ON THE VARIANCES OF ASYMPTOTICALLY
NORMAL ESTIMATORS FROM COMPLEX SURVEYS

David A. Binder[1]

The problem of specifying and estimating the variance of
estimated parameters based on complex sample designs from
finite populations is considered. The results of this
paper are particularly useful when the paramtere estima-
tors cannot be defined explicitly as a function of other
statistics from the sample. It is shown how these results
can be applied to linear regression, logistic regression
and loglinear contingency table models.

1. INTRODUCTION

In recent years, there has been an increasing demand for using survey

data to estimate the parameters of traditional models such as regres-

sion parameters, discriminant functions, logit and probit parameters

and others. However, for many such surveys, the primary objectives

of the survey is the estimation of population or sub-population means,

totals, trends and so on. For this reason and because of operational

considerations, the survey design is often not a simple random sample,

but is more typically stratified and often multi-stage with possibly

unequal probabilities at certain stages of sampling.

Because of this, there has been much discussion (see, for example,

Sarndal;1978) on whether the sampling weights should be used in making

inferences about these model parameters. The answer seems to depend on

whether a superpopulation model is appropriate for all population units.

If this is the case, the inference on the superpopulation parameters is

often the primary concern. This leads to model-based inference, where,

for a given sample, the inferences do not depend on the sampling weights.

[1] D.A. Binder, Institutional and Agriculture Survey Methods Division,
Statistics Canada.

The question that comes to mind is: If the superpopulation   model is not appropriate, what parameters are we estimating?   It must be recognized that for many studies, particularly in the   social sciences, the model (e.g. linear regression) is only a convenient approximation of the real world and the parameters of that model (e.g. correlations and partial correlations) are often used to understand the   approximate interdependencies of the variables   rather than   having a   particular scientific interpretation.   Therefore, the parameters we are estimating do not necessarily refer to a true superpopulation   model, but are of a more descriptive nature.

In this paper, we adopt the view that we   are interested in making inferences about these "descriptive" parameters of the population.   For example, suppose $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ are $N \times p$ and $N \times 1$ matrices respectively, where each row of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ corresponds to a different individual of the population.   We are interested in the descriptive parameter, $\underset{\sim}{B}$, a $p \times 1$ vector satisfying the equations:

$$\underset{\sim}{X}^T \underset{\sim}{X} \underset{\sim}{B} = \underset{\sim}{X}^T \underset{\sim}{Y} \tag{1.1}$$

This view of descriptive parameters is the same as that taken by Frankel (1971) and Kish and Frankel (1974).

The usual estimation of such parameters normally takes into account the sampling weights. If we denote by π_i the probability that the i-th unit in the sample is sampled and let $\underset{\sim}{\Pi} = \text{diag} (\pi_1, \ldots, \pi_n)$, then the weighted parameter estimate for $\underset{\sim}{B}$ satisfies:

$$\underset{\sim}{x}^T \underset{\sim}{\Pi}^{-1} \underset{\sim}{x} \underset{\sim}{B} = \underset{\sim}{x}^T \underset{\sim}{\Pi}^{-1} \underset{\sim}{y}, \tag{1.2}$$

where $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are $n \times p$ and $n \times 1$ matrices respectively, the rows of which correspond to the sampled rows of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$.

Suppose, now, an estimator of a population parameter can be expressed as:

$$\hat{\theta} = g(z_1, \ldots, z_k), \tag{1.3}$$

where $E(z_i) = Z_i$. Here, $\hat{\theta}$ is an estimator of $g(Z_1, \ldots, Z_k)$. Following Tepping (1968) and Woodruff (1971), a Taylor series expansion for $\hat{\theta}$ yields:

$$V[\hat{\theta}] \doteq V[\sum_{i=1}^{k} (\frac{\partial g}{\partial Z_i})(z_i - Z_i)] . \qquad (1.4)$$

These formulae are exemplified for estimation of regression coefficients (1.1) by Tepping (1968). However, the expressions resulting from (1.4) for the variances of the regression coefficients are somewhat complicated compared to those derived by Fuller (1975).

In this paper we consider parameters which are not defined through an explicit equation such as (1.3), but instead are defined implicitly as $U(Z,\theta) = 0$. A simple example showing the distinction would be the ratio parameter:

$$R = \frac{\Sigma Y_k}{\Sigma X_k} ,$$

which could also be defined implicitly as:

$$\Sigma Y_k - R\Sigma X_k = 0.$$

When we deal with some models such as indirect loglinear models or logistic regression models, the parameters can be defined only through implicit relationships. The extension of Tepping's (1968) results for this case is fairly straightforward,but does not appear in its general form at present in the literature. There are, however, specific examples of its application; see, for example Fuller (1975) and Freeman and Koch (1976).

In Section 2 we give the general framework and the main results of the paper. A number of models are exemplified in Section 3.

