A GENERALIZED WEIGHTING PROCEDURE BASED ON LINEAR MODELS

Jelke G. Bethlehem and Wouter J. Keller
Netherlands Central Bureau of Statistics

1. Introduction

A sample survey is an instrument for making inferences about a finite population using observations on only a part of the elements in the population. In order to improve the precision of estimates of population parameters and to correct for biases caused by problems met in the sampling process, often some kind of weighting is carried out. Weights are assigned to the observed elements in such a way that proper estimates are obtained by simple summation of the weighted observations.

Bailar et al. (1978) describe weighting as a currently used adjustment method to correct for a possibly existing bias. Also Platek and Gray (1980) and Lindström et al. (1979) present weighting as an important method to correct for bias due to non-respons. Even if no problems are encountered in the sample survey process it may still be worthwhile to perform some kind of weighting. Post-stratification (see e.g. Holt and Smith, 1979) is a well-known and much used weighting method which often produces estimators which are much more precise than direct estimators such as the sample mean multiplied by the population size.

This paper presents a general framework for weighting based on estimators constructed from linear models. It will be shown that classical weighting emerges from the theory as a special case. Due to the generality of the theory it presents a number of other possibilities for weighting which are especially useful in situations where classical weighting may cause trouble. More details on the theory can be found in Bethlehem and Keller (1983).

In sections 2 and 3 the basic notations are introduced. Section 4 presents the regression estimator. In section 5 the theory is applied to simple random sampling. Section 6 shows that post-stratification is a special case of the theory. Section 7 offers a solution for the situation in which post-stratification causes trouble. Applications of the theory are given in section 8. In section 9 a computer program is discussed and section 10 gives some suggestions for models for weighting which are not included in the theory.

2. Population and sample

Let the targetpopulation U of the sample survey consist of N identifiable elements, which may be labeled 1,2,...,N. Associated with each element k are the (unknown) value y_k of a quantitative targetvariable and the p-vector $x_k=(x_{k1},x_{k2},\ldots,x_{kp})'$ of values of p auxiliary variables. Let $y=(y_1,y_2,\ldots,y_N)'$ denote the vector of all values of the target variable. Let X be the N×p-matrix of values of the auxiliary variables, the k-th row of X corresponding to x_k'.

We assume the objective of the sample survey to be estimation of the population mean

$$\bar{y} = 1_N'y/N \,, \qquad (2.1)$$

where 1_N is the N-vector consisting of one's. The p-vector of population means of the p auxiliary variables is denoted by

$$\bar{x} = X'1_N/N \,. \qquad (2.2)$$

We restrict ourselves to sampling without replacement. In that case a sample from the finite population U can be denoted by an N×N-diagonal matrix T. The k-th diagonale element t_k of T assumes the value 1 of the corresponding element is in the sample, and t_k assumes the values 0 if this is not the case. The expected value of T is equal to

$$E(T) = \Pi \,, \qquad (2.3)$$

where Π is the N×N-diagonal matrix of first order inclusion probabilities $\pi_1,\pi_2,\ldots,\pi_N$. For the computation of the variance of estimators we will make use of the N×N-matrix Δ of which the ij-th element is equal to

$$\delta_{ij} = \frac{\pi_{ij}}{\pi_i\pi_j} - 1 \,, \qquad (2.4)$$

where π_{ij} is the second order inclusion probability and π_{ii} is to be taken equal to π_i. Furthermore for variance estimation it is convenient to use the N×N-matrix Λ of which the ij-th element is equal to

$$\lambda_{ij} = \left(\frac{1}{\pi_i\pi_j} - \frac{1}{\pi_{ij}}\right) . \qquad (2.5)$$

Observe that in this notation the Horvitz-Thompson (1952) estimator for the population mean for $\bar{y}$ can be written as

$$\hat{\bar{y}}_{HT} = 1_N'\Pi^{-1}Ty/N \,, \qquad (2.6)$$

with a variance equal to

$$V(\hat{\bar{y}}_{HT}) = y'\Delta y/N^2 \,. \qquad (2.7)$$

An estimate of the variance by Horvitz and Thompson (1952) is equal to

$$\hat{V}(\hat{\bar{y}}_{HT}) = y'T\Lambda Ty/N^2 \,. \qquad (2.8)$$

3. The regression model

If auxiliary variables are correlated with the target variable, they can be used to construct precise estimators. Such a relationship implies that for a suitably chosen p-vector $\beta=(\beta_1,\beta_2,\ldots,\beta_p)'$ of regression coefficients, the elements in the N-vector

