International Statistical Review, **51** (1983), pp. 279–292. Longman Group Limited/Printed in Great Britain

On the Variances of Asymptotically Normal Estimators from Complex Surveys

David A. Binder

Institutional & Agriculture Survey Methods Division, Statistics Canada, R.H. Coats Building, 11th Floor, Tunney's Pasture, Ottawa, Ontario, Canada K1A OT6

Summary

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 parameter estimators 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 log linear contingency table models. An example of the application of the technique to the Canada Health Survey is given.

Key words: Generalized linear models; Logistic regression; Log linear models; Regression.

1 Introduction

Because of the multivariate nature of many sample surveys and the increase in the use of multivariate models by researchers and analysts, sample survey data are frequently being used to estimate the parameters of multivariate models. Some examples of analyses which are carried out on sample survey data are regression analysis, discriminant analysis, logit and probit analysis and log linear analysis of contingency table data.

However, many of these surveys were designed primarily to estimate means, totals, proportions or ratios for certain populations. Also, because of operational and efficiency considerations, the survey design is usually stratified and often multistage with unequal probabilities of selection at certain stages.

There has been some discussion on whether the sampling weights should be used in making inferences about these model parameters; for example, Särndal (1978). The answer seems to depend on whether a superpopulation model applies to all population units. If so, model-based inference is appropriate because of certain optimality properties under the assumed model. This, of course, does not preclude the possibility that some of the design variables, e.g. stratum identification, measure of size, may be included in the model. The question that comes to mind is: if the superpopulation model is not appropriate, what parameters are we estimating? For example, what is the meaning of a regression coefficient if the linear model is not valid?

It must be recognized that for many studies, particularly in the social sciences, the model is only a convenient approximation to the real world and the parameters of that model are often used to understand the approximate interdependencies, rather than having a scientific interpretation.

In this paper we consider the problem of finding the asymptotic design-based sampling distribution for parameters which are defined as functions of the data values in the finite population. For example, suppose $\mathbf{X}$ and $\mathbf{Y}$ are $N \times p$ and $N \times 1$ matrices respectively, where each row of $\mathbf{X}$ and $\mathbf{Y}$ corresponds to a different individual of the population. In the linear regression context, we are interested in the sampling distribution of an estimator for the parameter $\mathbf{B}$ defined by $\mathbf{X}^T\mathbf{X}\mathbf{B} = \mathbf{X}^T\mathbf{Y}$, rather than the superpopulation parameter $\boldsymbol{\beta}$ arising from a linear model. This view is the same as that taken by Frankel (1971) and Kish & Frankel (1974).

In the next section we look at some examples of parameters which may be of interest in sample surveys. We pay particular attention to the generalized linear model (Nelder & Wedderburn, 1972). These parameters are then put into a more general class of estimation problems. An heuristic solution is given in § 3 for the asymptotic distribution of this general class of parameters. The Appendix gives a more formal proof of the main results. In § 4 we show some implications of this general solution to parameters from regression, logistic regression and log linear models of categorical data. A discriminant analysis and logistic regression analysis is discussed in the context of the Canada Health Survey in § 5.

Although this paper emphasizes the results for sampling from finite populations, many of the results can be applied to infinite populations as well, particularly when estimates of the sampling variance, without resorting to the structure of a model, are desired.

2 Parameters of interest

2.1 *Implicit and explicit parameters*

When sampling from finite populations, the usual parameters which are estimated are population or subpopulation totals, means, proportions or ratios. For a population consisting of N units, we suppose that associated with the kth unit is a q-dimensional data vector $\mathbf{Z}_k = (Z_{1k}, \ldots, Z_{qk})^T$. Now, often the population parameters of interest can be expressed explicitly as functions of population totals for variables associated with $\mathbf{Z}$. An example would be ratios of totals within certain subpopulations. To determine the asymptotic variances for these explicitly defined parameters, see Tepping (1968) and Woodruff (1971).

However, for many analyses of multivariate data, it is more convenient to define the parameters of interest as implicit functions of population totals. In § 1 we mentioned the regression parameter $\mathbf{B}$, defined by

$$\mathbf{X}^T\mathbf{X}\mathbf{B} = \mathbf{X}^T\mathbf{Y}, \tag{2.1}$$

where $\mathbf{X}$ is an $N \times p$ matrix and $\mathbf{Y}$ is an $N \times 1$ vector. It is true that $\mathbf{B}$ could be defined explicitly by $\mathbf{B} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$. We assume that $\mathbf{X}$ is of full rank. However, in this paper we will show that the implicit definition in (2.1) gives more tractable results for the variances of the estimated regression coefficients.

2.2 *Generalized linear models*

The generalized linear models, described by Nelder & Wedderburn (1972), provide examples of population parameters which are defined implicitly. Here, the parameter definition is motivated by parametric models for infinite populations. The finite population parameters of interest will be the maximum likelihood estimates for the infinite population based on all the data values in our finite population. For example, in normal regression analysis, the finite population parameter of interest is $\mathbf{B}$ defined by (2.1).

The generalized linear model is defined as follows. Given parameters θ and ϕ, the density function for observation y from an infinite population is given by

$$p(y;\theta,\phi)=\exp[\alpha(\phi)\{y\theta-g(\theta)+h(y)\}+\gamma(\phi,y)], \tag{2.2}$$

where $\alpha(\phi)>0$. Note that $E(Y)=g'(\theta)=\mu(\theta)$ and $V(Y)=\mu'(\theta)/\alpha(\phi)$. Now for a given vector $\mathbf{x}=(x_1,\ldots,x_p)$, we assume that $\theta=f(\sum x_i\beta_i)$, where $f(.)$ is a known differentiable function and $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_p)^{\mathrm{T}}$ is unknown. For observations $(y_k,\mathbf{x}_k^{\mathrm{T}})=(y_k,x_{1k},\ldots,x_{pk})$, where $k=1,\ldots,N$, the maximum likelihood solution for $\boldsymbol{\beta}$ is the solution of

$$\sum_{k=1}^{N}[y_k-\mu\{f(\mathbf{x}_k^{\mathrm{T}}\boldsymbol{\beta})\}]f'(\mathbf{x}_k^{\mathrm{T}}\boldsymbol{\beta})x_{ik}=0 \quad (i=1,\ldots,p). \tag{2.3}$$

If $\mu(.)$ and $f'(.)$ are strictly monotone functions then the solution to $\boldsymbol{\beta}$ is unique, whenever the $N\times p$ matrix of x values is of full rank.

