is true. On the basis of this statistic, the model is not rejected.

Similar statistics and tests can be obtained for the personal satisfaction data obtained from the metro area using the same procedure as for the nonmetro area. The Cholesky decomposition of \( \Sigma^{-1}_{m2} \) is given by

\[ \Sigma^{-1}_{m2} = R^T R, \]

where

\[ R = \begin{bmatrix} 6.5881 & 4.2164 \\ 0 & 4.4721 \end{bmatrix}. \]

The covariance matrix

\[ \hat{V}_{YY2} = \begin{bmatrix} 1.6845 & -0.6401 \\ -0.6401 & 1.5579 \end{bmatrix}. \]

The generalized least squares estimator defined in (4.2.71) is

\[ \hat{C}_{B2} = 1.6212. \]

The value of \( \hat{C}_{B2} \) is identical to that previously computed in (4.4.44).
The matrix of estimated covariances of the estimated covariances is given in Table 4.23. The cluster standard errors of the estimated covariances for the transformed variables are given in the last column of Table 4.24.

The matrix \( \hat{D}_{w2} \) consisting of the elements of Table 4.23 is used in expression (4.2.74). The resulting generalized least squares estimator for the metro area is

\[ \hat{C}_{w2} = 1.6319. \]

The estimated variance of \( \hat{C}_{w2} \), following (4.4.57), is

\[ \hat{V}\{\hat{C}_{w2}\} = 0.0801, \]

and the lack-of-fit statistic from (4.4.58) is

\[ X^2_{w2} = 6.3518. \]

One estimate of the variance of \( \hat{C}_{b2} \), following (4.4.59), is

\[ \hat{V}\{\hat{C}_{b2}\} = .0596. \]

Another estimate of the variance of \( \hat{C}_{b2} \), using the elements of \( \hat{V}_{yy2} \) in Table 4.23, is

\[ \hat{V}\{\hat{C}_{b2}\} = .07900. \]

The associated lack-of-fit statistic
Table 4.23. Cluster estimated covariance matrix of the transformed covariances for the metro area (personal satisfaction)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Estimated covariance</th>
<th>Identification</th>
<th>Estimated covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,11</td>
<td>.13712</td>
<td>12,12</td>
<td>.06500</td>
</tr>
<tr>
<td>11,12</td>
<td>-.00839</td>
<td>12,22</td>
<td>-.05215</td>
</tr>
<tr>
<td>11,22</td>
<td>-.04572</td>
<td>22,22</td>
<td>.19289</td>
</tr>
</tbody>
</table>

Table 4.24. Transformed estimated covariance matrix for proportions for the metro area (personal satisfaction)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Model estimate of ( \hat{\Omega}_{Y2} )</th>
<th>Elements of ( \hat{\Omega}_{Y2} )</th>
<th>Model standard error</th>
<th>Cluster standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>1.6212</td>
<td>0.04352</td>
<td>.3975</td>
<td>.3703</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>-0.03200</td>
<td>.2811</td>
<td>.2550</td>
</tr>
<tr>
<td>22</td>
<td>1.6212</td>
<td>0.05000</td>
<td>.3975</td>
<td>.4392</td>
</tr>
</tbody>
</table>

\[
X_{B2}^2 = (H_2^\dagger D_{B2}^{-1} \hat{\Sigma} D_{B2}^{-1} C_{B2} A \hat{\Sigma} D_{B2}^{-1}) (0.07900)^{-1}
\]

\[
= 5.2372
\]

is approximately distributed as a chi-square random variable with two degrees of freedom when the null model is true. Since neither \( X_{w2}^2 \) nor \( X_{B2}^2 \) is extremely large, the Dirichlet-Multinomial model appears to be reasonable for the metro area data on personal satisfaction.
Since the covariance matrices $\hat{\Sigma}_j$ (j = 1, 2) in Tables 4.19 and 4.20 are nonsingular, another method of estimating $C_j$ based on expression (4.2.64) is

$$\hat{C}_{j_{\text{wls}}} = (\hat{\Sigma}_j \bar{\Sigma}_j)^{-1} \hat{\Sigma}_j \bar{\Sigma}_j (j = 1, 2).$$

Then, for the nonmetro data on personal satisfaction the estimator is

$$\hat{C}_{1_{\text{wls}}} = 1.96148,$$

and for the metro data the estimator is

$$\hat{C}_{2_{\text{wls}}} = 1.19843.$$

A test of fit for the model can be obtained by

$$X^2_{j_{\text{wls}}} = (\hat{\Sigma}_j \bar{\Sigma}_j)^{-1} - \hat{C}_{j_{\text{wls}}} \hat{\Sigma}_j \bar{\Sigma}_j (j = 1, 2),$$

which is approximately distributed as a chi-square random variable with two degrees of freedom. The numerical values are

$$X^2_{1_{\text{wls}}} = .0788$$

and

$$X^2_{2_{\text{wls}}} = .7452.$$
suggesting that the model is a good fit for both areas. Note that $X_{2w}^2$ is considerably smaller than $X_{B2}^2$ and $X_{w2}^2$. This may be due to the fact that $X_{2w}^2$ uses a moment estimator for the covariance matrix of the estimated covariances. The other statistics, $X_{B2}^2$ and $X_{w2}^2$ use normal assumptions to estimate the covariance matrix for the estimated covariances. These normal assumptions may not be good approximations for small or moderate numbers of clusters even when the Dirichlet-Multinomial model is correct. Consequently, the $X_{B2}^2$ and $X_{w2}^2$ lack-of-fit tests may have a tendency to be too large because inadequacies in the estimation of the covariance matrix for the estimated covariances. The same phenomenon is observed for the nonmetro data, $X_{1w}^2$ is smaller than $X_{B1}^2$ and $X_{w1}^2$.

A similar analysis as carried out for the personal satisfaction data was done for the community satisfaction data. The estimated vector of proportions for the nonmetro area is

$$\hat{\pi}_1 = (0.25000, .61458, .13542)'$$

and the estimated vector of proportions for the metro area is

$$\hat{\pi}_2 = (0.25773, .60825, .13402)'$$

The hypothesis of interest is

$$H_0: \pi_j = \pi_0, \quad j = 1, 2,$$

where $\pi_0$ is an unknown probability vector of dimension three.
and $\Pi_j$ is the probability vector of the $j^{th}$ subpopulation. Under the Dirichlet-Multinomial model, the estimated common probability vector (4.3.46) is

$$\hat{\Pi}_0 = \left[ \frac{2}{\sum \sum N \hat{c}_{ij} - 1} \right] \sum \sum N \hat{c}_{ij} \hat{\Pi}_j .$$

(4.4.62)

Several methods of estimating $C_j$ were given in Section 4.2. These different methods of obtaining an estimator of $C_j$ are now considered. The factor $C_j$ is a measure of the clustering effect.