$$\varepsilon = y - X\beta \qquad (3.1)$$

of residuals vary less than the values of target variable itself. Observe that all quantities in (3.1) are fixed numbers. There are no random variables. An obvious criterion to measure the variation of the residuals is the residual sum of squares

$$\varepsilon'\varepsilon = (y-X\beta)'(y-X\beta) \ . \qquad (3.2)$$

Application of ordinary least squares results in

$$\beta = (X'X)^{-1}X'y \ , \qquad (3.3)$$

as the value for which (3.2) is minimized. In general the vector β will not be known, particularly because the vector y is unknown. The obvious solution to this problem is to estimate β from the sample by

$$\hat{\beta} = (X'\Pi^{-1}TX)^{-1}X'\Pi^{-1}Ty \ . \qquad (3.4)$$

The estimator $\hat{\beta}$ is not unbiased, but it can be shown that the bias is of order $n^{-\frac{1}{2}}$. So $\hat{\beta}$ is approximately unbiased for large samples.

4. The regression estimator

It is not our first objective to estimate β. What we need is an estimator for the population mean $\bar{y}$. We define the regression estimator of y by

$$\hat{\bar{y}}_R = 1_N'X\hat{\beta}/N = \bar{x}'\hat{\beta} \ . \qquad (4.1)$$

Since $\hat{\beta}$ is a consistent estimator of β, $X\hat{\beta}$ is a consistent estimator of $X\beta$. But (4.1) is a consistent estimator of $\bar{y}$ if, and only if $\bar{x}'\beta=\bar{y}$ and that is the case if, and only if there exists a p-vector c of fixed numbers such that $Xc=1_N$.

Under the restriction $Xc=1_N$ the regression estimator can be written in the somewhat different form

$$\hat{\bar{y}}_R = 1_N'\Pi^{-1}Ty/N + (1_N'X - 1_N'\Pi^{-1}TX)\hat{\beta}/N \ . \qquad (4.2)$$

We will call estimator (4.2), or its equivalent (4.1) the <u>regression estimator</u>. In case of simple random sampling and use of only one auxiliary variable estimator (4.2) reduces to the simple regression estimator as e.g. given in Cochran (1977). The estimator presented in (4.2) can be considered as a generalized version of this simple regression estimator. It is generalized in two ways: more than one auxiliary variable can be used and any without replacement sampling design may be applied.

The regression estimator is not an unbiased estimator. However, it can be shown that its bias is of order 1/n. The variance of the estimator can be approximated by

$$V(\hat{\bar{y}}_R) \doteq \varepsilon'\Delta\varepsilon/N^2 \ . \qquad (4.3)$$

Furthermore (4.3) can consistently be estimated by

$$\hat{V}(\hat{\bar{y}}_R) = \hat{\varepsilon}T\Lambda T\hat{\varepsilon}/N^2 \qquad (4.4)$$

where $\hat{\varepsilon}=y-X\hat{\beta}$.

The use of the proposed regression estimator implicitly produces weights which are to be assigned to the observations. Introducing the N-vector of weights

$$w = \Pi^{-1}TX(X'\Pi^{-1}TX)^{-1}\bar{x} \ , \qquad (4.5)$$

and recalling that $\hat{\bar{y}}_R$ can be written as

$$\hat{\bar{y}}_R = \bar{x}'(X'\Pi^{-1}TX)^{-1}X'\Pi^{-1}Ty \ , \qquad (4.6)$$

it is obvious that

$$\hat{\bar{y}}_R = w'y \ . \qquad (4.7)$$

If we used the weights to estimate the means of the auxiliary variables we would get

$$w'X = \bar{x}'(X'\Pi^{-1}TX)^{-1}X'\Pi^{-1}TX = \bar{x}' \ . \qquad (4.8)$$

Indeed the weights balance the sample such that the sample distribution of the auxiliary variables agrees the population distribution of these variables. The regression estimator is a proper means for computing weights.