Examples of models which fit into this structure are normal linear regression, probit and logit analysis, log linear analysis of contingency tables, and estimation of variance components. In the finite population context, we assume that for the kth individual we observe $(Y_k,X_{1k},\ldots,X_{pk})$. We define the population parameter $\mathbf{B}$ as the solution of

$$\sum_{k=1}^{N}[Y_k-\mu\{f(\mathbf{X}_k^{\mathrm{T}}\mathbf{B})\}]f'(\mathbf{X}_k^{\mathrm{T}}\mathbf{B})X_{ik}=0 \quad (i=1,\ldots,p). \tag{2.4}$$

Again, we emphasize that although the parameter $\mathbf{B}$ has been motivated by infinite population concepts, we are only concerned with the design-based sampling distribution of $\hat{\mathbf{B}}$, defined in terms of the finite population values.

2.3 General parameter definition

Expression (2.4) could be more generally written as:

$$\sum_{k=1}^{N}\mathbf{u}(Y_k,\mathbf{X}_k;\mathbf{B})=0, \tag{2.5}$$

where $\mathbf{u}$ is a p-dimensional vector-valued function of the data $(Y_k,\mathbf{X}_k)$ and the parameter $\mathbf{B}$.

For the remainder of this paper, we will consider the finite population parameter $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_p)$ defined by an expression of the form

$$\mathbf{W}_N(\boldsymbol{\theta})=\sum_{k=1}^{N}\mathbf{u}(\mathbf{Z}_k;\boldsymbol{\theta})-\mathbf{v}(\boldsymbol{\theta})=0, \tag{2.6}$$

where $\mathbf{Z}_k=(Z_{1k},\ldots,Z_{qk})$, the data values for the kth unit, and $\mathbf{Z}$ is the $N\times q$ matrix of the data values for all units in the finite population. We include in expression (2.6) the term $\mathbf{v}(\boldsymbol{\theta})$ to allow for explicitly defined parameters. The distinction is necessary because for a given $\boldsymbol{\theta}$ we know the value of $\mathbf{v}(\boldsymbol{\theta})$, but $\mathbf{u}(\mathbf{Z}_k;\boldsymbol{\theta})$ is only known for those units in the sample. This has implications on the estimation of $\boldsymbol{\theta}$ from a sample, as discussed in § 3.

3 Parameter estimation and variance estimation

3.1 Description

In this section we give our estimator for $\boldsymbol{\theta}$ defined by (2.6). For this estimator, we derive its asymptotic variance based on Taylor expansions. We also find an estimator for the asymptotic variance. The theoretical justification for the validity of the approach is given in the Appendix, with the required assumptions. The main assumptions are:

(a) a parameter space which contains a neighbourhood of the parameter of interest;
(b) a sequence of sample designs and populations which admits asymptotically normal estimators for certain population totals and consistent estimators for the variance of the estimate of the totals;
(c) some continuity and limiting conditions on $\mathbf{W}_N(\theta)$ and its partial derivatives;
(d) a continuity condition on the variance of the estimated total.

3.2 *Estimation of* $\boldsymbol{\theta}$

The parameter $\boldsymbol{\theta}$ is defined in (2.6). For any given value of $\boldsymbol{\theta}$, the term $\sum \mathbf{u}(\mathbf{Z}_k;\boldsymbol{\theta})$, where the sum is over $k=1,\dots,N$, is the population total of functions of the data values $\mathbf{Z}_1,\dots,\mathbf{Z}_k$. We represent the estimator for these totals by $\hat{\mathbf{U}}(\boldsymbol{\theta})$. This is simply the estimator of a total based on data values $\mathbf{u}(\mathbf{Z}_1;\boldsymbol{\theta}),\dots,\mathbf{u}(\mathbf{Z}_n;\boldsymbol{\theta})$. We assume that $\hat{\mathbf{U}}(\boldsymbol{\theta})$ is asymptotically normal with mean $\mathbf{U}(\boldsymbol{\theta})=\sum \mathbf{u}(\mathbf{Z}_k;\boldsymbol{\theta})$, where the sum is over $k=1,\dots,N$, and variance $\boldsymbol{\Sigma}_U(\boldsymbol{\theta})$. As well, a consistent estimator is assumed to exist for $\boldsymbol{\Sigma}_U(\boldsymbol{\theta})$, denoted by $\hat{\boldsymbol{\Sigma}}_U(\boldsymbol{\theta})$. We let

$$\hat{\mathbf{W}}(\boldsymbol{\theta})=\hat{\mathbf{U}}(\boldsymbol{\theta})-\mathbf{v}(\boldsymbol{\theta}). \tag{3.1}$$

Now, $\hat{\boldsymbol{\theta}}$ is defined by the equations:

$$\hat{\mathbf{W}}(\hat{\boldsymbol{\theta}})=\mathbf{0}. \tag{3.2}$$

For example, in the linear regression model, we define the parameter $\hat{\mathbf{B}}$ by $\mathbf{S}_{XX}\hat{\mathbf{B}}-\mathbf{S}_{XY}=0$, where $\mathbf{S}_{XX}$ and $\mathbf{S}_{XY}$ are estimators for $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ and $\mathbf{X}^{\mathrm{T}}\mathbf{Y}$ given in (2.1). This is equivalent to (3.1) and (3.2), where

$$\mathbf{u}(y_k,\mathbf{x}_k;\mathbf{B})=-(y_k-\mathbf{x}_k^{\mathrm{T}}\mathbf{B})\mathbf{x}_k, \quad \mathbf{v}(\mathbf{B})=0.$$

Other parameter estimators based on the generalized linear models of § 2.2 will be given in § 4.

3.3 *Variance of* $\hat{\theta}$

To obtain the variance of $\hat{\boldsymbol{\theta}}$, we take a Taylor expansion of $\hat{\mathbf{W}}(\hat{\boldsymbol{\theta}})$ at $\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ is the population parameter value. We obtain

$$\mathbf{0}=\hat{\mathbf{W}}(\hat{\boldsymbol{\theta}})\simeq\hat{\mathbf{W}}(\boldsymbol{\theta}_0)+\frac{\partial\hat{\mathbf{W}}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_0}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0),$$

so that

$$\hat{\mathbf{W}}(\boldsymbol{\theta}_0)\simeq-\frac{\partial\hat{\mathbf{W}}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_0}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0).$$