One estimator of $C_j$ is the Brier estimator, $\hat{C}_{jB}$, given in expression (4.2.52). The numerical value of $\hat{C}_{jB}$ for the nonmetro area is

$$\hat{C}_{1B} = 1.5955$$

and for the metro area is

$$\hat{C}_{2B} = 1.0140 .$$

For the nonmetro area, the number of categories is $I = 3$, the number of clusters is $S_1 = 20$, the cluster size is $n_1 = 5$. For the metro area, the number of categories is also $I = 3$, the number of clusters is $S_2 = 20$, the cluster size is $n_2 = 5$. Using the values for $\hat{C}_{1B}$ and $\hat{C}_{2B}$, the estimated probability vector is
\[ \hat{\Pi}_0 = \left[ \sum_{i=1}^{2} \left( \sum_{j=1}^{\hat{N}_i} \hat{a}_i \right)^{-1} \right]^{-1} = (0.254124, 0.611203, 0.134673) \] (4.4.63)

The weights for this linear combination are

\[ \hat{N}_1 \hat{C}_1 B \left( \sum_{j=1}^{\hat{N}_1} \hat{C}_1^{-1} \right)^{-1} = 0.4653 \]

and

\[ \hat{N}_2 \hat{C}_2 B \left( \sum_{j=1}^{\hat{N}_2} \hat{C}_2^{-1} \right)^{-1} = 0.5347. \]

The statistic for testing the equality of the vectors of proportions for the two areas is

\[ X^2_{DMI} = 0.0145. \]

When the model is true, \( X^2_{DMI} \) is distributed as a chi-square random variable with two degrees of freedom. This test provides no evidence for rejecting the equality of the vectors of proportions.

The second method of constructing \( \hat{C}_1 B \) is based on the generalized least squares technique. The estimator of the covariance matrix for a simple random sample of five observations from a multinomial distribution with probability vector \( \hat{N}_1 \) is
The vector \( \hat{\mathbf{x}}_1 \) obtained from \( \hat{\Sigma}_{ml} \) forms the right side of the regression equation

\[
\hat{\mathbf{x}}_1 = \mathbf{C}_1 \hat{\mathbf{x}}_1 + \epsilon. \tag{4.4.64}
\]

The estimated covariance matrix constructed using the cluster sampling formula (4.2.66) is

\[
\hat{\Sigma}_{DML} = \begin{bmatrix}
446.32 & -387.37 & -22.485 \\
-387.37 & 585.27 & -208.46 \\
-22.48 & -208.246 & 230.731
\end{bmatrix} \times 10^{-4}.
\]

The upper left 2x2 matrix of \( \hat{\Sigma}_{DML} \) is

\[
\hat{V}_{ppl} = \begin{bmatrix}
446.32 & -387.37 \\
-387.37 & 585.27
\end{bmatrix} \times 10^{-4}.
\]

A graph of the elements of \( \hat{V}_{ppl} \), denoted by \( \hat{V}_{x_1} \), versus the elements of \( \hat{\Sigma}_{ml} \), denoted by \( \hat{\Sigma}_{x_1} \), is given in Figure 4.9. In Figure 4.9, all points lie in the first and second quadrant and none of the covariance estimates are positive.
Figure 4.9. A graph of the elements of the estimated covariance matrix under cluster sampling, $\hat{V}$, versus the elements of the estimated covariance matrix under Multinomial sampling, $\hat{W}$, for the nonmetro area (community satisfaction).
Table 4.25a. Cluster estimated covariance matrix of sample covariances for the nonmetro area (community satisfaction)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Cluster covariance estimate x 10^4</th>
<th>Identification</th>
<th>Cluster covariance estimate x 10^4</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,11</td>
<td>0.83012</td>
<td>12,12</td>
<td>1.80022</td>
</tr>
<tr>
<td>11,12</td>
<td>-1.00203</td>
<td>12,22</td>
<td>-2.70536</td>
</tr>
<tr>
<td>11,22</td>
<td>1.35774</td>
<td>22,22</td>
<td>4.32784</td>
</tr>
</tbody>
</table>

Table 4.25b. Cluster estimated covariance matrix of sample covariances for the metro area (community satisfaction)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Cluster covariance estimate x 10^4</th>
<th>Identification</th>
<th>Cluster covariance estimate x 10^4</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,11</td>
<td>1.0236</td>
<td>12,12</td>
<td>1.4077</td>
</tr>
<tr>
<td>11,12</td>
<td>-1.0316</td>
<td>12,22</td>
<td>-1.3513</td>
</tr>
<tr>
<td>11,22</td>
<td>0.8396</td>
<td>22,22</td>
<td>2.4592</td>
</tr>
</tbody>
</table>

For the metro area, similar statistics can be obtained based on the same procedure for the nonmetro area. Assuming a multinomial model, the estimated covariance matrix for the metro area data is

\[ A_{\Sigma m2} = \begin{bmatrix} 375.00 & -300.00 \\ -300.00 & 480.00 \end{bmatrix} \times 10^{-4}, \]

and
\[ \hat{\Sigma}_2 = (375.00, -300.00, 480.00) \times 10^{-4}. \]

Similarly, from \( \hat{\Sigma}_{DM2}^{\prime} \),

\[ \hat{\Sigma}_2 = (373.68, -336.84, 547.37) \times 10^{-4}. \]

A graph of \( \hat{\Sigma}_2 \) versus \( \hat{\Sigma}_2 \) is given in Figure 4.10. As in the case of the nonmetro data, there are no second quadrant points in Figure 4.10.

Similar statistics can be obtained for the nonmetro area (community satisfaction) based on the generalized regression technique. The Cholesky decomposition of \( \hat{\Sigma}_{ml}^{-1} \) is given by

\[ \hat{\Sigma}_{ml}^{-1} = \hat{R}^t \hat{R}, \]

where

\[ \hat{R} = \begin{pmatrix} 7.51982 & 4.74936 \\ 0 & 4.60678 \end{pmatrix}. \]

Then,

\[ \hat{V}_{YY} = \begin{pmatrix} 1.07703 & -0.0614154 \\ -0.0614154 & 1.24207 \end{pmatrix}. \]

The generalized least squares estimator, defined in (4.2.71), is

\[ \hat{C}_{Bl} = 1.15955, \]

which is identical to the value of \( \hat{C}_{1b} \) obtained in (4.4.44).
Figure 4.10. A graph of the elements of the estimated covariance matrix under cluster sampling, $\hat{V}$, versus the elements of the estimated covariance matrix under Multinomial sampling, $\hat{W}$, for the metro area (community satisfaction).
The matrix of estimated covariances of the estimated covariances is given in Table 4.26. The cluster standard errors of the estimated covariances for the transformed variables are given in the last column of Table 4.27.