2. GENERAL FRAMEWORK

2.1 Framework

The population units are labelled 1, ..., N. Associated with the i-th unit we have a q-dimensional data vector $\underset{\sim}{X}_i$. We have a parameter space $\Theta \subseteq R^p$. The parameter $\underset{\sim}{\theta}_o = (\theta_{1o}, \ldots, \theta_{po})$ is defined by the p equations:

$$U_i(\underset{\sim}{X}, \underset{\sim}{\theta}_o) = \sum_{k=1}^{N} u_i(\underset{\sim}{X}_k, \underset{\sim}{\theta}_o) - v_i(\underset{\sim}{\theta}_o) = 0, \tag{2.1}$$

for i=1, ..., p. We assume that equations (2.1) define $\underset{\sim}{\theta}_o$ uniquely in Θ. We also assume that $\partial u_i(\underset{\sim}{X}, \underset{\sim}{\theta})/\partial \underset{\sim}{\theta}$ and $\partial v_i(\underset{\sim}{\theta})/\partial \underset{\sim}{\theta}$ exist in a neighbourhood of $\underset{\sim}{\theta}_o$. A simple example of (2.1) is where θ_o is a population total, and we have $U(\underset{\sim}{X}, \theta_o) = \sum_{k=1}^{N} X_k - \theta_o$. Here, $u(X_k, \theta_o) = X_k$ and $v(\theta_o) = \theta_o$.

We select a sample of the units, according to some probability distribution defined on the set of all non-empty subsets of $\{1, \ldots, N\}$. We denote by $\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n$ the selected values of $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_N$. We assume that for any $\underset{\sim}{\theta} \in \Theta$, we can construct a consistent, asymptotically normal estimator of $U_i(\underset{\sim}{X}, \underset{\sim}{\theta})$. We denote this estimator by $\hat{U}_i(\underset{\sim}{x}, \underset{\sim}{\theta})$. For example, for many without replacement sampling schemes,

$$\hat{U}_i(\underset{\sim}{x}, \underset{\sim}{\theta}) = \sum_{k=1}^{n} u_i(\underset{\sim}{x}_k, \underset{\sim}{\theta})/\pi_k - v_i(\underset{\sim}{\theta}) \tag{2.2}$$

will be a consistent asymptotically normal estimator, where π_k is the probability of inclusion for the k-th unit.

We let $\sigma_{ij}(\underset{\sim}{X}, \underset{\sim}{\theta}) = \text{Cov}[\hat{U}_i(\underset{\sim}{x}, \underset{\sim}{\theta}), \hat{U}_j(\underset{\sim}{x}, \underset{\sim}{\theta})]$. For example, for estimator (2.2), we have:

$$\sigma_{ij}(\underset{\sim}{X}, \underset{\sim}{\theta}) = \sum_{k=1}^{N} \sum_{\ell=1}^{N} u_i(\underset{\sim}{X}_k, \underset{\sim}{\theta}) \, u_j(\underset{\sim}{X}_\ell, \underset{\sim}{\theta}) (\pi_{k\ell} - \pi_k \pi_\ell)/\pi_k \pi_\ell, \tag{2.3}$$

where $\pi_{k\ell}$ is the probability that the k-th and ℓ-th units in sample.

We let $\underset{\sim}{\Sigma}(\underset{\sim}{X},\underset{\sim}{\theta})$ be the p×p matrix with entries $\sigma_{ij}(\underset{\sim}{X},\underset{\sim}{\theta})$, and $\underset{\sim}{\hat{\Sigma}}(\underset{\sim}{x},\underset{\sim}{\theta})$ be a consistent estimator for $\underset{\sim}{\Sigma}$. Now, for any given $\underset{\sim}{\theta}$,

$$U_i(\underset{\sim}{X},\underset{\sim}{\theta}) + v_i(\underset{\sim}{\theta}) = \sum_{k=1}^{N} u_i(\underset{\sim}{X_k},\underset{\sim}{\theta}),$$

so that estimators $\hat{U}_i(\underset{\sim}{X},\underset{\sim}{\theta})$ and $\underset{\sim}{\hat{\Sigma}}(\underset{\sim}{x},\underset{\sim}{\theta})$ can be specified for any design in which we can derive consistent asymptotically normal estimators of population totals and consistent estimators for the variances of the estimators of the totals.