Taking variances of both sides, we obtain in the limit

$$\boldsymbol{\Sigma}_U(\boldsymbol{\theta}_0)=\left[\frac{\partial\mathbf{W}_N(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_0}\right]\mathbf{V}(\hat{\boldsymbol{\theta}})\left[\frac{\partial\mathbf{W}_N(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_0}\right]^{\mathrm{T}},$$

or equivalently

$$\mathbf{V}(\hat{\boldsymbol{\theta}})=\left[\frac{\partial\mathbf{W}_N(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_0}\right]^{-1}\boldsymbol{\Sigma}_U(\boldsymbol{\theta}_0)\left[\frac{\partial W_N(\boldsymbol{\theta}_0)^{\mathrm{T}}}{\partial\boldsymbol{\theta}_0}\right]^{-1}, \tag{3.3}$$

providing $\partial \mathbf{W}_N/\partial \boldsymbol{\theta}_0$ is of full rank. An estimator of $\mathbf{V}(\hat{\boldsymbol{\theta}})$ is

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\theta}})=\left[\frac{\partial \hat{\mathbf{W}}(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right]^{-1}\hat{\boldsymbol{\Sigma}}_U(\hat{\boldsymbol{\theta}})\left[\frac{\partial \hat{\mathbf{W}}(\hat{\boldsymbol{\theta}})^{\mathrm{T}}}{\partial \hat{\boldsymbol{\theta}}}\right]^{-1}. \tag{3.4}$$

A formal justification of (3.3) and (3.4) is given in the Appendix. In § 4 we show how the results apply to some generalized linear models.

4 Estimation for linear models

4.1 Generalized linear models

We now consider the variance estimation for the parameters of the generalized linear models described in § 2.2. We concentrate on the important special case where $f(Y)=Y$. The parameter of interest $\mathbf{B}$ is defined by

$$\sum_{k=1}^{N}[Y_k-\mu(\mathbf{X}_k^{\mathrm{T}}\mathbf{B})]X_{ik}=0 \quad (i=1,\ldots,p). \tag{4.1}$$

Therefore

$$\mathbf{W}_N(\mathbf{B})=\sum_{k=1}^{N}[Y_k-\mu(\mathbf{X}_k^{\mathrm{T}}\mathbf{B})]\mathbf{X}_k, \tag{4.2}$$

where $\mathbf{X}_k$ is the $p\times 1$ vector of X values for the kth observation. If $\hat{\mathbf{H}}$ is an estimator of $\sum_k Y_k\mathbf{X}_k$ and $\hat{\mathbf{Q}}(\mathbf{B})$ is an estimator of $\sum_k \mu(\mathbf{X}_k^{\mathrm{T}}\mathbf{B})\mathbf{X}_k$, then the estimator for $\mathbf{B}$ is given by the solution to

$$\hat{\mathbf{Q}}(\hat{\mathbf{B}})=\hat{\mathbf{H}}. \tag{4.3}$$

Now,

$$\frac{\partial \mathbf{W}_N(B)}{\partial \mathbf{B}}=-\sum_{k=1}^{N}\mu'(\mathbf{X}_k^{\mathrm{T}}\mathbf{B})\mathbf{X}_k\mathbf{X}_k^{\mathrm{T}}=\mathbf{X}^{\mathrm{T}}\boldsymbol{\Lambda}(\mathbf{B})\mathbf{X},$$

where $\mathbf{X}$ is the $N\times p$ matrix of X values and

$$\boldsymbol{\Lambda}(\mathbf{B})=\operatorname{diag}[\mu'(\mathbf{X}_1^{\mathrm{T}}\mathbf{B}),\ldots,\mu'(\mathbf{X}_N^{\mathrm{T}}\mathbf{B})].$$

Therefore, we obtain from (3.3)

$$\mathbf{V}(\hat{\mathbf{B}})=[\mathbf{X}^{\mathrm{T}}\boldsymbol{\Lambda}(\mathbf{B})\mathbf{X}]^{-1}\boldsymbol{\Sigma}(\mathbf{B})[\mathbf{X}^{\mathrm{T}}\boldsymbol{\Lambda}(\mathbf{B})\mathbf{X}]^{-1}, \tag{4.4}$$

where $\boldsymbol{\Sigma}(\mathbf{B})$ is the variance of a total based on observations $\{e_k\mathbf{x}_k\}$, for $k=1,\ldots,n$. Here e_k is the residual: $y_k-\mu(\mathbf{x}_k^{\mathrm{T}}\mathbf{B})$. Now, let $\hat{\boldsymbol{\Sigma}}(\hat{\mathbf{B}})$ be a consistent estimator of a total based on observations $\{\hat{e}_k\mathbf{x}_k\}$, for $k=1,\ldots,n$. Here $\hat{e}_k$ is the estimated residual: $y_k-\mu(\mathbf{x}_k^{\mathrm{T}}\hat{\mathbf{B}})$. Also let $\hat{\mathbf{J}}(\hat{\mathbf{B}})$ be a consistent estimator of $\mathbf{X}^{\mathrm{T}}\boldsymbol{\Lambda}(\mathbf{B})\mathbf{X}$. This is possible since each entry of the matrix is itself a total. From (3.4) we obtain

$$\hat{\mathbf{V}}(\hat{\mathbf{B}})=[\hat{\mathbf{J}}^{-1}(\hat{\mathbf{B}})]\hat{\boldsymbol{\Sigma}}(\hat{\mathbf{B}})[\hat{J}^{-1}(\hat{\mathbf{B}})]. \tag{4.5}$$

Note that if $\hat{\mathbf{B}}$ is solved by Newton–Raphson iterations on equation (4.3), then $\hat{\mathbf{J}}(\hat{\mathbf{B}})$ is just the matrix of derivatives required at each iteration. In particular, the iterative step is defined by

$$\hat{\mathbf{B}}^{(i+1)}=\hat{\mathbf{B}}^{(i)}-\hat{\mathbf{J}}^{-1}(\hat{\mathbf{B}}^{(i)})[\hat{\mathbf{Q}}(\hat{\mathbf{B}}^{(i)})-\hat{\mathbf{H}}]. \tag{4.6}$$

4.2 Ordinary least squares regression

The application of these results to regression analysis is now quite straightforward. We have $\mu(\mathbf{X}_k^{\mathrm{T}}\mathbf{B}) = \mathbf{X}_k^{\mathrm{T}}\mathbf{B}$, so that

$$\mathbf{W}_N(\mathbf{B}) = \mathbf{X}^{\mathrm{T}}\mathbf{Y} - \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{B}, \quad \frac{\partial \mathbf{W}_N(\mathbf{B})}{\partial \mathbf{B}} = -\mathbf{X}^{\mathrm{T}}\mathbf{X}.$$

Thus

$$\hat{\mathbf{V}}(\hat{\mathbf{B}}) = \mathbf{S}_{XX}^{-1}\hat{\mathbf{\Sigma}}(\hat{\mathbf{B}})\mathbf{S}_{XX}^{-1}, \tag{4.7}$$

where $\mathbf{S}_{XX}$ is an estimate for $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ and $\hat{\mathbf{\Sigma}}(\hat{\mathbf{B}})$ is an estimate for the variance of a total based on observations $\{\hat{e}_k\mathbf{x}_k\}$, where $\hat{e}_k = y_k - \mathbf{x}_k^{\mathrm{T}}\hat{\mathbf{B}}$. Fuller (1975) has obtained these results for stratified and for two-stage sampling.