5. Simple random sampling

The first illustration of the theory, which was developed in section 4, is simple random sampling. Introducing y_s as the n-vector of sampled values of the target variable and X_s as the n×p-matrix of auxiliary variables corresponding to sampled elements, the regression estimator reduces to

$$\hat{\bar{y}}_R = \bar{y}_s + (\bar{x} - \bar{x}_s)'\hat{\beta} \ , \qquad (5.1)$$

in which $\bar{y}_s$ is the mean of the elements in y_s, $\bar{x}_s$ is the p-vector of sample means of the auxiliary variables, and

$$\hat{\beta} = (X_s'X_s)^{-1}X_s'y_s \ . \tag{5.2}$$

In (5.1) we once more recognize the simple regression estimator, be it that here $\bar{x}$, $\bar{x}_s$ and $\hat{\beta}$ are vectors instead of scalars. Working out the variance (4.13) in this case gives

$$V(\hat{\bar{y}}_R) \doteq \frac{1-f}{n}\,(y-X\beta)'(y-X\beta)/(N-1) \ . \tag{5.3}$$

This result confirms the approximation given by Cochran (1977), in case of the simple regression estimator. This approximated variance can consistently be estimated by

$$\hat{V}(\hat{\bar{y}}_R) = \frac{1-f}{n}\,(y_s-X_s\hat{\beta})'(y_s-X_s\hat{\beta})/(n-1) \ . \tag{5.4}$$

So the search for a precise estimator comes down to looking for a linear model which fits the data as good as possible. It means that we can use the variable selection techniques from the theory of linear models.

6. One-way-stratification

The use of the regression estimator is not restricted to quantitative auxiliary variables. In this section and the sections 7 and 8 we will explore the case of qualitative auxiliary variables. In this section we will consider the use of one such variable.

In order to include a qualitative variable in a linear model, it is replaced by as much dummy variables as it has categories. Suppose our auxiliary variable has p categories. Then it induces a division of the population into p non-overlapping sub-populations (strata). To each category there corresponds a dummy variable which assumes the value 1 if the particular element is contained in that stratum, and otherwise it assumes the value 0. So, for every element only one dummy variable assumes the value 1; all other values are 0. Consequently the matrix X consists of N rows, each row of which contains exactly one 1. The columns of X sum up to the sub-population totals $N_1,N_2,\ldots,N_p$, where $N_1+N_2+\ldots+N_p=N$.

From the thus structured population a simple random sample without replacement of size n is selected. So we can keep on using the notation of section 5. The columns of X_s will sum up to the (random) sample totals $n_1,n_2,\ldots,n_p$ in the strata, where $n_1+n_2+\ldots+n_p=n$. The vector of population means of the auxiliary variables is equal to $\bar{x}=(N_1,N_2,\ldots,N_p)'/N$ and the corresponding vector of sample means is equal to $\bar{x}_s=(n_1,n_2,\ldots,n_p)'/n$.

Due to the special structure of the matrix X the matrix $X_s'X_s$ is a diagonal matrix with diagonal elements equal to $n_1,n_2,\ldots,n_p$. Substitution in (5.2) results in

$$\hat{\beta} = (\bar{y}_s^{(1)},\bar{y}_s^{(2)},\ldots,\bar{y}_s^{(p)})' \ , \tag{6.1}$$

where $\bar{y}_s^{(h)}$ is the sample mean of the target variable in stratum h (h=1,2,...,p). Substitution of (6.1) in (5.1) gives as the regression estimator in this case

$$\hat{\bar{y}}_R = \sum_{h=1}^{p} N_h\bar{y}_s^{(h)}/N = \hat{\bar{y}}_{PS} \ , \tag{6.2}$$

in which $\hat{\bar{y}}_{PS}$ is the traditional post-stratification estimator. So post-stratification is a special case of the regression estimator. Since only one qualitative variable is used we will call this case <u>one-way-stratification</u>. Section 7 will deal with multiway stratification.

The estimator $\hat{\bar{y}}_{PS}$ is only defined if there is at least one observation available in every stratum. The same applies for the regression estimator. If there are no observations in one or more strata, then some of the diagonal elements of $X_s'X_s$ are zero, in which case $X_s'X_s$ is singular.

Application of (5.3) gives an approximated variance equal to

$$V(\bar{y}_R) \doteq \frac{1-f}{n}\sum_{h=1}^{p}\frac{N_h-1}{N-1}S_h^2 \ , \tag{6.3}$$

in which S_h^2 is the variance (with denominator N_h-1) in subpopulation h. This is a somewhat different expression than given by e.g. Cochran (1977), p. 135. The difference is mainly due to leaving out terms of order n^{-2} in (6.3).