### Table 4.26. Cluster estimated covariance matrix of the transformed covariances for the nonmetro area (community satisfaction)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Estimated covariance</th>
<th>Identification</th>
<th>Estimated covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,11</td>
<td>0.09786</td>
<td>12,12</td>
<td>0.08001</td>
</tr>
<tr>
<td>11,12</td>
<td>0.04738</td>
<td>12,22</td>
<td>0.01416</td>
</tr>
<tr>
<td>11,22</td>
<td>0.01331</td>
<td>22,22</td>
<td>0.10506</td>
</tr>
</tbody>
</table>

### Table 4.27. Transformed estimated covariance matrix for proportions for the nonmetro area (community satisfaction)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Model estimate of $\hat{V}_{YY2}$</th>
<th>Elements of $\hat{V}_{YY2}$</th>
<th>Model standard error</th>
<th>Cluster standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>1.15955</td>
<td>1.0770</td>
<td>0.33641</td>
<td>0.31283</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>-0.0614</td>
<td>0.23788</td>
<td>0.28286</td>
</tr>
<tr>
<td>22</td>
<td>1.15955</td>
<td>1.2421</td>
<td>0.33641</td>
<td>0.32414</td>
</tr>
</tbody>
</table>

Letting $\hat{D}_{wl}$ be the diagonal matrix composed of the diagonal elements of the matrix in Table 4.26, the estimated generalized least
squares estimator defined in (4.2.75) is

$$\hat{C}_{wl} = 1.15662.$$  

The estimated variance of \( \hat{C}_w \) (4.4.17) is

$$\hat{V}[\hat{C}_w] = 0.05067,$$

and the lack-of-fit statistic (4.2.75b) is

$$\chi^2_{wl} = 0.181363.$$

One estimated variance of \( \hat{C}_{Bl} \) following (4.4.19) is

$$\hat{V}[\hat{C}_{Bl}] = 0.05739.$$

A second estimate of the variance based on the elements of \( \hat{V}[\text{vech } v_{YY1}] \) in Table 4.26 is

$$\hat{V}[\hat{C}_{Bl}] = 0.05659,$$

and the standard error is 0.23789.

The lack-of-fit statistic following (4.4.58) is

$$\chi^2_{Bl} = 0.1870,$$

where the ratio of the diagonal elements of \( \hat{V}[\text{vech } \hat{V}_{YY1}] \) to the diagonal elements of
The statistic $X_{B1}^2$ is approximately distributed as a chi-square random variable with two degrees of freedom when the null hypothesis is true. On the basis of this the Dirichlet-Multinomial model is not rejected.

For the metro data on community satisfaction, the Cholesky decomposition of $\hat{\Sigma}_{m2}$ is given by

$$\hat{\Sigma}_{m2}^{-1} = \hat{R}_2 \hat{R}_2^\top$$

(4.4.65)

where

$$\hat{R}_2 = \begin{pmatrix} 7.30297 & 4.56435 \\ 0 & 4.56435 \end{pmatrix}.$$  

Let $\hat{V}_{YY2}$ denote the upper left 2x2 matrix of the estimated covariance of the linear combination of $\hat{\pi}_{t2}$'s defined by

$$\hat{\pi}_{t2} = \hat{R}_2 \hat{\pi}_{t2} - \hat{\pi}_2.$$  

(4.4.66)

Then,

$$\hat{V}_{YY2} = \hat{R}_2 \hat{V}_{pp2} \hat{R}_2^\top$$

$$= \begin{pmatrix} 0.88772 & 0.01754 \\ 0.01754 & 1.14035 \end{pmatrix}.$$
The generalized least squares estimator defined in (4.2.71) is

\[ \hat{C}_{b2} = 1.01404. \]

The estimator \( \hat{C}_{b2} \) is identical to that previously computed, as it should be.

The estimated covariance matrix for the estimated covariances of the transformed \( \hat{U}_{t2} \)'s was computed by applying formula (4.2.66) to the \( \hat{Y}_{tj} \)'s. The matrix of estimated covariances of the estimated covariances is given in Table 4.28. The cluster standard errors of the estimated covariances for the transformed variables are given in the last column of Table 4.29.

Table 4.28. Cluster estimated covariance matrix of the transformed covariances for the metro area

<table>
<thead>
<tr>
<th>Identification</th>
<th>Estimated covariance</th>
<th>Identification</th>
<th>Estimated covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,11</td>
<td>0.06897</td>
<td>12,12</td>
<td>0.01122</td>
</tr>
<tr>
<td>11,12</td>
<td>-0.00259</td>
<td>12,22</td>
<td>-0.06623</td>
</tr>
<tr>
<td>11,22</td>
<td>-0.03938</td>
<td>22,22</td>
<td>0.16967</td>
</tr>
</tbody>
</table>
Table 4.29. Transformed estimated covariance matrix for proportions for the metro area

<table>
<thead>
<tr>
<th>Identification</th>
<th>Model estimate of $\hat{V}_{YY1}$</th>
<th>Elements of $\hat{V}_{YY1}$</th>
<th>Model standard error</th>
<th>Cluster standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>1.01404</td>
<td>0.8877</td>
<td>0.3161</td>
<td>0.2626</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0.0175</td>
<td>0.2235</td>
<td>0.1059</td>
</tr>
<tr>
<td>22</td>
<td>1.01404</td>
<td>1.14035</td>
<td>0.3161</td>
<td>0.4119</td>
</tr>
</tbody>
</table>

Letting $\hat{D}_{w2}$ be the diagonal matrix composed of the diagonal elements of the matrix in Table 4.28, the estimated generalized least squares estimator defined in (4.2.75) is

$$\hat{C}_{w2} = 0.9607.$$  

The estimated variance of $\hat{C}_{w2}$ is

$$\hat{V}[\hat{C}_{w2}] = 0.0490,$$

and the lack-of-fit statistic is

$$\chi^2_{w2} = 0.2949.$$  

The estimate of $C_2$, obtained using the matrix $D_{w2}$, is smaller than that obtained using the matrix $D_B$. The estimate of $\hat{C}_{w2}$ is less than one, outside the allowable range of $1 \leq C \leq 5$.  