To demonstrate how the same technique could be used to obtain the asymptotic variance of other than the linear model coefficients, let us consider the variance of $\hat{R}^2$, an estimate of the coefficient of multiple determination. Note that Hidiroglou, Fuller & Hickman (1980, p. 78) give this variance in terms of a ratio of two random variables. We write the following equations which implicity define our parameters:

$$\mathbf{Y}^{\mathrm{T}}\mathbf{1} - N\bar{Y} = 0, \quad \mathbf{X}^{\mathrm{T}}\mathbf{Y} - \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{B} = \mathbf{0}, \quad (\mathbf{Y}^{\mathrm{T}}\mathbf{Y} - N\bar{Y}^2)(R^2 - 1) + \mathbf{Y}^{\mathrm{T}}\mathbf{Y} - \mathbf{Y}^{\mathrm{T}}\mathbf{X}\mathbf{B} = 0. \tag{4.8}$$

Here $\bar{Y}$ is the mean of the Y values and $\mathbf{Y}^{\mathrm{T}}\mathbf{Y} - N\bar{Y}^2 = \mathrm{SS}Y$ is the corrected sum of squares for the Y values. We combined all the rows of the left-hand side of (4.8) into the matrix $\mathbf{W}(\bar{Y}, \mathbf{B}, R^2)$. If $\mathbf{B}$ has p components then $\mathbf{W}$ is a $(p+2)\times(p+2)$ matrix. We let $\mathbf{S}_{XX}$, S_{YY} and $\mathbf{S}_{XY}$ be estimators for $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, $\mathbf{Y}^{\mathrm{T}}\mathbf{Y}$ and $\mathbf{X}^{\mathrm{T}}\mathbf{Y}$ respectively. We also let $\bar{Y}^*$ be the estimator for $\bar{Y}$. Our other estimated parameters then are

$$\hat{\mathbf{B}} = \mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}, \quad \hat{R}^2 = 1 - \frac{S_{YY} - \hat{B}^{\mathrm{T}}\mathbf{S}_{XY}}{S_{YY} - N\bar{Y}^{*2}}$$

Taking derivatives of the $\mathbf{W}$ matrix we obtain

$$\frac{\partial W}{\partial(\bar{Y}, \mathbf{B}, R^2)} = \begin{bmatrix} -N & 0 & 0 \\ 0 & -\mathbf{X}^{\mathrm{T}}\mathbf{X} & 0 \\ 2N\bar{Y}(1-R^2) & -\mathbf{Y}^{\mathrm{T}}\mathbf{X} & \mathrm{SS}Y \end{bmatrix} \tag{4.9}$$

with an inverse of

$$\begin{bmatrix} -N^{-1} & 0 & 0 \\ 0 & -(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} & 0 \\ 2\bar{Y}(1-R^2)/\mathrm{SS}Y & -\mathbf{B}^{\mathrm{T}}/\mathrm{SS}Y & 1/\mathrm{SS}Y \end{bmatrix}. \tag{4.10}$$

To construct the variance estimate, we first define a $(p+2)$-dimensional vector for the kth unit sampled, based on equations (4.8). We let

$$c_{1k} = y_k, \quad \begin{bmatrix} c_{2k} \\ \vdots \\ c_{p+1,k} \end{bmatrix} = (y_k - \mathbf{x}_k^{\mathrm{T}}\hat{\mathbf{B}})\mathbf{x}_k, \quad c_{p+2,k} = (\hat{R}^2 y_k - x_k^{\mathrm{T}}\hat{B})y_k. \tag{4.11}$$

We let $\hat{\Sigma}_C$ be the estimated variance matrix for the estimated total of the c vectors. Now the estimated variance of $\hat{R}^2$, using (3.4) is

$$\mathrm{V}(\hat{R}^2) = (S_{YY} - N\bar{Y}^*)^{-2}\mathbf{a}^{\mathrm{T}}\hat{\mathbf{\Sigma}}_C\mathbf{a},$$

where

$$\mathbf{a}^{\mathrm{T}} = [2\bar{Y}^*(1-\hat{R}^2) - \hat{\mathbf{B}}^{\mathrm{T}}\mathbf{1}].$$

4.3 *Logistic regression*

For the logistic regression model, the dependent variable y is dichotomous: that is 0 or 1. Given vector $\mathbf{x}$, the probability that $y=1$ is $\exp(\mathbf{x}^T\boldsymbol{\beta})/[1+\exp(\mathbf{x}^T\boldsymbol{\beta})]$. This model is one of the generalized linear models in § 2.2, where $g(\theta)=\log(1+e^\theta)$ and $f(Y)=Y$. Therefore

$$\mu(\mathbf{X}_k^T\mathbf{B})=\frac{\exp(\mathbf{X}_k^T\mathbf{B})}{1+\exp(\mathbf{X}_k^T\mathbf{B})},$$

and $\mathbf{W}_N(\mathbf{B})$ is given by (4.2).

The kth diagonal element of $\mathbf{\Lambda}(\mathbf{B})$ in (4.4) is $\mu(\mathbf{X}_k^T\mathbf{B})[1-\mu(\mathbf{X}_k^T\mathbf{B})]$. The results of § 4.1 now follow. In particular, the estimated variance of $\hat{B}$ is given by (4.5).

4.4 *Log linear models for categorical data*

An extension of logistic regression for multiple categories is given by the log linear 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 $\mathbf{a}_i$ such that the proportion of individuals in the ith category is approximately

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

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

$$\mathbf{W}(\mathbf{B})=\mathbf{A}^T\mathbf{N}-[\mathbf{A}^T\mathbf{p}(\mathbf{B})]\mathbf{1}^T\mathbf{N}=\mathbf{0}, \tag{4.12}$$

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

We let $\hat{\mathbf{N}}$ be a consistent asymptotically normal estimator of $\mathbf{N}$, with variance-covariance matrix $\mathbf{V}[\hat{\mathbf{N}}]$, estimated by the matrix $\hat{\mathbf{V}}[\hat{\mathbf{N}}]$. Our estimator, $\hat{\mathbf{B}}$, satisfies

$$\hat{\mathbf{W}}(\hat{\mathbf{B}})=\mathbf{A}^T\hat{\mathbf{N}}-[\mathbf{A}^T\mathbf{p}(\hat{\mathbf{B}})]\mathbf{1}^T\hat{\mathbf{N}}=\mathbf{0}. \tag{4.13}$$

This estimator was suggested by Freeman & Koch (1976). It may be less efficient than Imrey, Koch & Stokes (1981, 1982) functional asymptotic regression methodology; however, we need not calculate all the components of $\hat{\mathbf{V}}[\hat{\mathbf{N}}]$ to apply (4.13).