The Horvitz-Thompson estimator for (2.3) is:

$$\sum_{k=1}^{n} \sum_{\ell=1}^{n} u_i(\underset{\sim}{x_k},\underset{\sim}{\theta})\ u_j(\underset{\sim}{x_\ell},\underset{\sim}{\theta})(\pi_{k\ell} - \pi_k\pi_\ell)/\pi_k\pi_\ell\pi_{k\ell}. \tag{2.4}$$

In the case of fixed sample size, the Yates-Grundy estimator of (2.3) is:

$$\sum_{k<\ell} \left[\frac{u_i(\underset{\sim}{x_k},\underset{\sim}{\theta})}{\pi_k} - \frac{u_i(\underset{\sim}{x_\ell},\underset{\sim}{\theta})}{\pi_\ell}\right]\left[\frac{u_j(\underset{\sim}{x_k},\underset{\sim}{\theta})}{\pi_k} - \frac{u_j(\underset{\sim}{x_\ell},\underset{\sim}{\theta})}{\pi_\ell}\right] (\pi_k\pi_\ell - \pi_{k\ell}). \tag{2.5}$$

Letting $\underset{\sim}{U}(\underset{\sim}{X},\underset{\sim}{\theta})$ and $\underset{\sim}{\hat{U}}(\underset{\sim}{x},\underset{\sim}{\theta})$ be the p-dimensional vectors with components $U_i(\underset{\sim}{X},\underset{\sim}{\theta})$ and $\hat{U}_i(\underset{\sim}{x},\underset{\sim}{\theta})$ respectively, we define

$$\underset{\sim}{J}(\underset{\sim}{X},\underset{\sim}{\theta}) = \partial\underset{\sim}{U}(\underset{\sim}{X},\underset{\sim}{\theta})/\partial\underset{\sim}{\theta} \tag{2.6}$$

$$\underset{\sim}{\hat{J}}(\underset{\sim}{x},\underset{\sim}{\theta}) = \partial\underset{\sim}{\hat{U}}(\underset{\sim}{x},\underset{\sim}{\theta})/\partial\underset{\sim}{\theta}, \tag{2.7}$$

where $\underset{\sim}{J}$ and $\underset{\sim}{\hat{J}}$ are p×p partial derivative matrices. Assume that the matrices are continuous functions of $\underset{\sim}{\theta}$ and that the partial derivatives with respect to $\underset{\sim}{\theta}$ exist in a neighbourhood of $\underset{\sim}{\theta}_o$. Also assume $\underset{\sim}{\hat{J}}(\underset{\sim}{x},\underset{\sim}{\theta})$ is a consistent estimator of $\underset{\sim}{J}(\underset{\sim}{X},\underset{\sim}{\theta})$.

Our estimator for $\underset{\sim}{\theta}$ is given by $\underset{\sim}{\hat{\theta}}$, the solution to:

$$\hat{U}_i(\underset{\sim}{x},\underset{\sim}{\hat{\theta}}) = 0, \text{ for } i=1, \ldots, p. \tag{2.8}$$

We assume the sample size is sufficiently large so that the solution to (2.8) is unique in Θ. We show in the next section that the covariance matrix of $\hat{\underset{\sim}{\theta}}$ can be consistently estimated by:

$$[\hat{\underset{\sim}{J}}^{-1}(\underset{\sim}{x},\hat{\underset{\sim}{\theta}})]\ \hat{\underset{\sim}{\Sigma}}(\underset{\sim}{x},\hat{\underset{\sim}{\theta}})\ [\hat{\underset{\sim}{J}}^{-1}(\underset{\sim}{x},\hat{\underset{\sim}{\theta}})]^{T}.$$

2.2 Asymptotic Theory

Following the asymptotic arguments of Madow (1948), and Hájek (1960), we consider a sequence of populations indexed by t, with sizes $N^{(t)}$ and data $\underset{\sim}{X}^{(t)}$. We assume $N^{(t)} \to \infty$ as $t \to \infty$. For population t, we select a sample of size $n^{(t)}$ and observe data $\underset{\sim}{x}^{(t)}$. We let $\nu^{(t)} = E(n^{(t)})$ and assume

$$\lim_{t \to \infty} \nu^{(t)} = \infty$$

$$\lim_{t \to \infty} (N^{(t)} - \nu^{(t)}) = \infty$$

For any $\underset{\sim}{\theta}$ in a neighbourhood of $\underset{\sim}{\theta}_{o}^{(t)}$ we assume

$$[\nu^{(t)}]^{\frac{1}{2}}\ [\hat{\underset{\sim}{U}}(\underset{\sim}{x}^{(t)},\ \underset{\sim}{\theta}) - \underset{\sim}{U}(\underset{\sim}{X}^{(t)},\underset{\sim}{\theta})]/N^{(t)}$$

is asymptotically $N[\underset{\sim}{0},\underset{\sim}{S}(\underset{\sim}{\theta})]$, where

$$\underset{\sim}{S}(\underset{\sim}{\theta}) = \lim[\nu^{(t)}\ \underset{\sim}{\Sigma}(\underset{\sim}{X}^{(t)},\underset{\sim}{\theta})/\{N^{(t)}\}^{2}]$$

exists. We assume

$$\underset{\sim}{K}(\underset{\sim}{\theta}) = \lim \underset{\sim}{J}(\underset{\sim}{X}^{(t)},\underset{\sim}{\theta})/N^{(t)} \text{ exists and also}$$

$$\text{plim } \hat{\underset{\sim}{J}}(\underset{\sim}{x}^{(t)},\underset{\sim}{\theta})/N^{(t)} = \underset{\sim}{K}(\underset{\sim}{\theta}).$$