7. Multiway stratification

Application of post-stratification is not restricted to the use of one qualitative auxiliary variable. The theory is equally well applicable for more than one qualitative auxiliary variable. Suppose we have m such variables with numbers of categories equal to $p_1,p_2,\ldots,p_m$. Now every combination of values of the auxiliary variables induces a stratum, the total number of strata being equal to $p=p_1\times p_2\times\ldots\times p_m$. If the m qualitative auxiliary variables are replaced by p dummy variables then the theory of section 6 can be applied.

If the theory of linear models is restricted to use of qualitative independent variables it is usually called analysis of variance. That is why the terminology we are going to introduce has its roots in the analysis of variance. The auxiliary variables correspond to factors and the strata to cells. Stratification in which strata are constructed on the basis of all possible combinations of values of the auxiliary variables corresponds to an analysis of variance in which the model contains the highest order interaction. That is the reason why we call this type of post-stratification <u>complete multiway stratification</u>. Complete multiway stratification is not always practicable. Two major problems may be present itself.

The first problem is the problem of empty strata. If there is so much auxiliary information available that it allows complete multiway stratification, the resulting number of strata may be so large that there exists a risk of empty strata in the sample. In daily practice this problem is usually solved by collapsing strata. As this is frequently done by hand it is a time consuming process.

The second problem is created by lack of auxiliary information. Sometimes there is auxiliary information available, but is not detailed enough to allow complete multiway stratification. For instance, complete multiway stratification by sex, age, marital status and region requires knowledge of the population totals for each combination of sex, age, marital status and region. If the totals are known for each combination of sex, age, marital status and separately for each region, than complete stratification is not possible.

Incomplete multiway stratification offers a way out of the situations described above. If the highest order interactions are removed from the model and replaced by lower order interactions then in many cases the problems disappear. To describe an incomplete multiway stratification it is convenient to use a simple notational language. Complete multiway stratification by sex, age, marital status and region is denoted by SEX×AGE×MARITAL STATUS×REGION. In incomplete multiway stratification which uses the population totals for each combination of sex, age and marital status on the one hand and the population totals for each region on the other hand is denoted by (SEX×AGE×MARITAL STATUS)+REGION. The rows of the matrix X will in this last example not contain one 1, but two 1's. One set of dummy variables indicate the combination of sex, age and marital status, and another set of dummy variables denote the region. A consequence of such stratification designs is that the matrix $X_s'X_s$ can become singular. However, this singularity can easily be removed by deleting redundant columns from X.

Stratification comes down to estimation of the parameter vector β. The number of parameters to be estimated is smaller in incomplete stratification than in complete stratification. So incomplete stratification decreases estimation problems. For instance, if sex has 2 categories, age 10 categories, marital status 4 categories and region 11 categories, then SEX×AGE× MARITAL STATUS×REGION comes down to estimation of 880 parameters, whereas (SEX×AGE×MARITAL STATUS)+ REGION requires at most 91 parameters to be estimated. On the other hand the model behind the incomplete stratification might not fit as well as the model behind the complete stratification. However, we believe that in practical situations the incomplete stratification model is still based on so many parameters that this model fits nearly as well as the complete stratification model.

8. Examples

The theory was tested on census data of a Dutch municipality. Availability of data on the total population enabled us to compute variances. For the population two auxiliary variables are known: sex (2 categories) and age (5 categories). Complete multiway stratification (SEX× AGE) would result in 5×2=10 strata. In incomplete stratification (SEX+ AGE) there would in fact be two stratifications, one by sex and one by age. The design matrix has 6 columns. The first column indicates the constant term in the model, the second column indicates sex and the remaining four columns age. The matrix $X_s'X_s$ not only contains the sample sizes in the 10 possible combinations of sex and age but also the marginal sample totals of sex and age. In general $X_s'X_s$ contains interactions of order 1 higher than the matrix X_s. The population vector $\bar{x}$ only contains the 6 marginal means corresponding to the column means of X, i.e. unity, the proportion of of male, the proportion of people under 20, etc. The auxiliary information was used to estimate the number of people with a job, based on a simple random sample without replacement of size 100.

Tabel 8.1 contains the (approximated) variances of 5 different types of stratification. If possible, variances from the classical theory are also given. In case of normal stratification the differences between the variances are mainly created by leaving out terms of order 1/n in the regression estimator. When this term is also left out in the classical variance the resulting variance is very close to the variance of the regression estimator. It is obvious that with regard to the variance, incomplete multiway stratification takes a position between complete multiway stratification and oneway stratification.