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

$$\frac{\partial\mathbf{W}}{\partial\mathbf{B}}=-(\mathbf{1}^T\mathbf{N})\mathbf{A}^T\mathbf{H}(\mathbf{B})\mathbf{A}.$$

Therefore the asymptotic variance matrix for $\hat{\mathbf{B}}$ is given by

$$\mathbf{V}[\hat{\mathbf{B}}]=(\mathbf{N}^T\mathbf{1})^{-2}(\mathbf{A}^T\mathbf{H}(\mathbf{B})\mathbf{A})^{-1}A^T(\mathbf{I}-\mathbf{p}(\mathbf{B})\mathbf{1}^T)V[\hat{\mathbf{N}}](\mathbf{I}-\mathbf{1}\mathbf{p}(\mathbf{B})^T)\mathbf{A}(\mathbf{A}^T\mathbf{H}(\mathbf{B})\mathbf{A})^{-1}. \tag{4.14}$$

This expression can sometimes be simplied as follows. If it can be assumed that $\mathbf{N}/\mathbf{N}^T\mathbf{1}\simeq\mathbf{p}(\mathbf{B})$, then for $\hat{\boldsymbol{\pi}}=\hat{\mathbf{N}}/\hat{\mathbf{N}}^T\mathbf{1}$ we have

$$\mathbf{V}[\hat{\boldsymbol{\pi}}]\simeq(\mathbf{N}^T\mathbf{1})^{-2}(\mathbf{I}-\mathbf{p}(\mathbf{B})\mathbf{1}^T)\mathbf{V}[\hat{\mathbf{N}}](\mathbf{I}-\mathbf{1}\mathbf{p}(\mathbf{B})^T),$$

so that

$$\mathbf{V}[\mathbf{B}]\simeq(\mathbf{A}^T\mathbf{H}(\mathbf{B})\mathbf{A})^{-1}\mathbf{A}^T\mathbf{V}[\hat{\boldsymbol{\pi}}]\mathbf{A}(\mathbf{A}^T\mathbf{H}(\mathbf{B})\mathbf{A})^{-1}. \tag{4.15}$$

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

$$\mathbf{V}[\mathbf{p}(\hat{\mathbf{B}})]=\mathbf{H}(\mathbf{B})\mathbf{A}\mathbf{V}[\hat{\mathbf{B}}]\mathbf{A}^{\mathrm{T}}\mathbf{H}(\mathbf{B}).$$

The estimators of $\mathbf{V}[\hat{\mathbf{B}}]$ and $\mathbf{V}[\mathbf{p}(\mathbf{B})]$ are similar expressions, where $\mathbf{N}$ and $\mathbf{B}$ are replaced by $\hat{\mathbf{N}}$ and $\hat{\mathbf{B}}$ respectively. These assume that $\hat{\mathbf{V}}[\hat{\mathbf{N}}]$ 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} equal 1 if kth unit in ith category and equal 0 otherwise, for $k=1,\ldots,N$ and $i=1,\ldots,q$. Let $\mathbf{y}_k^{\mathrm{T}}=(y_{k1},\ldots,y_{kq})$. To construct the variance estimate, we define an r-dimensional vector $\mathbf{c}$ for the kth unit sampled, based on (4.12). We let $\mathbf{c}_k=\mathbf{A}^{\mathrm{T}}[\mathbf{I}-\mathbf{p}(\hat{\mathbf{B}})\mathbf{1}^{\mathrm{T}}]\mathbf{y}_k$. Letting $\hat{\mathbf{\Sigma}}_C$ be the estimated variance matrix for the estimated total of the c vectors we obtain

$$\hat{\mathbf{V}}[\hat{\mathbf{B}}]=(\hat{\mathbf{N}}^{\mathrm{T}}\mathbf{1})^{-2}(\mathbf{A}^{\mathrm{T}}\mathbf{H}(\hat{\mathbf{B}})\mathbf{A})^{-1}\hat{\mathbf{\Sigma}}_C(\mathbf{A}^{\mathrm{T}}\mathbf{H}(\mathbf{B})\mathbf{A})^{-1}.$$

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.

5 Canada Health Survey

5.1 *Description of the survey*

The Canada Health Survey, 1978–1979 (Health and Welfare Canada and Statistics Canada, 1981) was a national household survey, conducted jointly by Statistics Canada and Health and Welfare Canada to provide information on the health status of Canadians. The sample design was a highly stratified multistage sample to select households and all residents within selected households comprised the respondents. In total, 10,571 dwellings or 31,668 persons were selected, after removing the 14% household nonresponse. The data we use in this study is restricted to the population of persons 15 years of age or older who responded to the self-administered questionnaire. After removing the person-level nonresponse for the self-administered questionnaire, we have 20,726 respondents in the sample. The sampling weights were adjusted for household and person-level nonresponse by age and sex groupings at the provincial levels using census population projections.

For the purpose of variance estimation, we assumed that the primary sampling units within strata were selected with replacement. There are a total of 44 strata and 100 primary sampling units. Each stratum contained 2, 3 or 4 primary sampling units. We assume that the number of strata is sufficiently large and all moments are bounded so that the central limit theorem of Krewski & Rao (1981, Lemma 3.1) applies.

5.2 *Explaining physician use and nonuse*

In this paper, we consider the problem of explaining the users and nonusers of physician services. Our dependent variable is:

$$\text{PHYSUSE}\begin{cases}0 & \text{nonuser of physician services over 12 month period,}\\ 1 & \text{user of physician services over 12 month period.}\end{cases}$$

To explain the use or nonuse of physician services, we consider the following variables:

$$\text{AGE2024}\begin{cases}1 & 20\leqslant \text{age}\leqslant 24,\\ 0 & \text{otherwise;}\end{cases}\quad \text{AGE2544}\begin{cases}1 & 25\leqslant \text{age}\leqslant 44,\\ 0 & \text{otherwise;}\end{cases}$$

AGE4564 $\begin{cases} 1 & 45 \leqslant \text{age} \leqslant 64, \\ 0 & \text{otherwise;} \end{cases}$ AGEGE65 $\begin{cases} 1 & \text{age} \geqslant 65, \\ 0 & \text{otherwise;} \end{cases}$

SEX $\begin{cases} 0 & \text{male,} \\ 1 & \text{female;} \end{cases}$

AS2024–ASGE65, age–sex interactions;