Also, we assume

$$\lim[\text{rank } \{\underset{\sim}{J}(\underset{\sim}{X}^{(t)},\underset{\sim}{\theta})\}] = \text{plim}[\text{rank } \{\hat{\underset{\sim}{J}}(\underset{\sim}{x}^{(t)},\underset{\sim}{\theta})\}] = p.$$

We define $\hat{\underset{\sim}{\theta}}^{(t)}$ to satisfy

$$\hat{\underset{\sim}{U}}(\underset{\sim}{x}^{(t)},\ \hat{\underset{\sim}{\theta}}^{(t)}) = 0.$$

By a Taylor series expansion, we obtain

$$\hat{\underset{\sim}{U}}(\underset{\sim}{x}^{(t)},\ \underset{\sim}{\theta}_{o}^{(t)}) \doteq - \hat{\underset{\sim}{J}}(\underset{\sim}{x}^{(t)},\ \hat{\underset{\sim}{\theta}}^{(t)})\ (\hat{\underset{\sim}{\theta}}^{(t)} - \underset{\sim}{\theta}_{o}^{(t)}). \tag{2.9}$$

Since the left hand side of (2.9) is asymptotically normal, we have that
$$(n^{(t)})^{\frac{1}{2}}\ (\hat{\underset{\sim}{\theta}}^{(t)} - \underset{\sim}{\theta}_{o}^{(t)})$$
is asymptotically $N[\underset{\sim}{0},\ \underset{\sim}{G}(\underset{\sim}{\theta}_{o})]$, where $\underset{\sim}{S}(\underset{\sim}{\theta}_{o}) = \underset{\sim}{K}(\underset{\sim}{\theta}_{o})\ \underset{\sim}{G}(\underset{\sim}{\theta}_{o})\ [\underset{\sim}{K}(\underset{\sim}{\theta}_{o})]^{T}.$

Therefore,

$$G(\theta_0) = [K^{-1}(\theta_0)] \; S(\theta_0) \; [K^{-1}(\theta_0)]^T \tag{2.10}$$

and a consistent estimator for $G(\theta_0)$ is :

$$n^{(t)} [\hat{J}^{-1}(x,\hat{\theta})] \; \hat{\Sigma}(x,\hat{\theta}) \; [\hat{J}^{-1}(x,\hat{\theta})]^T. \tag{2.11}$$

Hence, when the functional form of $\hat{U}(x,\theta)$ and $\hat{\Sigma}(x,\theta)$ is specified, we need only derive the matrix $\hat{J}(X,\theta_0)$ and its estimator $\hat{J}(x,\hat{\theta})$ to use these results.

3.   EXAMPLES

3.1   Introduction

In this section we consider in detail the implication of the general formulation given in Section 2 with respect to estimating the variances of certain population parameter estimators.   In particular, we discuss ratios, regression coefficients and log linear models for categorical data.   Other models, such as probit models could be analyzed analogously.

In general, we use the following notation.   If W_1, ..., W_N are population values, with $W = \Sigma W_k$, then on selecting a sample w_1, ..., w_n, we have an unbiased estimator of W given by $\hat{W}$.   We let $V(\hat{W})$ represent the covariance matrix for $\hat{W}$ and $\hat{V}(\hat{W})$ a consistent estimator of $V(\hat{W})$.   The particular form of this estimator will depend on the sample design; for example, multi-stage stratified, pps with replacement, etc. .

3.2   Ratios

Suppose we are interested in $R = \Sigma X_{k2}/\Sigma X_{k1}$.   We define

$$U(X,R) = \Sigma X_{k2} - R \Sigma X_{k1}.$$

Therefore, for without replacement sampling, we have :

$$\hat{U}(x,R) = \hat{X}_2 - R\hat{X}_1.$$

Setting $\hat{U}(\underset{\sim}{x},\hat{R}) = 0$, we obtain

$$\hat{R} = \hat{X}_2/\hat{X}_1 . \tag{3.1}$$

We define $W_k = X_{k2} - R \, X_{k1}$.

Since, $J(\underset{\sim}{X},R) = -\Sigma X_{k1}$, we have that $V(\hat{R})$ is approximately $V(\hat{W})/(\Sigma X_{k1})^2$. This is estimated by $\hat{V}(\hat{W})/\hat{X}_1^2$. In the case of stratified sampling, this yields the same result as in Woodruff (1971).

