THE UNIVERSITY OF MANITOBA

MULTI-PHASE SAMPLING FOR STRATIFICATION AND TWO-STAGE SAMPLING

by

V. Ramachandran

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

Department of Statistics

WINNIPEG, MANITOBA

May 1973

ACKNOWLEDGEMENT

I take great pleasure in expressing my deep sense of gratitude to Dr. J. N. K. Rao, Department of Statistics, for his most valuable guidance and constructive criticism throughout the period of this study and during the writing of this thesis. I am also thankful to Dr. S. K. Sinha and Dr. R. J. MacKay, Department of Statistics, and Dr. Dhuruvarajan, Department of Economics, who examined this thesis extensively and recommended very helpful suggestions as to style and content.

ABSTRACT

In the present thesis we first review some of the major aspects of double sampling for stratification and two-stage sampling. We propose a simple method of three-phase sampling for stratification which is an extension to double sampling for stratification. Using mathematical programming, an algorithm is developed for the determination of optimum allocation of available resources. The subsampling of non-respondents with three attempts is obtained as a special case of our three-phase sampling scheme for stratification. An alternative method of double sampling for two-stage sampling is proposed and our method leads to solutions for the determination of optimal design.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION AND REVIEW

The purpose of statistical surveys is to obtain information about populations. By 'population' we understand a group of units defined according to the aims of a survey. A sampling method is a scientific and objective procedure of selecting units from the population and provides a sample that is expected to be representative of the population as a whole. A random sampling method makes it possible to estimate the population totals, averages or proportions while reducing at the same time the size of survey operations. Survey sampling literature describes several methods of using certain auxiliary information to attain more efficient sampling and estimation procedures.

If auxiliary information is not available but can be collected rather inexpensively on a somewhat large scale, it may pay to collect such information in the first instance and then take a sample for the measurement of y. This method of sampling is known as Double Sampling and the theory was first given by Neyman (1938). In double sampling for stratification the first sample is used in grouping the selected units into a temporary sampling frame divided into strata.

A preliminary simple random sample $s(n')$ of size n' is selected without replacement from N units of the population and the auxiliary character x alone is observed. The selected units are classified into L strata according to x. Let n_h' denote the number of units in $s(n')$ falling into stratum h ($h = 1, 2, \ldots, L$; $\sum_{h=1}^{L} n_h' = n'$). A sub-sample $s(n_h)$ of size n_h is drawn from $s(n_h')$ by simple random sampling (srs) without replacement for each h, and the character of interest, y, is observed.

In the literature (e.g., Cochran, 1963, pp.328-34), n_h is often implicitly assumed to be a constant and optimal n' and n_h are determined. This assumption, however, is inconsistent with the sampling procedure because, as pointed out by Singh and Singh (1965), n_h is bounded above by the random variable n_h' which varies from 0 to $\min(n', N_h)$, where N_h is the total number of units in stratum h. Singh and Singh proposed three procedures which are free of inconsistency:

(i) the $s(n_h)$ are selected with replacement and all units used in the estimator;

(ii) as (i), but with only distinct units used;

(iii) sub-sampling is without replacement, the size being $\min(n_h', n_h)$.

Procedures (i) and (ii), however, lead to loss in efficiency and (iii) results in an unwieldy variance formula which is not suitable for the determination of optimal n' and n_h. Rao (1973) proposed a method which is simple, free from the inconsistency and leads to a simple variance formula. Essentially, Rao's method employs without replacement sampling with $n_h = \nu_h n_h'$ where ν_h is a constant such that $0 < \nu_h \leq 1$, and with determination of optimal n' and ν_h. This method leads to simple solutions for the optimal design of analytical surveys involving comparison of group means, when the groups are not identifiable in advance.

A possible extension of double sampling for stratification would be Three-Phase sampling for double stratification. Robson and King (1952) proposed such a scheme and utilized it in the estimation of magazine readership. The scheme can be summarized as follows. The elements of a finite population of size N are characterized by three variables X, Z and Y. It is supposed that the character Y is more difficult to measure than Z which, in turn, is more difficult to measure than X. We take a large sample s(n') of size n' by simple random sampling without replacement and observe x alone. The selected units are classified into L strata according to x. Let n'_h denote the number of units in s(n') falling into stratum h ($h = 1, 2, \ldots, L$; $\sum_{h=1}^{L} n'_h = n'$). A subsample of size n_h is drawn from $s(n'_h)$ by simple random sampling without replacement independently for each h and the character z is observed. The selected units are then sub-classified into k_h strata according to z. Let n_{hi} denote the number of units in $s(n_h)$ falling into the i^{th} substratum of the h^{th} primary stratum ($i = 1, 2, \ldots, k_h$, $h = 1, 2, \ldots, L$, $\sum_{i=1}^{k_h} n_{hi} = n_h$, $\sum_{h=1}^{L} n_h = n$). A subsample $s(m_{hi})$ of size m_{hi} is drawn from $s(n_{hi})$ by simple random sampling without replacement independently for each (h, i) and the character of interest y is observed.