Table 8.1. Variance of the estimator of the number of people with a job for several types of stratification (n=100)

stratification	variance of the estimator	classical variance
no stratification	34816	34816
SEX	30488	30799
AGE	24302	25389
SEX+AGE	20007	
SEX×AGE	17378	15819

We also illustrate our approache with a small example from sampling practice. The Dutch Criminal Victimization Survey 1982 estimates, among other things, the percentage of crime victims in the Netherlands. From one of the strata, the city of Rotterdam, a simple random sample was selected. The number of observations was 540. Three auxiliary variables were available for weighting purposes: age (11 categories), sex (2 categories) and marital status (2 categories). Complete multiway stratification (age×sex×marital status) was not possible due to a number of empty cells. The next best stratification was obtained by simultaneous consideration of crossings of two variables. Since age×marital status also produced some empty cells, incomplete stratification (AGE×SEX)+(SEX×MARITAL STATUS) was carried out. So only the population totals are needed for every combination of age and sex, and every combination of sex and marital status. Table 8.2 contains the results for a number of possible stratifications. As the stratification uses more auxiliary information the precision of the estimates increases. This was to be expected. However, there is also an increasing shift of the estimate. Weighting does not only increase the precision, but apparently is also able to reduce a bias caused by nonsampling errors such as non-response.

Table 8.2. The percentage of victims in Rotterdam in the Dutch Criminal Victimization Survey 1982

stratification	estimate	stimated standard error
no stratification	40.74	2.11
AGE	42.11	2.03
SEX	40.73	2.11
MARITAL STATUS	40.82	2.11
AGE×SEX	42.45	2.02
SEX×MARITAL STATUS	40.82	2.11
AGE+SEX×MARITAL STATUS	42.10	2.03
(AGE×SEX)+(SEX×MARITAL STATUS)	42.40	2.02

9. A computer program

At the moment the Netherlands Central Bureau of Statistics is implementing the theory of weighting. The program LINWEIGHT automatically selects a suitable stratification. To be able to do this, the program requires two data files: a file with sample data (target variables, auxiliary variables, selection weights) and a file with available population information. Furthermore the user must specify the sampling design. The program can handle stratified two-stage samples. Given the population data and the specified minimal number of observations per post-stratum the program selects a proper weighting scheme and computes estimates of population means and their estimateds standard errors.

The deck-setup of LINWEIGHT agrees to a large extend to the SPSS-conventions. This facilitates use of the program in an environment where SPSS is the most important data analysis package. An example of a deck-setup is given below.

```
RUN NAME          ESTIMATION OF THE MEAN
                     INCOME
SAMPLE FILE       PROV, MUN, MUNWGT, SEX, AGE,
                     INCOME, WGT
N OF CASES        10240
POPULATION FILE   VARIABLES=PROV, SEX, TOT /
                  FREQUENCY=TOT / CASES=22 /
                  VARIABLES=PROV, AGE, TOT /
                  FREQUENCY=TOT / CASES= 110
SAMPLING DESIGN   STRATUM=PROV /
                  CLUSTER=MUN /
                  CLUSTER SAMPLING=PSS /
                  CLUSTERWEIGHT=MUNWGT
                  SAMPLING=SRS / WEIGHT=WGT
STRATIFICATION    FILLING=15 /
                  AUXILIARIES=PROV(11),
                     SEX(2), AGE(10) /
                  TARGETS=INCOME
OPTIONS           1, 2, 3
READ INPUT DATA
FINISH
```

The RUN NAME card specifies an identifying text. The SAMPLE FILE card gives the names and order of appearances of the variables in sample file. N OF CASES gives the number of cases in the sample file. The POPULATION FILE describes the structure of the population information. In the above example apparently population totals are available for each combination of province and sex, and for each combination of province and age (and not for each combination of province, sex and age). The variable TOT contains the population total for each combination of values of the auxiliary variables in the population file. The SAMPLING DESIGN card specifies the sampling design. In this case a stratified two-stage sample is selected. Strata are identicated by the variable PROV, clusters by the variable MUN. Clusters are selected with unequal probabilities and the selection weights are contained in the variable MUNWGT. Within clusters elements are selected simple random without replacement and the selection weights are contained in the variable WGT. The STRATIFICATION card indicates how the weighting scheme must be selected: PROV, SEX and AGE are auxiliary variables (number of categories within brackets), and each post-stratum must contain at least 15 observations. For the target variable INCOME estimates of the mean and standard error are computed.