3.3 Regression Coefficients and R

Suppose our data matrix $\underset{\sim}{X}$ is partitioned into $[\underset{\sim}{Z}|\underset{\sim}{Y}]$, the first column of $\underset{\sim}{Z}$ being the vector of 1's. The vector $\underset{\sim}{Y}$ is N×1. We have parameters of interest θ, $\underset{\sim}{B}$, and R^2 defined by:

$$U_1 = \theta - \underset{\sim}{Y}^T \underset{\sim}{1} = 0, \tag{3.2a}$$

$$\underset{\sim}{U}_2 = \underset{\sim}{Z}^T \underset{\sim}{Z} \underset{\sim}{B} - \underset{\sim}{Z}^T \underset{\sim}{Y} = 0, \tag{3.2b}$$

$$U_3 = (\underset{\sim}{Y}^T\underset{\sim}{Y}-N^{-1}\theta^2)(R^2-1) + \underset{\sim}{Y}^T \underset{\sim}{Y} - \underset{\sim}{Y}^T \underset{\sim}{Z} \underset{\sim}{B} = 0. \tag{3.2c}$$

Here, $\underset{\sim}{B}$ denotes the vector of regression coefficients, R^2 is the coefficient of multiple determination and θ is the total of the Y's . We first consider the case where N is known. We let $SSY = \underset{\sim}{Y}^T\underset{\sim}{Y} - N^{-1}\theta^2$. We also define $\underset{\sim}{S}_{ZZ}$ as the estimator for $\underset{\sim}{Z}^T\underset{\sim}{Z}$, S_{YY} the estimator for $\underset{\sim}{Y}^T\underset{\sim}{Y}$ and $\underset{\sim}{S}_{ZY}$ the estimator for $\underset{\sim}{Z}^T\underset{\sim}{Y}$. We therefore have :

$$\hat{\theta} = \hat{Y}, \tag{3.3a}$$

$$\hat{\underset{\sim}{B}} = \underset{\sim}{S}_{ZZ}^{-1} \underset{\sim}{S}_{ZY}, \tag{3.3b}$$

$$\hat{R}^2 = 1 - \frac{S_{YY} - \hat{\underset{\sim}{B}}^T \underset{\sim}{S}_{ZY}}{S_{YY} - N^{-1}\hat{Y}^2}, \tag{3.3c}$$

and

$$\underset{\sim}{J} = \partial \underset{\sim}{U}(\underset{\sim}{Z},\underset{\sim}{Y},\underset{\sim}{B},R^2,\theta)/\partial(\underset{\sim}{B},\hat{R},\theta) = \begin{bmatrix} \underset{\sim}{0}^T & 0 & 1 \\ \underset{\sim}{Z}^T\underset{\sim}{Z} & \underset{\sim}{0} & \underset{\sim}{0} \\ -\underset{\sim}{Y}^T\underset{\sim}{Z} & SSY & 2\bar{Y}(1-R^2) \end{bmatrix},$$

where $\bar{Y} = \theta/N$.

Therefore,

$$
\underset{\sim}{J}^{-1} = \begin{bmatrix} \underset{\sim}{0} & (\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1} & \underset{\sim}{0} \\ -2\bar{Y}(1-R^2)/SSY & \underset{\sim}{B}^T/SSY & 1/SSY \\ 1 & \underset{\sim}{0}^T & 0 \end{bmatrix} .
$$

Now, letting $\underset{\sim}{W}_k^T(B) = (Z_{k1} e_k, \ldots, Z_{kp} e_k)$, where $e_k = Y_k - \underset{j}{\Sigma} Z_{kj} B_j$, we obtain :

$$
\underset{\sim}{V}[\underset{\sim}{\hat{B}}] \doteq (\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1} \underset{\sim}{V}[\underset{\sim}{\hat{W}}(B)](\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1} . \tag{3.4}
$$

This is a direct consequence of (2.10). Note that the set of $\underset{\sim}{W}_k(B)$ vectors corresponds to $\underset{\sim}{U}_2$ in (3.2b). Fuller (1975) obtains the same result for stratified or two-stage stratified sampling.

To estimate (3.4) we use :

$$
\underset{\sim}{\hat{V}}[\underset{\sim}{\hat{B}}] = \underset{\sim}{S}_{ZZ}^{-1} \underset{\sim}{\hat{V}}[\underset{\sim}{\hat{W}}(\underset{\sim}{\hat{B}})] \underset{\sim}{S}_{ZZ}^{-1} .
$$

We can also estimate the variance of $\hat{R}^2$. If $\underset{\sim}{W}_k^T(B,R^2) = [Y_k, Z_{k1} e_k, \ldots, Z_{kp} e_k, Y_k(\underset{j}{\Sigma} Z_{kj} B_j - R^2 Y_k)]$ and $\underset{\sim}{c}^T = [-2\hat{Y}(1-\hat{R}^2)/N, \underset{\sim}{\hat{B}}^T, 1]/(S_{YY}-N^{-1} \hat{Y}^2)$, we obtain:

$$
\hat{V}[\hat{R}^2] \doteq \underset{\sim}{c}^T \underset{\sim}{\hat{V}}[\underset{\sim}{\hat{W}}(\underset{\sim}{\hat{B}},\hat{R}^2)] \underset{\sim}{c} . \tag{3.5}
$$