QUINT1–QUINT4, a 0–1 variable indicating family income quintile of respondent;

INCUNK $\begin{cases} 1 & \text{family income unknown,} \\ 0 & \text{otherwise;} \end{cases}$

WHITEC $\begin{cases} 1 & \text{employed, white collar worker,} \\ 0 & \text{otherwise;} \end{cases}$

BLUECOL $\begin{cases} 1 & \text{employed, blue collar worker,} \\ 0 & \text{otherwise;} \end{cases}$

OCCUNK $\begin{cases} 1 & \text{employed, unknown occupation,} \\ 0 & \text{otherwise;} \end{cases}$

SEXWH, SEX * WHITEC; SEXBLUE, SEX * BLUECOL; SEXOU, SEX * OCCUNK;

MARRIED $\begin{cases} 1 & \text{married,} \\ 0 & \text{not married;} \end{cases}$ SEPDIVW $\begin{cases} 1 & \text{separated, widowed or divorced,} \\ 0 & \text{otherwise;} \end{cases}$

MSUNK $\begin{cases} 1 & \text{marital status unknown,} \\ 0 & \text{otherwise;} \end{cases}$

SELFMH, number of conditions in medical history among heart trouble, high blood pressure, stroke, diabetes, cancer;

SELFUNK $\begin{cases} 1 & \text{if SELFMH} = 0 \text{ and at least 1 response was ‘not sure’,} \\ 0 & \text{otherwise;} \end{cases}$

HLTHPROB, number of health problems found in the survey;

DRUGUSE, $\begin{cases} 1 & \text{used drug on advice of medical doctor in last 2 days,} \\ 0 & \text{otherwise;} \end{cases}$

DRUGUNK $\begin{cases} 1 & \text{DRUGUSE} = 0 \text{ and used drug in last 2 days but} \\ & \text{does not know if on advice of medical doctor,} \\ 0 & \text{otherwise;} \end{cases}$

ACCIDENT, number of accidents in 12 month period;

NOACC $\begin{cases} 1 & \text{if ACCIDENT} = 0, \\ 0 & \text{otherwise;} \end{cases}$

DISABILT, number of disability days (bed days, activity loss days and cut down days) in last 2 weeks

DIS∅1 $\begin{cases} 0 & \text{if DISABILT} = 0, \\ 1 & \text{otherwise;} \end{cases}$

MAJCITY $\begin{cases} 1 & \text{if resides in community with population 100,000–999,999,} \\ 0 & \text{otherwise;} \end{cases}$

URBAN $\begin{cases} 1 & \text{if resides in community with population 1000,000 or more,} \\ 0 & \text{otherwise;} \end{cases}$

PHPRATIO, the ratio of physicians to 100 persons in the population by province (1978 data).

A summary of the weighted means, and standard deviations for these variables is given in Table 1.

Table 1
Variable of the study

Variable	Mean	St. dev.	Variable	Mean	St. dev.	Variable	Mean	St. dev
PHYSUSE	0.7692	0.4213	QUINT3	0.1724	0.3777	SELFMH	0.2533	0.5744
AGE2024	0.1266	0.3325	QUINT4	0.2007	0.4005	SELFUNK	0.0635	0.2438
AGE2544	0.3700	0.4828	INCUNK	0.0542	0.2265	HLTHPROB	1.3384	1.6286
AGE4564	0.2546	0.4356	WHITEC	0.3603	0.4823	DRUGUSE	0.3104	0.4626
AGEGE65	0.1154	0.3195	BLUECOL	0.1899	0.3922	DRUGUNK	0.0004	0.0201
SEX	0.5092	0.4999	OCCUNK	0.0042	0.0648	ACCIDENT	0.1191	0.4216
AS2024	0.0635	0.2436	SEXWH	0.1836	0.3872	NOACC	0.9008	0.2990
AS2544	0.1854	0.3886	SEXBLUE	0.0225	0.1483	DISABILT	0.6803	2.5370
AS4564	0.1303	0.3366	SEXOU	0.0018	0.0427	DISØ1	0.1197	0.3246
ASGE65	0.0647	0.2461	MARRIED	0.6321	0.4822	MAJCITY	0.2937	0.4555
QUINT1	0.1727	0.3780	SEPDIVW	0.0988	0.2985	URBAN	0.3296	0.4701
QUINT2	0.1695	0.3752	MSUNK	0.0168	0.1286	PHPRATIO	0.1805	0.0173

Using PHYSUSE as our dependent variable, we performed both regression and logistic regression analyses, using the methodologies described in §§ 4.2 and 4.3. Table 2 summarizes the results of the analysis with tests of significance.

If one pays attention to only the sign of the coefficients the two models give consistent results. From a purely descriptive point of view, both models tend to lead to the same basic conclusions:

(a) females in fertility years are more likely to be users than their male counterparts;
(b) there may be a marginal income effect, with the lower middle classes being users of physicians services less often;
(c) medical history, use of drugs, health problems, disability days, are all important determinants of physician use;
(d) marital status is an important variable, single people being lower users than others;
(e) if the person has had at least one accident he is more likely to use physician services;
(f) rural residents are less likely to be users than their city counterparts.

5.3 *Qualitative diagnostics*

The information contained in Table 2 does not indicate adequately which of the two models are preferable. A simple qualitative diagnostic would be to compare the predicted probabilities of use of physician services with the observed probabilities, with certain probability groupings. In Table 3, we cross-tabulate the 20,726 respondents.