The program is written in PASCAL. It is still in development. Interested readers may contact the authors.

10. Other models

The optimal value of β in (3.3) is determined by ordinary least squares. This method assumes the order of magnitude of the residuals in ε to be roughly of the same magnitude. In cases where the order of magnitude of residuals differs in a known way, β can better be estimated by the method of generalized least squares:

$$\beta = (X'V^{-1}X)^{-1}XV^{-1}y \ , \qquad (10.1)$$

where V is an N×N-diagonal matrix of known constants such that the residuals in

$$y = X\beta + V^{\frac{1}{2}}\varepsilon \qquad (10.2)$$

are of roughly the same size. The theory of the regression estimator can be developed in the same way as the theory in the previous sections, where now

$$\hat{\beta} = (X'V^{-1}\Pi^{-1}TX)^{-1}X'V^{-1}\Pi^{-1}Ty \ . \qquad (10.3)$$

The ratio estimator is a special case of the generalized regression estimator. Using one auxiliary variable and assuming that the order of magnitude of the residuals is proportional to the square root of the value of the auxiliary variable, turns the estimator into the ratio estimator. Approximated variance and estimated variance agree to the results given in e.g. Cochran (1977).

Another approach to weighting based on linear models is given by Bethlehem and Keller (1982). If practical problems affect the inclusion probabilities, it is better to estimate the true inclusion probabilities on the basis of the sample. To that end a model is estimated which relates the inclusion probabilities to the available auxiliary information. An estimator of the population mean is now obtained by replacing the true, but unknown, inclusion probabilities in the usual Horvitz-Thompson estimator by the estimated inclusion probabilities.

The previous approaches are based on the use of linear models. There is no reason to restrict weighting to linear models. In fact, since weights should be non-negative, a multiplicative model might be more appropriate. Examples of the use of such models can be found e.g. in Chapman (1976) and Bailar et al. (1978). In the literature on sampling theory this method is usually called 'raking', but in other statistical literature the method based on the multiplicative model also appears under the names 'RAS-technique' and 'iterative proportional fitting'. Just as the concept of post-stratification can be extended to linear models for weighting, can the concept of raking be extended to loglinear models for weighting.

11. Conclusions

In this paper we proposed a general method for the computation of weighting schemes. The theory not only includes all standard estimation techniques, but also offers possibilities for solving the problem of empty strata or the problem of the lack of too detailed auxiliary information.

Since the weights are the result of application of the theory of linear models, techniques for selection of variable can be used to select the best possible stratification. Efficient application of these weighting methods in practice requires a computer program for the selection of the best auxiliary variables and the best model based on the selected auxiliary variables. If computation of weights can be performed automatically, time is saved and the cost of analysis of the sample survey data is reduced.

The theory is developed for treatment of one target variable. The extension to estimation of the mean of a vector of target variables is straightforward. Instead of a variance of the estimate it will produce a covariance matrix of the estimates. This extension is particularly useful if the sample data is used for multivariate analysis.

References

Bailar, B.A., L. Bailey and C. Corby, 1978, A comparison of some adjustment and weighting procedures for survey data. In: N. Krishnan Namboodiri, ed., Survey sampling and measurement (Academic Press, Cambridge).

Bethlehem, J.G. and W.J. Keller, 1982, Linear models for weighting of sample survey data. Internal CBS-report (Netherlands Central Bureau of Statistics, Voorburg).

Bethlehem, J.G. and W.J. Keller, 1983, Weighting sample survey data using linear models. Internal CBS-report (Netherlands Central Bureau of Statistics, Voorburg).

Chapman, D.W., 1976, A survey of non-response imputation procedures. Proceedings of the American Statistical Association, Social Statistic Section, pp. 245-251.

Cochran, W.G., 1977, Sampling techniques, 3rd edition (Wiley, New York).

Holt, D. and T.M.F. Smith, 1979, Post stratification. Journal of the Royal Statistical Society A 142, pp. 33-36.

Horvitz, D.G. and D.J. Thompson, 1952, A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association 47, 663-685.

Lindström, H., J. Wretman, G. Forsman and C. Cassel, 1979, Standard methods for nonresponse treatment in statistical estimation (National Central Bureau of Statistics, Stockholm).

Platek, R. and G.B. Gray, 1980, Imputation methodology (total survey error) (Statistics Canada, Ottawa).