One estimator of the variance of $\hat{C}_{B2}$, based on expression (4.4.44),
is

\[ V_{\widehat{C}_{B2}} = 0.03997. \]

A second variance estimator can be computed under the assumption that 
\( V[\text{vech } \hat{V}_{YY}] \) is a diagonal matrix proportional to \( D_B \), which is

\[ D_B = (2,1,2). \]

For the metro data, the ratio of the diagonal elements of \( V[\text{vech } \hat{V}_{YY}] \) to the diagonal elements of \( D_B \) is 0.04997. This provides an alternative estimate of the variance

\[ V_{\widehat{C}_{B2}} = 0.04997, \]

and the standard error of \( \hat{C}_{B2} \) is 0.2236. The two methods of estimating the variance of \( \hat{C}_{B2} \) result in similar values.

If \( \Sigma_2 \) is known, a lack-of-fit test based on expression (4.2.73)

is

\[ X^2_{B2} = (H_B D_B^{-1} H_B - \hat{C}_{B2} A_B^{-1} D_B A_B^{-1})(0.04997)^{-1} \]

\[ = 0.32546. \] (4.4.67)

This statistic is approximately distributed as a chi-square random variable with two degrees of freedom when the null model is true.

On the basis of \( X^2_{B2} \) the model is considered to be reasonable.

Since \( \hat{\Sigma}_j \), the estimated covariance matrix of \( \hat{\Sigma}_j \) (Tables 4.25a
and 4.25b), is nonsingular another method of estimating $C_j$, based on expression (4.2.64), is

$$\hat{C}_{j \text{wls}} = (\hat{H}_j)^{-1} S_j \hat{H}_j^{-1} \hat{S}_j \hat{C}_{j \text{wls}}^{-1} \hat{S}_j \hat{H}_j^{-1}.$$  

Then, for the nonmetro area the estimator is

$$\hat{C}_{1 \text{wls}} = 1.1138$$

and for the metro area the estimator is

$$\hat{C}_{2 \text{wls}} = .9387.$$  

A test of fit for the model can be obtained by

$$x^2_{j \text{wls}} = (\hat{H}_j)^{-1} S_j \hat{H}_j^{-1} \hat{S}_j \hat{C}_{j \text{wls}}^{-1} \hat{S}_j \hat{H}_j^{-1}$$

which is approximately distributed as a chi-square random variable with three degrees of freedom. The numerical values are

$$x^2_{1 \text{wls}} = .3215 \text{ (nonmetro)}$$

and

$$x^2_{2 \text{wls}} = .1311 \text{ (metro)},$$

suggesting that the model is reasonable for both areas.

A summary of the $C$ estimators for the nonmetro and the metro areas for the personal satisfaction data is given in Table 4.30a. The
test statistic

\[ x^2_{DMI} = \sum_{j=1}^{2} \frac{\hat{C}_j}{N} \sum_{i=1}^{3} \frac{\hat{A}_i}{\hat{N}_{ij} - \hat{N}_{io}}^2 \hat{N}_{io} \]

is computed for each of the pairs of \( C_j \) estimators and values are given in the fourth column of Table 4.30a. The estimator \( \hat{C}_P \) is one for both the metro and nonmetro areas. The value of \( x^2_{DMI} \) using the pair of \( \hat{C}_P \) estimators is equivalent to the Pearson test statistic value for independence. The values of \( x^2_{DMI} \), given in Table 4.30a, suggest that there is no difference in personal satisfaction for the metro and nonmetro areas at the five percent level.

Table 4.30a. A summary of \( C_j \) estimators and the test statistic value for testing the hypothesis \( H_0: \Pi_{ij} = \Pi_o \) for the nonmetro and metro area (personal satisfaction)

<table>
<thead>
<tr>
<th>C estimators</th>
<th>Nonmetro</th>
<th>Metro</th>
<th>( x^2_{DMI} )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{C}_B )</td>
<td>2.0768</td>
<td>1.6212</td>
<td>4.593</td>
<td>( p &gt; .10 )</td>
</tr>
<tr>
<td>( \hat{C}_W )</td>
<td>1.9727</td>
<td>1.6319</td>
<td>4.714</td>
<td>( .05 &lt; p &lt; .10 )</td>
</tr>
<tr>
<td>( \hat{C}_P )</td>
<td>1.00</td>
<td>1.00</td>
<td>8.514</td>
<td>( .025 &lt; p &lt; .01 )</td>
</tr>
<tr>
<td>( \hat{C}_{WLS} )</td>
<td>1.9615</td>
<td>1.1984</td>
<td>5.373</td>
<td>( .10 &lt; p &lt; .05 )</td>
</tr>
</tbody>
</table>

A summary of the \( C_j \) estimators and the test statistic values for the nonmetro and the metro areas for community satisfaction is given in Table 4.30b. The p-values for the test statistics, which are approxi-
Table 4.30b. A summary of $C_j$ estimators and the test statistic value for testing the hypothesis $H_0: \pi_j = \pi_o$ for the nonmetro and metro area (community satisfaction)

<table>
<thead>
<tr>
<th>C estimators</th>
<th>Nonmetro</th>
<th>Metro</th>
<th>$X^2_{DMI}$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{C}_B$</td>
<td>1.5955</td>
<td>1.014</td>
<td>.0145</td>
<td>p &gt; .50</td>
</tr>
<tr>
<td>$\hat{C}_W$</td>
<td>1.1566</td>
<td>(0.9607)</td>
<td>.0146</td>
<td>p &gt; .50</td>
</tr>
<tr>
<td>$\hat{C}_P$</td>
<td>1.000</td>
<td>1.000</td>
<td>.0158</td>
<td>p &gt; .50</td>
</tr>
<tr>
<td>$\hat{C}_{WLS}$</td>
<td>1.1138</td>
<td>(.9387)</td>
<td>.0149</td>
<td>p &gt; .50</td>
</tr>
</tbody>
</table>

Naturally distributed as chi-square random variables with two degrees of freedom, indicate that the hypothesis $(H_0: \pi_j = \pi_o)$ should not be rejected. The estimators $\hat{C}_W$ and $\hat{C}_m$ for the metro data given in parentheses in Table 4.30b are less than one. In computing $X^2_{DMI}$, $\hat{C}_W$ and $\hat{C}_m$ are assumed to be one for the metro data. The $\hat{C}$ estimates for the metro data are not significantly different from one.

The largest value of $X^2_{DMI}$ occurs when $\hat{C}_P$ estimates are used.

In this example, there is a moderately large number of clusters relative to the number of categories. However, for any cluster there are no more than five observations and a considerable number of observed zeros. In spite of the small cluster sample size, the estimators for the $C_j$'s and the goodness-of-fit statistics agree quite well. This was not the case in the previous examples where there were few clusters and the cluster sample sizes were large. From these few examples, it
appears that a large number of clusters is more important than large cluster sample sizes for the estimation and testing procedures to be reliable.