For the case where N is unknown (e.g. the primary sampling units are geographic areas), we have the additional equation:

$$
U_4 = N - \Sigma 1 . \tag{3.6}
$$

Adding the appropriate row and column to $\underset{\sim}{J}$ and inverting, we obtain the following results for estimating $V[\hat{R}^2]$.
We let

$$
\underset{\sim}{W}_k^T(B,R^2) = [Y_k, Z_{k1} e_k, \ldots, Z_{kp} e_k, Y_k(\underset{j}{\Sigma} Z_{kj} B_j - R^2 Y_k), 1]
$$

and

$$
\underset{\sim}{c}^T = [-2\hat{Y}(1-\hat{R}^2)/\hat{N}, \underset{\sim}{\hat{B}}^T, 1, \hat{Y}^2(1-\hat{R}^2)/\hat{N}^2]/(S_{YY} - \hat{N}^{-1} \hat{Y}^2) .
$$

We then have $\hat{V}[\hat{R}^2]$ is given by (3.5) for these new values of $\underset{\sim}{W}_k(B, R^2)$ and $\underset{\sim}{c}$.

3.4 Logistic Regression

As in the previous section, we assume the data matrix $\underset{\sim}{X}$ can be partitioned into $[\underset{\sim}{Z}|\underset{\sim}{Y}]$, but now $\underset{\sim}{Y}$ is a vector of 0's and 1's. In the traditional statistical framework, the logistic regression model for $\underset{\sim}{Y}$ conditional on $\underset{\sim}{Z}$ asserts that $Y_1, \ldots, Y_N$ are independent with $\Pr(Y_k = 1) = p_k(\underset{\sim}{\beta})$, where :

$$p_k(\underset{\sim}{\beta}) = \frac{\exp(\underset{\sim}{\beta}^T \underset{\sim}{z}_k)}{1 + \exp(\underset{\sim}{\beta}^T \underset{\sim}{z}_k)} \quad . \tag{3.7}$$

Letting $\underset{\sim}{B}$ be the maximum likelihood estimator for $\underset{\sim}{\beta}$, we have that $\underset{\sim}{B}$ satisfies

$$\underset{\sim}{U} = \underset{\sim}{Z}^T \underset{\sim}{P}(B) - \underset{\sim}{Z}^T \underset{\sim}{Y} = 0 , \tag{3.8}$$

where $\underset{\sim}{P}(\underset{\sim}{B})^T = [p_1(\underset{\sim}{B}), \ldots, p_N(\underset{\sim}{B})]$.
For a given finite population, we define $\underset{\sim}{B}$ as our parameter of interest.

We let $\underset{\sim}{C}(B)$ be our estimate for $\underset{\sim}{Z}^T \underset{\sim}{P}(B)$ and $\underset{\sim}{S}_{ZY}$ our estimate for $\underset{\sim}{Z}^T \underset{\sim}{Y}$. Therefore, $\hat{\underset{\sim}{B}}$ satisfies $\underset{\sim}{C}(\hat{\underset{\sim}{B}}) = \underset{\sim}{S}_{ZY}$. These equations must be solved iteratively in general. We also have

$$\underset{\sim}{J} = \frac{\partial \underset{\sim}{U}}{\partial \underset{\sim}{B}} \quad .$$

The (i,j)th component of $\underset{\sim}{J}$ is $\sum_k z_{ki} z_{kj} p_k(\underset{\sim}{B}) [1 - p_k(\underset{\sim}{B})]$. We denote the estimator of $\underset{\sim}{J}$ by $\hat{\underset{\sim}{J}}$.

To estimate the variance of $\hat{\underset{\sim}{B}}$, we let

$$\underset{\sim}{w}_k^T = (z_{k1} \hat{e}_k, \ldots, z_{kr} \hat{e}_k)$$

where $\hat{e}_k = p_k(\hat{\underset{\sim}{B}}) - Y_k$. The estimator for $\underset{\sim}{V}[\hat{\underset{\sim}{B}}]$ is given by :

$$\hat{\underset{\sim}{J}}^{-1} \hat{V}(\hat{\underset{\sim}{w}}) \hat{\underset{\sim}{J}}^{-1} \quad .$$

3.5 Loglinear Models for Categorical Data

Suppose that each member of the population belongs to exactly one of q distinct categories. Associated with category i we have an $r \times 1$ vector $\underset{\sim}{a}_i$ such that the proportion of individuals in the i-th category is approximately

$$p_i(\underset{\sim}{\beta}) = \frac{\exp(\underset{\sim}{a}_i^T \underset{\sim}{\beta})}{\sum_j \exp(\underset{\sim}{a}_j^T \underset{\sim}{\beta})} \ .$$