Our method in Chapter 2 is essentially an extension of Rao's method with

$$n_h = \nu_h n'_h \qquad 0 < \nu_h \leq 1,\ h = 1, 2, \ldots, L;$$

$$m_{hi} = \lambda_{hi} n_{hi} \qquad 0 < \lambda_{hi} \leq 1,\ i = 1, 2, \ldots, k_h,\ h = 1, 2, \ldots, L,$$

where ν_h and λ_{hi} are constants, with the determination of optimal n', ν_h and λ_{hi}. The method of Robson and King employs

$$n_h = n'_h \frac{n}{n'} \qquad h = 1, 2, \ldots, L$$

and

$$m_{hi} = n_{hi} \frac{m}{n} \qquad i = 1, 2, \ldots, k_h,\ h = 1, 2, \ldots, L$$

which clearly is a special case of our general scheme by putting $\nu_h = \frac{n}{n'}$ and $\lambda_{hi} = \frac{m}{n}$.

Hansen and Hurwitz (1946) were the first to deal with the problem of incomplete samples in mail surveys. In many of these surveys the response rate is very low and many doubts are normally associated with estimates resulting therefrom. In the classical non-response theory, a sample of size n', selected by simple random sampling without replacement, is contacted by ordinary field methods. A sub-sample of size

$$n_2 = n'_2/k \qquad (k \geq 1),$$

from the n'_2 $(= n' - n_1)$ nonrespondents, again selected by simple random sampling without replacement, is enumerated by more intensive efforts. Rao (1973) has shown that the procedure is a special case of double sampling method for stratification with $L = 2$, $\nu_1 = 1$ and $\nu_2 = 1/k$.

The subsampling of nonrespondents with three attempts is a special case of our 3-phase sampling scheme for stratification. A sample of size n' is selected by simple random sampling without replacement. A subsample of size $n_2 = n_2'/k$ $(k \geq 1)$ from the n_2' $(= n' - n_1')$ nonrespondents again selected by simple random sampling without replacement, is contacted by more intensive efforts. Once again a subsample of size $m_{22} = n_{22}/\ell$ $(\ell \geq 1)$ from the n_{22} $(= n_2 - n_{21})$ nonrespondents, selected by simple random sampling without replacement, is enumerated with serious efforts. This procedure is clearly a special case of our three-phase sampling method for stratification with $L = 2$, $\nu_1 = 1$, $\nu_2 = \frac{1}{k}$, $\lambda_{11} = \lambda_{12} = \lambda_{21} = 1$ and $\lambda_{22} = \frac{1}{\ell}$.

A difficulty often encountered in cluster or two-stage sampling is an inadequate knowledge of the composition of the clusters or primary units. Dalenius (1957) describes a situation of this kind that occurred in a Swedish survey of car-owners carried out by the Central Bureau of Statistics in 1954. The car registers on which the selection was to be based were not arranged by community but by licence number, and for reasons of economy, it was considered highly desirable to use two-stage sampling. Among the procedures considered was double sampling for two-stage sampling, the basic idea of which was to form a temporary sampling frame based on a relatively large preliminary sample in which car owners were identified as to community. The second sampling phase consisted of two stage sampling with communities as primary sampling units.

We take a large sample $s(n')$ of size n' by simple random sampling without replacement and observe x alone. The selected units are classified into L' clusters according to x. Let n'_h denote the number of units in $s(n')$ falling into cluster h $(h = 1, 2, \ldots, L', \sum_{h=1}^{L'} n'_h = n')$. From the L' clusters, we select a sample $s(k)$ of size k with probability proportional to $\frac{n'_h}{h'}$ (with replacement). A sub-sample $s(n_h)$ of size n_h is drawn from $s(n'_h)$ by srs independently each time a primary unit h is drawn in sample. Berg (1972) proposed the selection of n_h units from $s(n'_h)$ by simple random sampling with replacement. To avoid the loss in efficiency due to sampling with replacement, we consider

$$n_h = n'_h \nu_h \qquad (0 < \nu_h < 1), \; h = 1, 2, \ldots, L$$

and select $s(n_h)$ by srs without replacement. The optimal n', k and ν_h are obtained in Chapter 3.

CHAPTER II

2.1 THREE-PHASE SAMPLING FOR STRATIFICATION

We make the customary assumption that n' is so large that $\Pr(n_h' = 0) = 0$ for all h. In the second-phase sampling similarly we assume that n_h $(h = 1, 2, \ldots, L)$ is so large that $\Pr(n_{hi} = 0) = 0$ $(i = 1, 2, \ldots, k_h,\ h = 1, 2, \ldots, L)$. The following symbols refer to stratum h: $W_h = \frac{N_h}{N}$, $w_h = \frac{n_h}{n'}$, $\bar{y}_h$ is the mean for $s(n_h)$, $\bar{Y}_h$ is the population mean, S_h^2 is the population variance, $S_h'^2$ is the unbiased estimator of S_h^2 based on $s(n_h')$. The following symbols refer to the i^{th} secondary stratum of the h^{th} primary stratum: $W_{hi} = \frac{N_{hi}}{N}$, $w_{hi}^* = \frac{n_{hi}}{n_h}$, $\bar{y}_{hi}$ is the mean for $s(m_{hi})$, $\bar{y}_{hi}^*$ is the mean based on $s(n_{hi})$, $\bar{Y}_{hi}$ is the population mean, S_{hi}^2 is the population variance, s_{hi}^{*2} is the unbiased estimator of S_{hi}^2 based on $s(n_{hi})$, $s_{hi}'^2$ is the unbiased estimator of S_{hi}^2 based on the n_{hi}' non-observable units.