Table 2
Regression and logistic regression

Variable	d.f.	Ordinary regression Coefficient	χ^2	Logistic regression Coefficient	χ^2
Age	4		12.499		19.232
AGE2024	1	0.0339	1.894	0.0750	0.333
AGE2544	1	−0.0049	0.055	−0.1928	3.364
AGE4564	1	−0.0052	0.060	−0.2920	6.650
AGEGE65	1	0.0354	2.283	−0.0527	0.130
SEX	1	0.0847	17.304	0.3797	12.494
AGE SEX	4		39.760		36.001
AS2024	1	0.0591	7.108	0.6573	20.479
AS2544	1	0.0431	2.396	0.4452	8.417
AS4564	1	−0.0520	4.761	−0.1458	1.084
ASGE65	1	−0.0926	13.376	−0.4370	6.483
Income	5		15.060		14.642
QUINT1	1	−0.0066	0.337	−0.0186	0.053
QUINT2	1	−0.0258	6.481	−0.1395	5.288
QUINT3	1	−0.0167	1.880	−0.0907	1.302
QUINT4	1	−0.0062	0.349	−0.0327	0.203
INCUNK	1	−0.0655	7.300	−0.3486	8.004
Occupation	3		14.386		8.614
WHITEC	1	−0.0113	0.522	−0.0285	0.111
BLUECOL	1	0.0047	0.639	0.0682	0.430
OCCUNK	1	0.2043	12.457	1.2967	7.018
Occupation & sex	3		16.746		11.501
SEXWH	1	0.0470	5.574	0.2743	4.498
SEXBLUE	1	0.0297	1.146	0.1627	0.727
SEXOU	1	−0.1881	4.064	−1.0611	1.990
Marital status	3		46.715		45.752
MARRIED	1	0.0753	43.920	0.4748	44.880
SEPDIVW	1	0.0643	18.888	0.4653	22.394
MSUNK	1	0.0036	0.011	0.0386	0.041
History	2		43.244		36.700
SELFMH	1	0.0399	31.085	0.4309	24.579
SELFUNK	1	−0.0403	3.273	−0.2387	3.610
HLTHPROB	1	0.0293	111.499	0.3280	81.554
Drugs	2		421.086		272.175
DRUGUSE	1	0.1363	404.953	1.0059	253.074
DRUGUNK	1	0.1038	0.300	0.4053	0.145
Accidents	2		180.560		106.372
ACCIDENT	1	−0.0115	1.481	−0.0707	0.482
NOACC	1	−0.1718	108.135	−1.4489	71.118
Disability days	2		11.401		29.052
DISABILT	1	−0.0002	0.012	0.0797	6.228
DIS∅1	1	0.0408	11.381	0.1334	0.925
Community size	2		10.725		11.751
MAJCITY	1	0.0295	8.129	0.1975	8.230
URBAN	1	0.0371	8.636	0.2585	9.755
PHPRATIO	1	0.1864	0.060	1.0817	0.540
Intercept	1	0.6823	229.857	1.1800	13.671

Table 3
Cross-classification of predicted and observed probabilities

		Predicted probability range (%) 30–40	40–50	50–60	60–70	70–80	80–90	90–100	100+	Total
Ordinary regression	Number	0	116	2728	5513	4597	3866	2506	1400	20726
	% users	—	35.3	49.5	63.0	79.3	90.1	95.5	97.1	76.0
Logistic regression	Number	102	1325	3260	2980	3189	4200	5670	—	20726
	% users	36.3	42.6	55.6	64.4	76.2	86.1	94.7	—	76.0

We see that the observed proportion of users in each probability range closely matches the expected numbers for the logistic regression analysis; whereas, the ordinary regression behaves poorly, especially at the extremes. Therefore, heuristically at least, the logistic model is preferable.

6 Discussion

The techniques described in the paper have been given previously for some specific models; see, for example, Fuller (1975) and Freeman & 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 & Stokes (1981, 1982). Their functional asymptotic regression methodology also falls within the general framework described here, with respect to variance derivation and estimation.

Acknowledgements

I am grateful to J.N.K. Rao, M.A. Hidiroglou, G.G. Koch and W.A. Fuller for their comments to an earlier draft of this paper. I would also like to thank R. Broyles, P. Manga, D. Angus and A. Charette for their advice on developing the models of § 5. I am grateful to D. Paton for helping in the development of the computer programs.

Appendix

Asymptotic distribution of $\hat{\boldsymbol{\theta}}$

In this appendix, we provide conditions for the asymptotic normality of $\hat{\boldsymbol{\theta}}$, defined by (3.2) and for the validity of variance formulae (3.3) and (3.4).

First we suppose there is a sequence of populations, indexed by $t = 1, 2, 3, \ldots$. The population sizes are N_t. The function $\mathbf{W}_{N_t}(\boldsymbol{\theta})$ is defined for all $\boldsymbol{\theta} \in \Theta$, the parameter space.

Condition 1. We assume

$$\lim_{t\to\infty} \frac{\mathbf{W}_{N_t}(\boldsymbol{\theta})}{N_t}$$

exists and equals $\boldsymbol{\Omega}(\boldsymbol{\theta})$.

Condition 2. We assume $\boldsymbol{\Omega}(\boldsymbol{\theta})$ is a one-to-one function, so that $\boldsymbol{\Omega}^{-1}(.)$ exists.

Condition 3. There exists a $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ such that $\boldsymbol{\Omega}(\boldsymbol{\theta}_0) = \mathbf{0}$.

Condition 4. The parameter space Θ contains a neighbourhood of $\boldsymbol{\theta}_0$.

We now suppose that a sequence of samples of sizes $n_1, n_2, \ldots$ are taken and that for each sample we construct the function $\hat{\mathbf{W}}(\boldsymbol{\theta})$, defined in (3.1).

Condition 5. We assume

$$\frac{n^{\frac{1}{2}}}{N_t} [\hat{\mathbf{W}}(\boldsymbol{\theta}) - \mathbf{W}_{N_t}(\boldsymbol{\theta})]$$

converges in distribution to the normal law with mean 0 and positive-definite variance matrix $\boldsymbol{\Phi}(\boldsymbol{\theta})$ as $t \to \infty$, for all $\boldsymbol{\theta}$ in a neighbourhood of $\boldsymbol{\theta}_0$.

Conditions required for condition 5 to be satisfied are given in the literature; see, for example, Madow (1948), Hajék (1960, 1964), Rosen (1972), von Bahr (1972), Fuller (1975), Krewski & Rao (1981).

Condition 6. We assume $\hat{\mathbf{W}}(.)$ is totally differentiable in a neighbourhood of $\boldsymbol{\theta}_0$.

Condition 7. We assume

$$\lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0,\, t\to\infty}\left[N_t^{-1}\left(\frac{\partial\hat{\mathbf{W}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\right]$$

exists and equals

$$\left.\frac{\partial\boldsymbol{\Omega}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}=\mathbf{J}(\boldsymbol{\theta}_0).$$

Condition 8. The matrix $\mathbf{J}(\boldsymbol{\theta}_0)\boldsymbol{\Phi}(\boldsymbol{\theta}_0)\mathbf{J}(\boldsymbol{\theta}_0)$ is full rank.

Condition 9. We assume $\boldsymbol{\Phi}(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$.

Condition 10. We assume $\partial\boldsymbol{\Omega}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is a continuous function of $\boldsymbol{\theta}$.

Condition 11. A consistent estimator for $\boldsymbol{\Phi}(\boldsymbol{\theta})$ is $\hat{\phi}(\theta)$, assumed to exist under the sample design.