We let $\underset{\sim}{p}(\underset{\sim}{\beta})^T = [p_1(\underset{\sim}{\beta}), \ldots, p_q(\underset{\sim}{\beta})]$ and $\underset{\sim}{N}^T = (N_1, \ldots, N_q)$, where N_i is the number of individuals in the i-th category. Now, if the population were generated from a multinomial distribution with probabilities $\underset{\sim}{p}(\underset{\sim}{\beta})$, the maximum likelihood estimator for $\underset{\sim}{\beta}$, given by $\underset{\sim}{B}$, satisfies:

$$\underset{\sim}{U} = \underset{\sim}{A}^T \underset{\sim}{N} - [\underset{\sim}{A}^T \underset{\sim}{p}(\underset{\sim}{B})] \underset{\sim}{1}^T \underset{\sim}{N} = 0,$$

where $\underset{\sim}{A}$ is a $q \times r$ matrix with i-th row being $\underset{\sim}{a}_i^T$. We consider $\underset{\sim}{B}$ as our parameter of interest for any given finite population.

We let $\hat{\underset{\sim}{N}}$ be a consistent asymptotically normal estimator of $\underset{\sim}{N}$, with variance-covaraince matrix $V[\hat{\underset{\sim}{N}}]$ and estimated matrix $\hat{V}[\hat{\underset{\sim}{N}}]$. Our estimator, $\hat{\underset{\sim}{B}}$, satisfies:

$$\underset{\sim}{A}^T \hat{\underset{\sim}{N}} - [\underset{\sim}{A}^T \underset{\sim}{p}(\hat{\underset{\sim}{B}})] \underset{\sim}{1}^T \hat{\underset{\sim}{N}} = 0. \tag{3.9}$$

This estimator was suggested by Freeman and Koch (1976). It may be less efficient than Imrey, Koch and Stokes (1981, 1982) functional asymptotic regression methodology; however, we need not calculate all the components of $\hat{V}[\hat{\underset{\sim}{N}}]$ to apply (3.9).

Let $\underset{\sim}{D}(\underset{\sim}{B})$ be $\text{diag}[\underset{\sim}{p}(\underset{\sim}{B})]$ and $\underset{\sim}{H}(\underset{\sim}{B}) = \underset{\sim}{D}(\underset{\sim}{B}) - \underset{\sim}{p}(\underset{\sim}{B}) \underset{\sim}{p}(\underset{\sim}{B})^T$. We have:

$$\underset{\sim}{J} = \frac{\partial \underset{\sim}{U}}{\partial \underset{\sim}{B}} = - (\underset{\sim}{1}^T \underset{\sim}{N}) \underset{\sim}{A}^T \underset{\sim}{H}(\underset{\sim}{B}) \underset{\sim}{A} \ .$$

Therefore the asymptotic variance matrix for $\hat{\underset{\sim}{B}}$ is given by:

$$\underset{\sim}{V}[\hat{\underset{\sim}{B}}] = (\underset{\sim}{N}^T\underset{\sim}{1})^{-2} \; (\underset{\sim}{A}^T \; \underset{\sim}{H}(B) \; \underset{\sim}{A})^{-1}$$

$$\underset{\sim}{A}^T(\underset{\sim}{I}-\underset{\sim}{p}(B)\underset{\sim}{1}^T) \; V[\hat{\underset{\sim}{N}}] \; (\underset{\sim}{I}-\underset{\sim}{1} \; \underset{\sim}{p}(B)^T) \; \underset{\sim}{A}(\underset{\sim}{A}^T \; \underset{\sim}{H}(B) \; \underset{\sim}{A})^{-1} . \quad (3.10)$$

This expression can sometimes be simplified as follows. If it can be assumed that $\underset{\sim}{N}/\underset{\sim}{N}^T\underset{\sim}{1} \doteq \underset{\sim}{p}(B)$, then for $\hat{\underset{\sim}{\pi}} = \hat{\underset{\sim}{N}}/\hat{\underset{\sim}{N}}^T\underset{\sim}{1}$ we have:

$$\underset{\sim}{V}[\hat{\underset{\sim}{\pi}}] \doteq (\underset{\sim}{N}^T\underset{\sim}{1})^{-2} \; (\underset{\sim}{I}-\underset{\sim}{p}(B)\underset{\sim}{1}^T) \; \underset{\sim}{V}[\hat{\underset{\sim}{N}}](\underset{\sim}{I}-\underset{\sim}{1} \; \underset{\sim}{p}(B)^T),$$

so that

$$\underset{\sim}{V}[B] \doteq (\underset{\sim}{A}^T \; \underset{\sim}{H}(B) \; \underset{\sim}{A})^{-1} \; \underset{\sim}{A}^T \; \underset{\sim}{V}[\hat{\underset{\sim}{\pi}}] \; \underset{\sim}{A}(\underset{\sim}{A}^T \; \underset{\sim}{H}(B) \; \underset{\sim}{A})^{-1} . \quad (3.11)$$