THEOREM 1

The estimator

$$\bar{y}_{3p} = \sum_{h=1}^{L} \sum_{i=1}^{k_h} \frac{n_h'}{n'} \frac{n_{hi}}{n_h} \bar{y}_{hi}$$

$$= \sum_{h=1}^{L} \sum_{i=1}^{k_h} w_h w_{hi}^* \bar{y}_{hi} \qquad (2.1.1)$$

is unbiased for

$$\bar{Y} = \sum_{h=1}^{L} \sum_{i=1}^{k_h} W_{hi} \bar{Y}_{hi}$$

with variance

$$V(\bar{y}_{3p}) = (\frac{1}{n'} - \frac{1}{N})S^2 + \sum_{h=1}^{L} W_h \frac{S_h^2}{n'} (\frac{1}{\nu_h} - 1)$$

$$+ \sum_{h=1}^{L} \sum_{i=1}^{k_h} \frac{W_{hi}S_{hi}^2}{n'\nu_h} (\frac{1}{\lambda_{hi}} - 1) \qquad (2.1.2)$$

where '3p' denotes 'three-phase'.

Proof: For a given $\underset{\sim}{n}' = (n_1', n_2', \ldots, n_L')$ and $s(n')$, $s(n_h)$ is a simple random sample from $s(n_h')$ in stratum h. Let E_2 and V_2 respectively denote the expectation and variance operators for a give $\underset{\sim}{n}'$ and $s(n')$. Similarly, for a given $\underset{\sim}{n}_h = (n_{h1}, \ldots, n_{hk_h})$ and $s(n_h)$ in h^{th} stratum, $s(m_{hi})$ is a simple random sample from $s(n_{hi})$ in the i^{th} secondary stratum of the h^{th} primary stratum. Consequently, we denote E_3 and V_3 respectively as the expectation and variance operators for a given $\underset{\sim}{n}'$ and $\underset{\sim}{n}_h$. Let E_1 and V_1 be the unconditional expectation and variance operators for $s(n')$. We have

$$E(\bar{y}_{3p}) = E_1E_2E_3(\bar{y}_3)$$

$$E_3(\bar{y}_{hi}) = \bar{y}_{hi}^*$$

$$E_2E_3(\bar{y}_{3p}) = E_2 \sum_{h=1}^{L} \frac{n_h'}{n'} \bar{y}_h = \sum_{h=1}^{L} \frac{n_h'}{n'} \bar{y}_h' = \bar{y}^1$$

and

$$E_1E_2E_3(\bar{y}_{3p}) = E_1(\bar{y}') = \bar{Y}.$$

Similarly,

$$V(\bar{y}_{3p}) = V_1E_2E_3(\bar{y}_{3p}) + E_1V_2E_3(\bar{y}_{3p}) + E_1E_2V_3(\bar{y}_{3p}) \qquad (2.1.3)$$

$$V_3(\bar{y}_{hi}) = (\frac{1}{m_{hi}} - \frac{1}{n_{hi}})s_{hi}^{*2}$$

we get

$$V(\bar{y}_{3p}) = V_1E_2\left\{\sum_{h=1}^{L} \frac{n_h'}{n'}\bar{y}_h\right\} + E_1V_2\left\{\sum_{h=1}^{L} \frac{n_h'}{n'}\bar{y}_h\right\}$$

$$+ E_1E_2\left\{\sum_{h=1}^{L}\sum_{i=1}^{k_h} w_h^2 w_{hi}^{*2}\left(\frac{1}{m_{hi}} - \frac{1}{n_{hi}}\right)s_{hi}^{*2}\right\} \qquad (2.1.4)$$

Now using the relationship

$$m_{hi} = \lambda_{hi}n_{hi},$$

$$V_2(\bar{y}_h) = \left(\frac{1}{n_h} - \frac{1}{n_h'}\right)s_h'^2$$

and simplifying, (2.1.4) becomes

$$V(\bar{y}_{3p}) = V_1(\bar{y}^1) + E_1 \sum_{h=1}^{L}\sum_{i=1}^{k_h} w_h^2\left(\frac{1}{nh} - \frac{1}{nh'}\right)s_h'^2$$

$$+ E_1 \sum_{h=1}^{L}\sum_{i=1}^{k_h} w_h^2 \frac{n_{hi}'}{n'n_h} s_{hi}'^2\left(\frac{1}{\lambda_{hi}} - 1\right). \qquad (2.1.5)$$

Finally, letting

$$n_h = \nu_h n_h'$$

and simplifying (2.1.5) leads to (2.1.2).

Corollary 1

For Double Sampling for stratification

$$\lambda_{hi} = 1,\ i = 1, 2, \ldots, k_h,\ h = 1, 2, \ldots, L$$

and (2.1.2) reduces to Rao's (1973) formula.