Table 4.31 summarizes the statistics obtained in Example 4.1 (works of Herodotus and Thucydides), Example 4.2 (habitat preference of wild turkeys), and Example 4.3 (personal and community satisfaction). The estimators \( \hat{C}_w \) and \( \hat{C}_B \) are very similar in value when the model is a good fit. The two estimates given for the standard error of \( \hat{C}_B \) are alike in value when the model is accepted. The lack-of-fit statistic, \( X_B^2 \), is usually smaller in value than lack-of-fit statistic, \( X_w^2 \). However, when the model is a good fit the two statistics \( (X_B^2, X_w^2) \) are very close in numerical value.

Examples 4.1, 4.2 and 4.3 differ in several ways. Example 4.1, concerning the author of Greek prose, had a large number of observations per cluster, but few clusters in relation to the number of categories, for the two subpopulations. Example 4.2, concerning the turkeys, had fewer observations per cluster and more subpopulations than in Example 4.1. However, some of the subpopulations had few clusters in relation to the number of categories, while others had a great deal of clusters in relation to the number of categories. Example 4.3, which concerns satisfaction levels in two areas, had several clusters in response to the number of categories. The clusters in Example 4.3 had very few observations in any cluster. The Dirichlet-Multinomial model may be a reasonable model for clustering. However, the lack-of-fit test which assumes normality performs well when there
Table 4.31. A summary of the statistics for Examples 4.1, 4.2 and 4.3

<table>
<thead>
<tr>
<th>Subpopulation</th>
<th>$\widehat{C}_w^a$</th>
<th>$\widehat{C}_B^b$</th>
<th>s.e($\widehat{C}_w^c$)</th>
<th>s.e($\widehat{C}_B^d$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herodotus</td>
<td>1.418</td>
<td>2.0054</td>
<td>0.3674</td>
<td>1.0817</td>
</tr>
<tr>
<td>Thucydides</td>
<td>0.3691</td>
<td>1.1724</td>
<td>0.0793</td>
<td>0.1853</td>
</tr>
<tr>
<td>Juvenile male</td>
<td>0.6004</td>
<td>3.1717</td>
<td>0.1321</td>
<td>0.6921</td>
</tr>
<tr>
<td>Adult male</td>
<td>1.1635</td>
<td>3.8863</td>
<td>1.0787</td>
<td>1.1848</td>
</tr>
<tr>
<td>Juvenile female</td>
<td>1.2880</td>
<td>3.0609</td>
<td>0.2715</td>
<td>0.7788</td>
</tr>
<tr>
<td>Adult female</td>
<td>3.4854</td>
<td>4.2013</td>
<td>0.4365</td>
<td>0.8116</td>
</tr>
<tr>
<td>Nonmetro (personal)</td>
<td>1.9727</td>
<td>2.0768</td>
<td>0.4838</td>
<td>0.6111</td>
</tr>
<tr>
<td>Metro (personal)</td>
<td>1.6319</td>
<td>1.6212</td>
<td>0.2830</td>
<td>0.2441</td>
</tr>
<tr>
<td>Nonmetro (community)</td>
<td>1.1566</td>
<td>1.5955</td>
<td>0.2261</td>
<td>0.2396</td>
</tr>
<tr>
<td>Metro (community)</td>
<td>0.9607</td>
<td>1.0140</td>
<td>0.2214</td>
<td>0.1999</td>
</tr>
</tbody>
</table>

$\widehat{C}_w^a$ is the C estimator (4.2.75a).
$\widehat{C}_B^b$ is the C estimator (4.2.71).
s.e($\widehat{C}_w^c$) is the standard error of $\widehat{C}_w$ (4.4.17).
s.e($\widehat{C}_B^d$) is the standard error of $\widehat{C}_B$ (4.4.19).
$s.e_1(\widehat{C}_B^d)$ is the standard error of $\widehat{C}_B$ (4.2.66).
$\chi^2_w$ is the lack-of-fit statistic (4.2.75b).
$\chi^2_B$ is the lack-of-fit statistic (4.2.73).
$S$ is the number of clusters.
$I$ is the number of categories.
$n$ is the size per cluster.
Model rejected at one percent level.
<table>
<thead>
<tr>
<th>$s.e_{2}^e(C_B)^{\Delta}$</th>
<th>$x_{w}^{f}$</th>
<th>$x_{B}^{g}$</th>
<th>$s^{h}$</th>
<th>$i$</th>
<th>$n^{j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7030</td>
<td>9.40</td>
<td>11.69</td>
<td>9</td>
<td>5</td>
<td>200</td>
</tr>
<tr>
<td>0.2311</td>
<td>$50.75^k$</td>
<td>22.11</td>
<td>8</td>
<td>5</td>
<td>200</td>
</tr>
<tr>
<td>0.6812</td>
<td>$55.62^k$</td>
<td>$41.82^k$</td>
<td>5</td>
<td>7</td>
<td>50</td>
</tr>
<tr>
<td>1.0688</td>
<td>$31.35^k$</td>
<td>21.34</td>
<td>4</td>
<td>7</td>
<td>50</td>
</tr>
<tr>
<td>0.5746</td>
<td>$42.86^k$</td>
<td>$46.68^k$</td>
<td>10</td>
<td>7</td>
<td>50</td>
</tr>
<tr>
<td>0.6361</td>
<td>23.45</td>
<td>20.84</td>
<td>21</td>
<td>7</td>
<td>50</td>
</tr>
<tr>
<td>0.5794</td>
<td>0.10</td>
<td>0.10</td>
<td>20</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>0.2811</td>
<td>6.35</td>
<td>5.24</td>
<td>20</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>0.2379</td>
<td>0.18</td>
<td>0.19</td>
<td>20</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>0.2235</td>
<td>0.29</td>
<td>0.32</td>
<td>20</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>
are a great number of clusters in relation to the number of categories. The number of observations per cluster makes very little difference to the performance of the model, once the number of clusters is greater than the product of the number of categories times half of one less than the number of categories. When this relationship among categories and number of clusters is satisfied, the corresponding estimators suggested throughout the examples are about equal.

Attention was given to constructing Wald Statistics for several hypotheses of the multinomial type. However, the estimation of the covariance matrix to be used in the construction of the Wald Statistics requires large numbers of primary sampling units.