We also have that the covariance matrix for $\underset{\sim}{p}(\hat{\underset{\sim}{B}})$, the estimated cell probabilities, is given by:

$$\underset{\sim}{V}[\underset{\sim}{p}(\hat{\underset{\sim}{B}})] = \underset{\sim}{H}(B) \; \underset{\sim}{A} \; \underset{\sim}{V}[\hat{\underset{\sim}{B}}] \; \underset{\sim}{A}^T \; \underset{\sim}{H}(B).$$

The estimators of $\underset{\sim}{V}[\hat{\underset{\sim}{B}}]$ and $\underset{\sim}{V}[\underset{\sim}{p}(B)]$ are similar expressions, where $\underset{\sim}{N}$ and $\underset{\sim}{B}$ are replaced by $\hat{\underset{\sim}{N}}$ and $\hat{\underset{\sim}{B}}$ respectively. These assume that $\hat{\underset{\sim}{V}}[\hat{\underset{\sim}{N}}]$ is readily available. For some problems where q is relatively large compared to r, it would be more efficient to proceed as follows. Let

$$Y_{ki} = 1 \quad \text{if k-th unit in i-th category}$$
$$= 0 \quad \text{otherwise},$$

for k=1, ..., N; i=1, ..., q. Let $\underset{\sim}{Y}_k^T = (Y_{k1}, \; ..., \; Y_{kq})$, and

$$\underset{\sim}{W}_k = \underset{\sim}{A}^T[\underset{\sim}{I}-\underset{\sim}{p}(\hat{\underset{\sim}{B}}) \; \underset{\sim}{1}^T] \; \underset{\sim}{Y}_k .$$

We then obtain:

$$\hat{\underset{\sim}{V}}[\hat{\underset{\sim}{B}}] = \quad (\hat{\underset{\sim}{N}}^T\underset{\sim}{1})^2 \; (\underset{\sim}{A}^T \; \underset{\sim}{H}(\hat{\underset{\sim}{B}}) \; \underset{\sim}{A})^{-1} \; \hat{\underset{\sim}{V}}(\hat{\underset{\sim}{W}}) \; (\underset{\sim}{A}^T \; \underset{\sim}{H}(B) \; \underset{\sim}{A})^{-1}.$$

We remark that the methodology described in this section can be readily extended to product-multinomial type models, where we have a log-linear model for $\{N_{ij}\}$, but the margins $\{\sum_j N_{ij}\}$ are known.

4. DISCUSSION

The techniques described in the paper have been described for some specific models; see, for example, Fuller (1975) and Freeman and Koch (1976). However, the general results are not explicitly described. Many standard statistical packages may be used for the estimation of the parameters of the models described, but the variances and tests of hypotheses given in these packages will not be valid.

The results of this paper depend on the assumption of asymptotic normality of the estimators. Empirical studies on the validity of these approximations are important.

An alternative methodology to estimating many of the parameters described here is given by Imrey, Koch and Stokes (1981, 1982). Their functional asymptotic regression methodology also falls within the general framework described here, with respect to variance derivation and estimation.

REFERENCES

[1]   Frankel, M.R. (1971), Inference from Survey Samples. University
       of Michigan, Ann Arbor.

[2]   Freeman, D.H. Jr., and Koch, G.G. (1976), "An Asymptotic Covariance
       Structure for Testing Hypotheses on Raked Contingency Tables from
       Complex Sample Surveys" , Proc. Amer. Statist. Ass. (Social
       Statistics Section), Part 1, 330-335.

[3]   Fuller, W.A. (1975), "Regression Analysis for Sample Survey", Sankya,
       Series C, 37, 117-132.

[4]   Hajek, J. (1960), "Limiting Distributions in Simple Random Sampling
       from a Finite Population", Publ. Math. Inst. Hung. Acad. Sci., 5,
       361-374.

[5]   Imrey, P.B., Koch, G.G., Stokes, M.E. (1981, 1982), "Categorical
       Data Analysis: Some Reflections on the Log Linear Model and
       Logistic Regression; Part I: Historical and Methodological
       Overview. Part II: Data analysis", International Statistical
       Review. To appear.

[6]   Kish, L., and Frankel, M.R. (1974), "Inference from Complex
       Samples", J. Roy. Statistic. Soc. B, 36, 1-22.

[7]   Madow, W.G. (1948), "On the Limiting Distribution of Estimates
       Based on Samples from Finite Universes", Ann. Math. Statist., 19,
       535-545.

[8]   Särndal, C.E. (1978), "Design-Based and Model-Based Inference in
       Survey Sampling", Scand. J. Statist., 5, 27-52.

[9]   Tepping, B.J. (1968), "The Estimation of Variance in Complex Surveys",
       Proc. Amer. Statist. Assoc. (Social Statistics Section), 11-18.

[10]   Woodruff, R.S. (1971), "A Simple Method for Approximating the
       Variance of a Complicated Estimate", J. Amer. Statist. Assoc.
       66, 411-414.