Corollary 2

For porportional allocation

$$\nu_h = \frac{n}{n'} \qquad h = 1, 2, \ldots, L$$

$$\lambda_{hi} = \frac{m}{n} \qquad i = 1, 2, \ldots, k_h,\ h = 1, 2, \ldots, L$$

(2.1.2) reduces to

$$V(\bar{y}_{3p}) = (\frac{1}{n'} - \frac{1}{N})S^2 + \frac{(n'-n)}{n} \sum_{h=1}^{L} W_h S_h^2$$

$$+ \frac{(n-m)}{nm} \sum_{h=1}^{L} \sum_{i=1}^{k_h} W_{hi} S_{hi}^2 . \qquad (2.1.6)$$

Corollary 3

For estimating the population proportion P possessing characteristic Y (that is, y = 0 or 1) with proportional allocation,

$$\nu_h = \frac{n}{n'} , \qquad \lambda_{hi} = \frac{m}{n}$$

$$\bar{Y} = P, \qquad \bar{Y}_h = P_h, \qquad \bar{Y}_{hi} = P_{hi}$$

($i = 1, 2, \ldots, k_h$, $h = 1, 2, \ldots, L$), and (2.1.2) reduces to

$$V_p(\bar{y}_{3p}) = \frac{N-n'}{n'(N-1)} P(1-P) + \frac{(n'-n)}{n} \sum_{h=1}^{L} W_h \frac{N_h}{(N_h-1)} P_h(1-P_h)$$

$$+ \frac{(n-m)}{nm} \sum_{h=1}^{L} \sum_{i=1}^{k_h} W_{hi} \frac{N_{hi}}{(N_{hi}-1)} P_{hi}(1-P_{hi}) \qquad (2.1.7)$$

which agrees with the formula of Robson and King (1952).

Optimum Allocation

We turn now to the determination of optimal n', $\underline{\nu} = (\nu_1, \nu_2, \ldots, \nu_L)$ and $\underline{\lambda} = (\lambda_{11}, \ldots, \lambda_{1k_1}; \ldots; \lambda_{L_1}, \ldots, \lambda_{Lk_L})$. The cost function is taken as

$$C = n'c' + \sum_{h=1}^{L} n_h c_h + \sum_{h=1}^{L} \sum_{i=1}^{k_h} m_{hi} c_{hi} \qquad (2.1.8)$$

where c' is smaller than c_h which in turn is smaller than c_{hi}. Since C in (2.1.8) is random, we take the expected cost

$$C^* = n'c' + \sum_{h=1}^{L} E(n_h)c_h + \sum_{h=1}^{L} \sum_{i=1}^{k_h} E(m_{hi})c_{hi}$$

$$= n'c' + n' \sum_{h=1}^{L} W_h c_h \nu_h + n' \sum_{h=1}^{L} \sum_{i=1}^{k_h} W_{hi} c_{hi} \nu_h \lambda_{hi} . \qquad (2.1.9)$$

Raó (1973) has shown that the the coefficient of variation (C.V.) of C would be very small for large n'. (2.1.2) may be written as

$$V(\bar{y}_{3p}) = -\frac{S^2}{N} + \frac{B}{n'} + \sum_{h=1}^{L} \frac{W_h A_h}{m_h} + \sum_{h=1}^{L} \sum_{i=1}^{k_h} \frac{W_{hi} S_{hi}^2}{\ell_{hi}} \qquad (2.1.10)$$

where

$$m_h = n'\nu_h \qquad 0 < m_h \leq n', \; h = 1, 2, \ldots, L$$

$$\ell_{hi} = m_h \lambda_{hi} \qquad 0 < \ell_{hi} \leq m_h, \; i = 1, 2, \ldots, k_h, \; h = 1, 2, \ldots, L$$

$$= n'\nu_h \lambda_{hi}$$

$$B = S^2 - \sum_{h=1}^{L} W_h S_h^2$$

and

$$A_h = S_h^2 - \sum_{i=1}^{k_h} \frac{W_{hi}}{W_h} S_{hi}^2 \qquad h = 1, 2, \ldots, L.$$

B and A_h are > 0 except possibly in pathological situations. Also

$$C^* = n'c' + \sum_{h=1}^{L} W_h c_h m_h + \sum_{h=1}^{L} \sum_{i=1}^{k_h} W_{hi} c_{hi} \ell_{hi}. \qquad (2.1.11)$$

We need to minimize (2.1.10) subject to (2.1.11) and the inequality constraints $0 < m_h \leq n'$, $0 \leq \ell_{hi} \leq m_h$ ($i = 1, 2, \ldots, k_h$, $h = 1, 2, \ldots, L$) and $n' > 0$. Minimization of a non-linear function subject to inequality constraints is called Mathematical Programming. A relative minimum would be a global minimum if the function to be minimized is a convex function and the Kuhn-Thucker (1951) necessary conditions are satisfied.

Using the results of Kuhn and Tucker, Thompson (1962) obtained the following theorem.

THEOREM 2

Minimization of a differentiable, strictly convex function $g(\underset{\sim}{w})$ subject to $A^{-1}\underset{\sim}{w} \geq 0$ leads to the solution $\hat{\underset{\sim}{w}}$ where

$$A^{-1}\hat{\underset{\sim}{w}} \geq 0$$

$$A'G(\hat{\underset{\sim}{w}}) \geq 0$$

and

$$\hat{\underset{\sim}{w}}'G(\hat{\underset{\sim}{w}}) = 0.$$

Equivalently, either

$$\underset{\sim}{b}_i'\hat{\underset{\sim}{w}} = 0 \quad \text{and} \quad \underset{\sim}{a}_i'G(\underset{\sim}{w}) = 0$$

or

$$\underset{\sim}{b}_i'\hat{\underset{\sim}{w}} > 0 \quad \text{and} \quad \underset{\sim}{a}_i'G(\underset{\sim}{w}') = 0 \qquad i = 1, 2, \ldots, p$$

where $\underset{\sim}{w} = (w_1, w_2, \ldots, w_p)$.