Chapter 4 presented a Dirichlet model that may be useful in cluster sampling. The Dirichlet model assumes that the covariance matrix under cluster sampling is a multiple of the covariance matrix under multinomial sampling. This assumption is somewhat restrictive, but when this restrictive assumption is satisfied or nearly satisfied, the estimation of the parameter of the distribution is of utmost importance. The methods presented to estimate the multiplier were based on regression techniques and approximate normality assumptions. When the number of clusters is small, the different methods results in estimates that differ considerably. Differences were present even when the sizes of the clusters were large. However, the methods gave similar results for the multiplier when the ratio of the number of clusters to the number of categories is 'large'. There is a need to investigate the robustness of the statistics and to provide guidance
on how large the ratio of the number of clusters to the number of categories must be for the methods to begin to tell the same thing. It is the opinion of the author that the amount of data required to estimate the parameter of the Dirichlet model is less than that needed for the construction of Wald Statistics.

The Dirichlet-Multinomial model provides test statistic values that are smaller than the usual Pearson statistic, thereby correcting for the clustering effect. The reduction factor for the test statistic value given by the Dirichlet-Multinomial assumption is inversely proportional to the number of clusters per subpopulation and the number of subpopulations compared. Despite the restrictiveness of the Dirichlet-Multinomial model, the model seemed to fit the data of some examples.

The lack-of-fit tests for the Dirichlet model constructed in Chapter 4 are asymptotic tests. There is a need to further investigate the performance of \( \chi^2_{DMI} \) both for a small number of clusters and a large number of clusters. The performance of \( \chi^2_{DMI} \) in situations where clustering is present, but the Dirichlet model is not satisfied should also be studied further. Research is required to adequately resolve the problem when cluster sizes are unequal.
5. REFERENCES


Okamoto, M. 1963. Chi-square statistic based on the pooled frequencies of several observations. Biometrika 50;524-528.


Pearson, K. 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Phil. Mag. 50;157-175.


Wald, A. 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans. Amer. Math. Soc. 54:426-482.


6. ACKNOWLEDGMENTS

It gives me great pleasure to acknowledge my debt to the faculty and staff of the Statistical Laboratory, at Iowa State University, for their contribution to my development as a statistician.

A number of people must be singled out for their special contribution to my achievement at Iowa State. Dr. Kenneth Koehler, one of my major professors, is chiefly responsible for my interest in categorical data. He has been a great encouragement and a consistent source of direction, criticism, and insight into the problems I have attempted to solve. Professor Wayne Fuller, my other major professor, has been very helpful and never without a word of advice.

I would also like to express my thanks to Professor Oscar Kempthorne who is in part responsible for my coming to Iowa State University. He has been very consistent in his interest and the well being of my family. My other committee members, Professors Vince Sposito and Roger Homer, have also been very helpful and thoughtful in my attitudes as an applied statistician.

I am grateful to Mrs. Sharon Shepard for her expert typing in the preparation of this dissertation.

I would like to thank the Government of Trinidad and Tobago who provided financial support in the first two (2) years of my pursuit of this degree. In addition, special thanks to Mr. Leo Pujadas the Director of Statistics (Trinidad and Tobago) who assisted in whatever way possible. To my sister Mrs. June Berment, I extend my deepest thanks for her continued financial support. Special thanks go to
Dr. George Jackson, Assistant to the Vice President for Student Affairs and Director of Minority Student Affairs for his continued support throughout my stay at Iowa State.

To my daughters Rhonda and Roxanne, I am very grateful for their childlike understanding towards the task that stood before me. Finally, to my wife, Patricia I owe the deepest debt. She has been very understanding and certainly has contributed greatly in making this achievement possible. To her, I dedicate this dissertation. To God be all glory for his strength and blessings.
## Table 7.1. The occurrence of *einai* in sentences in the works of Herodotus (Morton (1965))

<table>
<thead>
<tr>
<th>No. of occurrences of <em>einai</em></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. <em>einai</em></td>
<td>147</td>
<td>107</td>
<td>142</td>
<td>144</td>
<td>140</td>
<td>156</td>
<td>142</td>
<td>143</td>
<td>146</td>
<td>1,267</td>
</tr>
<tr>
<td>One</td>
<td>47</td>
<td>65</td>
<td>44</td>
<td>45</td>
<td>49</td>
<td>36</td>
<td>46</td>
<td>48</td>
<td>48</td>
<td>428</td>
</tr>
<tr>
<td>Two</td>
<td>6</td>
<td>22</td>
<td>13</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>11</td>
<td>8</td>
<td>5</td>
<td>89</td>
</tr>
<tr>
<td>Three</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Four or more</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>No. of sentences in sample</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>1,800</td>
</tr>
</tbody>
</table>

## Table 7.2. The occurrence of *einai* in sentences in the works of Thucydides (Morton (1965))

<table>
<thead>
<tr>
<th>No. of occurrences of <em>einai</em></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. <em>einai</em></td>
<td>126</td>
<td>129</td>
<td>141</td>
<td>131</td>
<td>129</td>
<td>141</td>
<td>147</td>
<td>147</td>
<td>1091</td>
</tr>
<tr>
<td>One</td>
<td>56</td>
<td>58</td>
<td>45</td>
<td>55</td>
<td>57</td>
<td>45</td>
<td>42</td>
<td>46</td>
<td>404</td>
</tr>
<tr>
<td>Two</td>
<td>16</td>
<td>8</td>
<td>12</td>
<td>9</td>
<td>11</td>
<td>9</td>
<td>5</td>
<td>16</td>
<td>82</td>
</tr>
<tr>
<td>Three</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Four or more</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>No. of sentences in sample</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>1600</td>
</tr>
</tbody>
</table>
Table 7.3. Juvenile male wild turkeys in Iowa

<table>
<thead>
<tr>
<th>Turkey</th>
<th>Brush</th>
<th>Conifer</th>
<th>Corn/bean</th>
<th>Grass</th>
<th>Lowland</th>
<th>Oakpole</th>
<th>Oaksaw</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>6</td>
<td>7</td>
<td>18</td>
<td>2</td>
<td>13</td>
<td>4</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>13</td>
<td>6</td>
<td>1</td>
<td>18</td>
<td>9</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>9</td>
<td>13</td>
<td>12</td>
<td>1</td>
<td>13</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>5</td>
<td>15</td>
<td>2</td>
<td>0</td>
<td>20</td>
<td>6</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>7</td>
<td>29</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>28</td>
<td>77</td>
<td>38</td>
<td>4</td>
<td>78</td>
<td>20</td>
<td>250</td>
</tr>
</tbody>
</table>