LEMMA 1. *The asymptotic distribution of* $n_t^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ *is the same as the asymptotic distribution of*

$$-\mathbf{J}^{-1}(\boldsymbol{\theta}_0)\frac{n_t^{\frac{1}{2}}}{N_t}\hat{\mathbf{W}}(\boldsymbol{\theta}_0).$$

Proof. For any $\boldsymbol{\theta}$ in the neighbourhood of $\boldsymbol{\theta}_0$, we have

$$\hat{\mathbf{W}}(\hat{\boldsymbol{\theta}})=\hat{\mathbf{W}}(\boldsymbol{\theta})+\frac{\partial\hat{\mathbf{W}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})-\boldsymbol{\varepsilon}\,\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|, \tag{A.1}$$

where $\boldsymbol{\varepsilon}\to\mathbf{0}$ as $\hat{\boldsymbol{\theta}}\to\boldsymbol{\theta}$. Since the left-hand side is zero, we may rewrite this as

$$\hat{\mathbf{W}}(\boldsymbol{\theta})+\frac{\partial\hat{\mathbf{W}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})=\boldsymbol{\varepsilon}\,\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|. \tag{A.2}$$

Now, since $\mathbf{W}_{N_t}(\boldsymbol{\theta}_0)/N_t\to\boldsymbol{\Omega}(\boldsymbol{\theta}_0)=\mathbf{0}$ as $t\to\infty$, and $N_t^{-1}\,\partial\hat{W}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\boldsymbol{\Omega}_0\to\mathbf{J}(\boldsymbol{\theta}_0)$ as $t\to\infty$, we have

$$\frac{n_t^{\frac{1}{2}}}{N_t}\hat{\mathbf{W}}(\boldsymbol{\theta}_0)+n_t^{\frac{1}{2}}\mathbf{J}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=\frac{n_t^{\frac{1}{2}}}{N_t}\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|\,\boldsymbol{\varepsilon}. \tag{A.3}$$

Using the same argument as in Rao (1973, p. 386), the right-hand side tends to 0 in probability. The result is thus proved, since by condition 8 the matrix $\mathbf{J}(\boldsymbol{\theta}_0)$ is full rank.

COROLLARY 1. *The asymptotic distribution of* $n_t^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ *is the normal law with mean* $\mathbf{0}$ *and variance matrix* $[\mathbf{J}^{-1}(\boldsymbol{\theta}_0)][\boldsymbol{\Phi}(\boldsymbol{\theta}_0)][\mathbf{J}^{-1}(\boldsymbol{\theta}_0)]^{\mathrm{T}}$. *This is the limiting value of* (3.3).

COROLLORY 2. *Let* F_t *be the distribution function for* $n_t^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$, *based on the* tth *sample and let* G_t *be the distribution function of a multivariate normal distribution with mean zero and variance matrix* $[\hat{\mathbf{J}}^{-1}(\hat{\boldsymbol{\theta}})][\hat{\boldsymbol{\Phi}}(\hat{\boldsymbol{\theta}})][\hat{\mathbf{J}}^{-1}(\hat{\boldsymbol{\theta}})]^{\mathrm{T}}$, *where* $\hat{\mathbf{J}}(\boldsymbol{\theta})=N_t^{-1}\,\partial\hat{W}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. *Then, by virtue of conditions* 9 *to* 11

$$\lim_{t\to\infty}\sup|F_t-G_t|=0.$$

This result shows the validity of using (3.4) to estimate (3.3). The proof is analogous to Rao (1973, p. 389).

References

Frankel, M.R. (1971). Inference from survey samples. University of Michigan, Ann Arbor.
Freeman, D.H. & Koch, G.G. (1976). An asymptotic covariance structure for testing hypotheses on raked contingency tables from complex sample surveys. In *Proc. Am. Statist. Assoc., Social Statist. Sect.*, 330–335.
Fuller, W.A. (1975). Regression analysis for sample survey. *Sankhyā* C **37,** 117–132.
Hajék, J. (1960). Limiting distributions in simple random sampling from a finite population. *Pub. Math. Inst. Hung. Acad. Sci.* **5,** 361–374.
Hajék, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Ann. Math. Statist.* **35,** 1491–1523.
Health and Welfare Canada and Statistics Canada (1981). *The Health of Canadians: Report of the Canada Health Survey*. Statistics Canada Catalogue No. 82–538E.
Hidiroglou, M.A., Fuller, W.A. & Hickman, R.D. (1980). Super Carp. Statistical Laboratory, Iowa State University.
Imrey, P.B., Koch, G.G. & Stokes, M.E. (1981). Categorical data analysis: Some reflections on the log linear model and logistic regression. Part I: Historical and methodological overview. *Int. Statist. Rev.* **49,** 265–283.
Imrey, P.B., Koch, G.G. & Stokes, M.E. (1982). Categorical data analysis: Some reflections on the log linear model and logistic regression. Part II: Data analysis. *Int. Statist. Rev.* **50,** 35–63.
Kish, L. & Frankel, M.R. (1974). Inference from complex samples. *J. R. Statist. Soc.* B **36,** 1–22.
Krewski, D. & Rao, J.N.K. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *Ann. Statist.* **9,** 1010–1019.
Madow, W.G. (1948). On the limiting distribution of estimates based on samples from finite universes. *Ann. Math. Statist.* **19,** 535–545.
Nelder, J.A. & Wedderburn, R.W.M. (1972). Generalized linear models. *J. R. Statist. Soc.* A **135,** 370–384.
Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, 2nd edition. New York: Wiley.
Rosen, B. (1972). Asymptotic theory for successive sampling with varying probabilities without replacement, I and II. *Ann. Math. Statist.* **43,** 373–397 and 748–776.
Särndal, C.E. (1978). Design-based and model-based inference in survey sampling. *Scand. J. Statist.* **5,** 27–52.
Tepping, B.J. (1968). The estimation of variance in complex surveys. *Proc. Am. Statist. Assoc., Social Statist. Sect.*, 11–18.
von Bahr, B. (1972). On sampling from a finite set of independent random variables. *Z. Wahr. verw. Geb.* **24,** 279–286.
Woodruff, R.S. (1971). A simple method for approximating the variance of a complicated estimate. *J. Am. Statist. Assoc.* **66,** 411–414.

Résumé

On discute le problème de la spécification et de l'estimation de la variance de paramètres estimés basés sur les plans d'échantillonnage complexes provenant de populations finies. Les résultats présentés dans cet article sont particulièrement utiles lorsque les estimateurs des paramétres ne sont pas définis explicitement comme étant une fonction des autres statistiques de l'échantillon. On montre comment ces résultats peuvent s'appliquer à la régression linéaire, la régression logistique et aux modèles linéaires logarithmiques de tableaux de contingence. Une example de l'application de la methodologie à l'Enquête Santé Canada est donnée.

[*Paper received May* 1982, *revised February* 1983]