$G(\underset{\sim}{w})$ is the vector of derivatives $\frac{\partial g(\underset{\sim}{w})}{\partial w_i}$, $i = 1, 2, \ldots, p$; $\underset{\sim}{b}_i'$ is the i^{th} row of A^{-1} and a_i' is the i^{th} row of A' and A is a non-singular p×p matrix.

Thompson's algorithm applies to our problem by noting

$$(n' - m_h) \geq 0 \qquad h = 1, 2, \ldots, L$$

$$(m_h - \ell_{hi}) \geq 0 \qquad i = 1, 2, \ldots, k_h,\ h = 1, 2, \ldots, L$$

and letting

$$\underset{\sim}{w}' = (n', m_1, \ell_{11}, \ell_{12}, \ldots, \ell_{1k_1}, \ldots, m_L, \ell_{L_1}, \ldots, \ell_{Lk_L})$$

$$g = -\frac{1}{N} + \frac{B}{n'} + \sum_{h=1}^{L} \frac{W_h A_h}{m_h} + \sum_{h=1}^{L} \sum_{i=1}^{k_h} \frac{W_{hi} S_{hi}^2}{\ell_{hi}}$$

$$+ \mu(c'n' + \sum_{h=1}^{L} W_h c_h m_h + \sum_{h=1}^{L} \sum_{i=1}^{k_h} W_{hi} c_{hi} \ell_{hi} - C^*). \qquad (2.1.12)$$

Also

$$A^{-1} = \begin{bmatrix} T_{11} & T_{12} & \cdots & T_{1L} \\ T_{21} & T_{22} & \cdots & T_{2L} \\ \cdots & \cdots & \cdots & \cdots \\ T_{L1} & T_{L2} & \cdots & T_{LL} \end{bmatrix}$$

where T_{ij} $(i = 1, 2,\ldots,L,\ j = 1, 2,\ldots,L)$ are partitioned matrices defined by

$$T_{11} = \begin{bmatrix} 1 & 0 & \cdots & & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 \cdots & 0 \\ 0 & 1 & 0 & -1 \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & \cdots & \cdots & -1 \end{bmatrix}_{(k_1+2)\times(k_1+2)}$$

$$T_{ii} = \begin{bmatrix} -1 & 0 & \cdots & 0 \\ 1 & -1 & \cdots & 0 \\ 1 & 0 & -1 \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & \cdots & -1 \end{bmatrix}_{(k_i+1)\times(k_i+1)} \qquad i = 2, 3,\ldots,L$$

$$T_{i1} = \begin{bmatrix} 1 & 0 \cdots 0 \\ 0 & 0 \cdots 0 \\ 0 \cdots & 0 \\ \cdots & \cdots \\ 0 \cdots & 0 \end{bmatrix}_{(k_i+1)\times(k_1+2)} \qquad i = 2, 3,\ldots,L$$

$$T_{ij} = 0_{(k_i+2)x(k_j+1)}$$

$$T_{ij} = 0_{(k_i+1)(k_j+1)} \quad \text{otherwise}$$

where 0 is a zero matrix of the appropriate order.

Similarly,

$$A' = \begin{bmatrix} R_{11} & R_{12} \cdots\cdots R_{1L} \\ R_{21} & R_{22} \cdots\cdots R_{2L} \\ \cdots & \cdots\cdots\cdots \\ R_{L1} & R_{L2} \qquad R_{LL} \end{bmatrix}$$

where R_{ij} ($i = 1, 2,\ldots,L$, $j = 1, 2,\ldots,L$) are partitioned matrices defined by

$$R_{11} = \begin{bmatrix} 1 & 1 \cdots\cdots\cdots\cdots 0 \\ 0 & -1 \cdots\cdots\cdots\cdots 0 \\ 0 & 0 \quad -1 \quad 0 \cdots 0 \\ \cdots & \cdots\cdots\cdots\cdots\cdots \\ 0 & 0 \cdots\cdots\cdots\cdots -1 \end{bmatrix}_{(k_1+2) \times (k_1+2)}$$

$$R_{ii} = \begin{bmatrix} -1 & -1 \cdots\cdots\cdots\cdots 0 \\ 0 & -1 \cdots\cdots\cdots\cdots 0 \\ 0 & 0 \quad -1 \quad 0 \cdots 0 \\ \cdots & \cdots\cdots\cdots\cdots\cdots \\ 0 & 0 \cdots\cdots\cdots\cdots -1 \end{bmatrix}_{(k_i+1) \times (k_i+1)} \qquad R_{ij} = \begin{bmatrix} 1 & 1 \quad 1 \cdots 1 \\ 0 & 0 \quad 0 \cdots 0 \\ \cdots & \cdots\cdots\cdots \\ 0 & 0 \cdots\cdots\cdots 0 \end{bmatrix}_{(k_i+2) \times (k_j+1)} \quad j = 2, 3,\ldots,L$$

$$R_{i1} = 0_{(k_i+1)x(k_1+2)}, \quad i = 2, 3,\ldots,L$$

$$R_{ij} = 0_{(k_i+1)(k_j+1)} \quad \text{otherwise}$$

where 0 is a zero matrix of the appropriate order.