Table 7.4. Adult male wild turkeys in Iowa

<table>
<thead>
<tr>
<th>Turkey</th>
<th>Brush</th>
<th>Conifer</th>
<th>Corn/bean</th>
<th>Grass</th>
<th>Lowland</th>
<th>Oakpole</th>
<th>Oaksaw</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>11</td>
<td>2</td>
<td>1</td>
<td>34</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>35</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0</td>
<td>23</td>
<td>0</td>
<td>7</td>
<td>16</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>12</td>
<td>4</td>
<td>26</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>3</td>
<td>50</td>
<td>16</td>
<td>14</td>
<td>111</td>
<td>0</td>
<td>200</td>
</tr>
</tbody>
</table>
Table 7.5. Juvenile female wild turkeys in Iowa

<table>
<thead>
<tr>
<th>Turkey</th>
<th>Brush</th>
<th>Conifer</th>
<th>Corn/bean</th>
<th>Grass</th>
<th>Lowland</th>
<th>Oakpole</th>
<th>Oaksaw</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>0</td>
<td>11</td>
<td>11</td>
<td>2</td>
<td>16</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>6</td>
<td>14</td>
<td>6</td>
<td>1</td>
<td>16</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>8</td>
<td>3</td>
<td>16</td>
<td>3</td>
<td>14</td>
<td>3</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>8</td>
<td>19</td>
<td>3</td>
<td>0</td>
<td>15</td>
<td>4</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>0</td>
<td>11</td>
<td>18</td>
<td>4</td>
<td>7</td>
<td>7</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>4</td>
<td>11</td>
<td>4</td>
<td>1</td>
<td>24</td>
<td>6</td>
<td>50</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>15</td>
<td>15</td>
<td>6</td>
<td>1</td>
<td>12</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>13</td>
<td>15</td>
<td>3</td>
<td>1</td>
<td>16</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>8</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>17</td>
<td>3</td>
<td>50</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>8</td>
<td>34</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>70</td>
<td>147</td>
<td>72</td>
<td>14</td>
<td>145</td>
<td>34</td>
<td>500</td>
</tr>
</tbody>
</table>
Table 7.6. Adult female wild turkeys in Iowa

<table>
<thead>
<tr>
<th>Turkey</th>
<th>Brush</th>
<th>Conifer</th>
<th>Corn/bean</th>
<th>Grass</th>
<th>Lowland</th>
<th>Oakpole</th>
<th>Oaksaw</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>7</td>
<td>15</td>
<td>3</td>
<td>1</td>
<td>18</td>
<td>3</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>16</td>
<td>7</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>1</td>
<td>20</td>
<td>4</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>6</td>
<td>15</td>
<td>5</td>
<td>0</td>
<td>17</td>
<td>6</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>4</td>
<td>9</td>
<td>7</td>
<td>2</td>
<td>21</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>14</td>
<td>18</td>
<td>9</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>7</td>
<td>23</td>
<td>7</td>
<td>0</td>
<td>11</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>6</td>
<td>13</td>
<td>7</td>
<td>0</td>
<td>19</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>6</td>
<td>14</td>
<td>3</td>
<td>1</td>
<td>22</td>
<td>3</td>
<td>50</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>9</td>
<td>14</td>
<td>2</td>
<td>1</td>
<td>18</td>
<td>6</td>
<td>50</td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>9</td>
<td>17</td>
<td>10</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>3</td>
<td>15</td>
<td>3</td>
<td>15</td>
<td>11</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>20</td>
<td>4</td>
<td>0</td>
<td>11</td>
<td>7</td>
<td>4</td>
<td>50</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>15</td>
<td>5</td>
<td>10</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>17</td>
<td>6</td>
<td>3</td>
<td>15</td>
<td>0</td>
<td>9</td>
<td>17</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>9</td>
<td>0</td>
<td>23</td>
<td>13</td>
<td>50</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>11</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>11</td>
<td>50</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>13</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>4</td>
<td>50</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>5</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>163</td>
<td>251</td>
<td>109</td>
<td>48</td>
<td>367</td>
<td>72</td>
<td>1050</td>
</tr>
</tbody>
</table>
Table 7.7. Personal satisfaction for nonmetro areas in Minnesota

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Unsatisfied</th>
<th>Satisfied</th>
<th>Very satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 7.8. Personal satisfaction for metro areas in Minnesota

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Unsatisfied</th>
<th>Satisfied</th>
<th>Very satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 7.9. Community satisfaction for nonmetro areas in Minnesota

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Unsatisfied</th>
<th>Satisfied</th>
<th>Very satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 7.10. Community satisfaction for metro areas in Minnesota

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Unsatisfied</th>
<th>Satisfied</th>
<th>Very satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 7.11. Socioeconomic status, intelligence and college plans for Wisconsin school boys

<table>
<thead>
<tr>
<th>IQ</th>
<th>College plans</th>
<th>Low economic level</th>
<th>Low middle economic level</th>
<th>Upper middle economic level</th>
<th>High economic level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Plans to go to college</td>
<td>17</td>
<td>29</td>
<td>55</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>No plans for college</td>
<td>413</td>
<td>316</td>
<td>257</td>
<td>105</td>
</tr>
<tr>
<td>Low middle</td>
<td>Plans to go to college</td>
<td>42</td>
<td>71</td>
<td>80</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>No plans for college</td>
<td>279</td>
<td>296</td>
<td>330</td>
<td>137</td>
</tr>
<tr>
<td>Upper middle</td>
<td>Plans to go to college</td>
<td>50</td>
<td>105</td>
<td>165</td>
<td>233</td>
</tr>
<tr>
<td></td>
<td>No plans for college</td>
<td>180</td>
<td>207</td>
<td>192</td>
<td>106</td>
</tr>
<tr>
<td>High</td>
<td>Plans to go to college</td>
<td>59</td>
<td>136</td>
<td>204</td>
<td>422</td>
</tr>
<tr>
<td></td>
<td>No plans for college</td>
<td>110</td>
<td>138</td>
<td>115</td>
<td>71</td>
</tr>
</tbody>
</table>
8. APPENDIX B