Details of the various cases with the necessary conditions to be satisfied are given in the Appendix. Here we outline the algorithm for the allocation of λ_{hi} and ν_h. Let us denote the j^{th} largest of $\frac{S_{hi}}{c_{hi}}$ in a stratum h as $\frac{S_{h(j)}}{c_{h(j)}}$; that is,

$$\frac{S_{h(1)}}{c_{h(1)}} \geq \frac{S_{h(2)}}{c_{h(2)}} \geq \cdots \geq \frac{S_{h(k_h)}}{c_{h(k_h)}} \qquad h = 1, 2,\ldots,L.$$

If $\frac{S_{h(1)}}{c_{h(1)}} < \frac{A_h}{c_h}$ (h = 1, 2,...,L), then arrange $\frac{A_h}{c_h}$ in decreasing order of magnitude as follows:

$$\frac{A_{(1)}}{c_{(1)}} \geq \frac{A_{(2)}}{c_{(2)}} \geq \ldots . \geq \frac{A_{(L)}}{c_{(L)}}.$$

From the conditions worked out in the Appendix for getting the optimum values of λ_{hi} and $\nu_{j(i)}$ $(h \neq j)$; we see that the allocations are independent of each other as we normally expect. The optimum values of λ_{hi} and ν_h are given for different regions in which $\frac{S_{hi}}{c_{hi}}$ and $\frac{A_h}{c_h}$ lie.

<u>1.</u> $0 < \frac{S^2_{h(1)}}{c_{h(1)}} \leq \frac{A_h}{c_h}$ $\qquad i = 1, 2,\ldots,k_h;\ h = 1, 2,\ldots,L.$

$$\lambda_{hi} = \sqrt{\frac{S^2_{hi}}{c_{hi}} \frac{c_h}{A_h}} \qquad i = 1, 2,\ldots,k_h;\ h = 1, 2,\ldots,L \tag{2.1.13}$$

In this region, it is clear that the allocations of ν_h are independent from the allocations of λ_{hi}. Here we have four possible cases.

<u>1.(a)</u> $0 < \frac{A_{(1)}}{c_{(1)}} \leq \frac{B}{c'}$

$$\nu_h = \sqrt{\frac{A_h}{c_h} \frac{c'}{B}} \qquad h = 1, 2,\ldots,L \tag{2.1.14}$$

<u>1.(b)</u> $\frac{A_{(2)}}{c_{(2)}} \leq \frac{B+W_{(1)}A_{(1)}}{c'+W_{(1)}c_{(1)}} < \frac{A_{(1)}}{c_{(1)}}$

$$\left.\begin{aligned} \nu_{(1)} &= 1 \\ \nu_{(h)} &= \sqrt{\frac{A_h}{ch} \frac{c'+W_{(1)}c_{(1)}}{B+W_{(1)}S^2_{(1)}}} \qquad h = 1, 2,\ldots,L. \end{aligned}\right\} \tag{2.1.15}$$

This means that if $\frac{A_{(1)}}{c_{(1)}} > \frac{B}{c'}$

and $\frac{A_{(2)}}{c_{(2)}} > \frac{B}{c'}$

then $\nu_{(1)} = 1$ and $\nu_{(2)} = 1$ if and only if

$$\frac{A_{(2)}}{c_{(2)}} \geq \frac{B+W_{(1)}A_{(1)}}{c'+W_{(1)}c_{(1)}} . \tag{2.1.16}$$

By intuition we would try to push both $\nu_{(1)} = \nu_{(2)} = 1$ without testing (2.1.16) and if (2.1.16) is not satisfied, this does not give the optimum solution as shown in the example at the end of this chapter.

1.(c) $$\frac{A_{(r+1)}}{c_{(r+1)}} \leq \frac{B+W_{(1)}A_{(1)}+\ldots+W_{(r)}A_{(r)}}{c'+W_{(1)}c_{(1)}+\ldots+W_{(r)}c_{(r)}} < \frac{A_{(L)}}{c_{(L)}}$$

$$\left.\begin{aligned} &\nu_{(1)} = \ldots\ldots = \nu_{(r)} = 1 \\ &\nu_{(h)} = \sqrt{\frac{A_{(h)}}{c_{(h)}} \frac{c'+W_{(1)}c_{(1)} + \ldots + W_{(r)}c_{(r)}}{B+W_{(1)}A_{(1)} + \ldots + W_{(r)}A_{(r)}}} \end{aligned}\right\} \tag{2.1.17}$$

$$h = r+1,\ldots,L$$

1.(d) $$\frac{B+W_{(1)}A_{(1)} + \ldots + W_{(L)}A_{(L)}}{c'+W_{(1)}c_{(1)} + \ldots + W_{(L)}c_{(L)}} < \frac{A_{(L)}}{c_{(L)}}$$

$$\nu_{(1)} = \nu_{(2)} = \ldots\ldots = \nu_{(L)} = 1. \tag{2.1.18}$$

2. When one or more λ_{hi} in one or more substratums are pushed to 1, then corresponding allocation for ν_h are affected.