One approximation to the distribution of $\hat{C}_B$ (4.2.71) is obtained by treating $\hat{\Sigma}_m$ as known (4.2.72). A superior approximation recognizes that $\hat{\Sigma}_m$ is a function of the $\hat{\mu}$'s. Let $\hat{\mu}$ be the $I-1$ dimensional probability vector

$$\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_{I-1})'$$

Then, the estimated covariance matrix for $\hat{\mu}$ under multinomial sampling is

$$\hat{\Sigma}_m = n^{-1} [\text{diag}(\hat{\mu}) - \hat{\mu}'\hat{\mu}]$$

and a small departure from $\Sigma_m$ is

$$\Delta \hat{\Sigma}_m = \hat{\Sigma}_m - \Sigma_m$$

$$= n^{-1} [\text{diag}(\hat{\mu}) - \text{diag}(\hat{\mu}) - \hat{\mu}'\hat{\mu} + \hat{\mu}'\hat{\mu}]$$

$$= n^{-1} [\text{diag}(\Delta \hat{\mu}) - \hat{\mu}'(\hat{\mu}-\hat{\mu}) - (\hat{\mu}-\hat{\mu})'\hat{\mu} + (\hat{\mu}-\hat{\mu})'(\hat{\mu}-\hat{\mu})]$$

$$= n^{-1} [\text{diag}(\Delta \hat{\mu}) - \hat{\mu}'(\Delta \hat{\mu}) - (\Delta \hat{\mu})'\hat{\mu}] + O_p(s^{-1})$$

where $\Delta \hat{\mu} = \hat{\mu}-\hat{\mu}$. Therefore, the error $\hat{C}_B$ (using 4.2.71) is

$$\hat{C}_B - C = (\hat{\Sigma}_B^{-1} \Lambda)^{-1} \hat{\Sigma}_B^{-1} (\hat{\mu}-\mu_0)$$
\begin{align*}
\text{where} \\
\mathbf{H} &= \operatorname{vech} \mathbf{V}^\wedge_{p p} \quad \text{(from 4.2.71)}
\end{align*}

and

\begin{align*}
\hat{\mathbf{R}} &= \mathbf{R} + O_p(S^{-\frac{1}{2}}).
\end{align*}

Also

\begin{align*}
\hat{\mathbf{C}}_{B - C} &= (A'^{-1} A)^{-1} A'^{-1} \mathbf{A}^\wedge \mathbf{D}^{-1} \mathbf{R} \mathbf{V}^\wedge_{p p} (A'^{-1} A)^{-1} \mathbf{A}^\wedge \mathbf{D}^{-1} \mathbf{R} + O_p(S^{-1}) \\
&= (A'^{-1} A)^{-1} A'^{-1} \mathbf{A}^\wedge \mathbf{D}^{-1} \mathbf{R} + O_p(S^{-1}),
\end{align*}

where

\begin{align*}
\mathbf{A}^\wedge &= \operatorname{vech}(\mathbf{R}((A^\wedge_{p p} - \mathbf{C}(A^\wedge_{m}))R^r}).
\end{align*}

Then, an estimator of the variance of the approximate distribution of $\hat{\mathbf{C}}_{B}$ is

\begin{align*}
\hat{\mathbf{V}}[\hat{\mathbf{C}}_{B}] &= (A'^{-1} A)^{-1} A'^{-1} \mathbf{A}^\wedge \mathbf{D}^{-1} \mathbf{A} (A'^{-1} A)^{-1},
\end{align*}

where $\hat{\mathbf{A}}_{dd}$ is an estimator of the covariance matrix of $\mathbf{A}^\wedge$. One estimator of \( \hat{\mathbf{V}}[\hat{\mathbf{A}^\wedge}] \) is
\[ \hat{\mathcal{V}}(\hat{\mathcal{A}}) = (s-1)^{-2} \sum_{t=1}^{S} (\hat{\mathcal{A}}_{t} - \overline{\mathcal{A}}) (\hat{\mathcal{A}}_{t} - \overline{\mathcal{A}})' , \]

where

\[ \hat{\mathcal{A}}_{t} = \text{vech}[\hat{R}(\hat{\mathcal{M}}_{t} - \overline{\mathcal{M}})'(\hat{\mathcal{M}}_{t} - \overline{\mathcal{M}}) + \hat{C}^{-1}_s \text{diag}(\hat{\mathcal{M}}_{t}) - \overline{\mathcal{M}}' \hat{\mathcal{M}}_{t} - \overline{\mathcal{M}}' \overline{\mathcal{M}}_{t} ] , \]

\[ \hat{\mathcal{A}} = s^{-1} \sum_{t=1}^{S} \hat{\mathcal{A}}_{t} , \]

\( \hat{\mathcal{M}}_{t} \) is the vector of proportions for the \( t \)th cluster, and \( \overline{\mathcal{M}} \) is the estimator of the first \( l \) elements of \( \mathcal{M} \). The second term in the definition of \( \hat{\mathcal{A}}_{t} \) is the derivative of \( \hat{\mathcal{M}}_{t} \) with respect to \( \mathcal{M} \), evaluated at \( \hat{\mathcal{M}} \) and multiplied by \( \hat{\mathcal{M}}_{t} \).

An alternative goodness-of-fit statistic is

\[ x^2 = s \hat{\Sigma}_{\hat{\mathcal{A}} \hat{\mathcal{A}}} , \]

where \( \hat{\Sigma}_{\hat{\mathcal{A}} \hat{\mathcal{A}}} \) is an estimator of the variance of \( \hat{\mathcal{A}} \) and

\[ \hat{\mathcal{A}}' = \text{vech}[\hat{R} \hat{\mathcal{V}} \hat{P}'] - \hat{C}_s \text{vech}[I] \]

\[ = \overline{\mathcal{M}}' - \hat{C}_s \overline{\mathcal{M}}' \]

\[ = \text{vech}[\hat{V}_{pp}'] - \hat{C}_s \overline{\mathcal{M}}' . \]
The statistics developed in this Appendix were used on the data in Example 4.1 for the first subpopulation (Herodotus). For the works of Herodotus, the diagonal elements of $\hat{\Sigma}_{dd}$ are (0.4841, 0.8241, 1.020, 0.4922, 2.1154, 2.8199, 1.1571, 2.3251, 1.4998, 0.3542). These elements are similar to the variances of the elements of $\hat{V}_{ppl}$ given in Table 4.4. The estimator of the variance of $\hat{C}_B$ according to the statistics developed here is

$$\hat{V}[\hat{C}_B] = 0.4784.$$  

This value of 0.4784 for $\hat{V}[\hat{C}_B]$ is considerably larger than the value of 0.1350 obtained under the normal model in (4.4.17). The off-diagonal elements of the covariance matrix contribute to $\hat{V}[\hat{C}_B]$. Because $\hat{\Sigma}_{dd}$ is singular, the chi-square statistic using $\hat{\Sigma}_{dd}^{-1}$ cannot be computed. If only the diagonal portion of $\hat{\Sigma}_{dd}$ is used, the value of the goodness-of-fit statistic is

$$x^2 = \hat{e}^T[\text{diag}(\hat{\Sigma}_{dd})^{-1}]\hat{e},$$

$$= 9.71,$$

where

$$\hat{e} = (-0.95, -1.14, -1.24, -0.35, 0.20, 2.10, 1.17, 0.95, 1.37, -0.20).$$