$$\frac{S^2_{si(2)}}{c_{si(2)}} \leq \frac{A_{si}+W_{si(1)}S^2_{si(1)}}{c_{si}+W_{si(1)}c_{si(1)}} < \frac{S^2_{si(1)}}{c^2_{si(1)}}$$

$$1 \le s_1,\ s_2, \ldots, s_r \le L$$

$$s_1 \ne s_2 \ne \ldots \ne s_r$$

$$i = 1,\ 2, \ldots, r.$$

$$\frac{S^2_{h(1)}}{c_{h(1)}} < \frac{A_{(h)}}{c_{(h)}} \qquad h = 1,\ 2, \ldots, L;\ h \ne s_1,\ s_2, \ldots, s_r$$

$$\left.\begin{aligned} &\lambda_{si(1)} = 1 \qquad i = 1,\ 2, \ldots, r \\ &\lambda_{si(j)} = \sqrt{\frac{S^2_{si(j)}}{c_{si(j)}} \frac{c_{si} + W_{si(1)}c_{si(1)}}{A_{si} + W_{si(1)}S^2_{si(1)}}} \\ &\qquad j = 2,\ 3, \ldots, k_{si};\ i = 1,\ 2, \ldots, r\ (\le L) \\ &\lambda_{h(i)} = \sqrt{\frac{S^2_{h(i)}}{c_{h(i)}} \frac{c_h}{A_h}} \qquad h = 1,\ 2, \ldots, L:\ h \ne s_1,\ s_2, \ldots, s_r; \\ &\qquad i = 1,\ 2, \ldots, k_h. \end{aligned}\right\} \quad (2.1.19)$$

Now the ν_h are no longer independent of λ_{hi}. Let us define

$$\frac{A^*_h}{c^*_h} = \frac{A_h + \delta_h W_{h(1)} S^2_{h(1)}}{c_h + \delta_h W_{h(1)} c_{h(1)}} \qquad h = 1,\ 2, \ldots, L$$

where

$$\left.\begin{aligned} \delta_h &= 1 \quad \text{if } h = s_1,\ s_2, \ldots, s_r \\ &= 0 \quad \text{otherwise} \end{aligned}\right\} \qquad (2.1.20)$$

Order the $\frac{A^*_h}{c^*_h}$ in decreasing order of magnitude and repeat the allocation we discussed earlier in 1.(a), 1.(b), 1.(c) and 1.(d) by replacing A_h by A^*_h and c_h by c^*_h.

3. $$\frac{S^2_{si(j+1)}}{c_{si(j+1)}} \le \frac{W_{si}A_{si} + \sum_{j=1}^{q_i} W_{si(j)} S^2_{si(j)}}{W_{si}c_{si} + \sum_{j=1}^{u_i} W_{si(j)} c_{si(j)}} < \frac{S^2_{si(j)}}{c_{si(j)}}$$

$j = 1, 2, \ldots, q_i$; $i = 1, 2, \ldots, r \leq L$; $1 \leq s_1 \leq \ldots \leq s_r \leq L$;

$s_1 \neq \ldots \neq s_r$; $1 \leq q_i \leq k_{si}$,

$$\left.\begin{array}{ll}
\dfrac{S^2_{h(1)}}{c_{h(1)}} < \dfrac{A_h}{c_h} & h = 1, 2, \ldots, L;\ h \neq s_1, s_2, \ldots, s_r \\[2ex]
\lambda_{si(j)} = 1 & j = 1, 2, \ldots, q_i,\ i = 1, 2, \ldots, r \leq L, \\
 & 1 \leq q_i \leq k_{si} \\[2ex]
\lambda_{si(k)} = \sqrt{\dfrac{S^2_{si(k)}}{c_{si(k)}} \dfrac{W_{si}c_{si} + \sum\limits_{j=1}^{u_i} W_{si(j)}c_{si(j)}}{W_{si}c_{si} + \sum\limits_{j=1}^{u_i} W_{si(j)}S^2_{si(j)}}} & \\[2ex]
\quad k = q_i + 1, \ldots, k_{si};\ i = 1, 2, \ldots, r & \\[2ex]
\lambda_{h(i)} = \sqrt{\dfrac{S^2_{h(i)}}{c_{h(i)}} \dfrac{c_h}{A_h}} & h = 1, 2, \ldots, L;\ h \neq s_1, s_2, \ldots, s_r
\end{array}\right\} \quad (2.1.21)$$

Once again define

$$\frac{A^*_h}{c^*_h} = \frac{W_h A_h + \delta_h \sum\limits_{j=1}^{q_i} W_{h(j)} S^2_{h(j)}}{W_h c_h + \delta_h \sum\limits_{j=1}^{q_i} W_{h(j)} c_{h(j)}} \qquad h = 1, 2, \ldots, L,$$

where δ_h is defined by (2.1.20). The allocation of ν_h would be on the same lines as in 1.(a), 1.(b), 1.(c) and 1.(d) by replacing A_h by A^*_h and c_h by c^*_h.

If the strata weights W_h, W_{hi} are not accurately known, we propose a different subsampling procedure which guarantees desired precision at a slightly higher cost than the optimum.

Following Srinath (1971), we choose

$$\left.\begin{aligned} n_h &= \frac{n_h'}{\alpha_h n' + n_h'} \qquad (h = 1, 2, \ldots, L) \\ m_{hi} &= \frac{n_{hi}^2}{\beta_{hi} n_h (1 - \frac{n_h}{n_h'}) + n_{hi}} \qquad i = 1, 2, \ldots, k_h;\ h = 1, 2, \ldots, L \end{aligned}\right\} \quad (2.1.22)$$

where α_h, β_{hi} are constants ≥ 0, fixed in advance; noting that the subsampling fractions $\frac{n_h}{n_h'}$, $\frac{m_{hi}}{n_{hi}}$ respectively, vary according to the observed values of n_h' and n_{hi} unlike the previous procedure. The variance of $\bar{y}_{3p}$ is given by

$$V(\bar{y}_{3p}) = (\frac{1}{n'} - \frac{1}{N})S^2 + \sum_{h=1}^{L} \frac{\alpha_h S_h^2}{n'} + \sum_{h=1}^{L} \sum_{i=1}^{k_h} \frac{\alpha_{hi} \beta_h S_h^2}{n'} \qquad (2.1.23)$$

which is independent of the weights W_h and W_{hi}. The expected cost under subsampling rule (2.1.22) is

$$\begin{aligned} E(C) &= n'c' + \sum_{h=1}^{L} E\left(\frac{n_h'^2}{\alpha_h n' + n_h'}\right) c_h + \sum_{h=1}^{L} \sum_{i=1}^{k_h} E\left(\frac{n_{hi}^2}{\beta_{hi} n_h (1 - \frac{n_h}{n_h'}) + n_{hi}}\right) \\ &\simeq n'c' + \sum_{h=1}^{L} \frac{n' W_h^2 c_h}{\alpha_h + W_h} + \sum_{h=1}^{L} \sum_{i=1}^{k_h} \frac{n' W_{hi}^2 c_{hi}}{\alpha_h \beta_{hi} + \frac{W_{hi}}{W_h} (\alpha_h + W_h)} \end{aligned} \qquad (2.1.24)$$

for sufficiently large n' and n_{hi}.

The optimal α_h and β_{hi} can be calculated in the usual way and is omitted here.

2.2 NON-RESPONSE THEORY

Rao (1973) has shown that the classical non-response theory is a special case of Double Sampling for stratification. The subsampling of non-respondents with three attempts is a special case of our three-phase sampling for stratification. A sample of size n' is selected by simple random sampling without replacement. A subsample of size $n_2 = n_2'/k$ $(k \geq 1)$ from the n_2' $(= n' - n_1')$ non-respondents, again selected by simple random sampling without replacement, is contacted by more intensive efforts. Once again a subsample of size $m_{22} = n_{22}/\ell$ $(\ell \geq 1)$ from the n_{22} $(= n_2 - n_{21})$ non-respondents, selected by simple random sampling without replacement, is enumerated with serious efforts. This procedure is clearly a special case of our three-phase sampling for stratification with

$$L = 2,\ \nu_1 = 1,\ \nu_2 = \frac{1}{k},\quad \lambda_{11} = \lambda_{12} = \lambda_{21} = 1,\quad \text{and} \quad \lambda_{22} = \frac{1}{\ell}.$$

The variance formula reduces to

$$V(\bar{y}_{3p}) = \left(\frac{1}{n'} - \frac{1}{N}\right)S^2 + \frac{W_2 S_2^2}{n'}(k-1) + W_{22}\frac{S_{22}^2}{n'}k(\ell-1). \qquad (2.2.1)$$

The expected cost is

$$C^* = n'\left(c' + W_1 c_1 + W_{11}c_{11} + W_{12}c_{12} + \frac{W_{21}c_{21}}{k}\right) + \frac{n'}{k}W_2 c_2 + \frac{n'}{k\ell}W_{22}c_{22} \qquad (2.2.2)$$

where c' is the unit cost of making the first attempt, c_1 is unit cost of processing the data in stratum 1, c_2 is the unit cost of processing the data in stratum 2, c_{11}, c_{12}, c_{21}, c_{22} being the unit cost of processing the data in the respective substratums. The choice $\nu_{(1)} = 1$, is not optimal if the cost ratio $\frac{c'}{c}$ is small since the optimal ν_1 may be < 1. Similary, if $\frac{c_{11}}{c_1}$, $\frac{c_{12}}{c_1}$, $\frac{c_{21}}{c_2}$ are small, then λ_{11}, λ_{12}, λ_{21} may not be optimal since the optimal λ_{11}, λ_{12}, λ_{21} may be < 1. The optimal $k = \frac{1}{\nu_2}$ and $\ell = \frac{1}{\lambda_{22}}$ is obtained from (2.1.13) to (2.1.21) as

$$\ell_{opt} = \left\{ \frac{c_{22}(S_2^2 - W_{22}S_{22}^2)}{S_{22}^2(c'+W_{21}c_{21})} \right\}^{\frac{1}{2}} \tag{2.2.3}$$

and

$$k_{opt} = \left\{ \frac{c_2}{S_2^2} \, \frac{S^2 - W_2 S_2^2}{c'+W_1c_1+W_{11}c_{11}+W_{12}c_{12}+W_{21}c_{21}} \right\}^{\frac{1}{2}} \tag{2.2.4}$$

Srinath's (1971) method for subsampling of non-respondents, when the rate of non-response is not accurately known, is a special case of our procedure with $L = 2$, $k_h = 2$, $h = 1, 2$.

2.3 EMPIRICAL ILLUSTRATION

We shall demonstrate by an artificial example that intuitional approach to allocation when inequality constraints are involved, may not always give the optimal solution. To simplify the discussion we consider only the case of Double Sampling for stratification.

Let $W_1 = .1$, $W_2 = .2$, $W_3 = .4$, $W_4 = .3$, $S_1^2 = 150$, $S_2^2 = 140$, $S_3^2 = 50$, $S_4^2 = 40$, $S^2 = 125$, $N = 2000$, $c' = \$5$, $c_1 = \$10$, $c_2 = \$12$, $c_3 = \$25$, $c_4 = \$40$, and $C^* = \$2000$. Now

$$\frac{S_{(1)}^2}{c_{(1)}} = 15, \quad \frac{S_{(2)}^2}{c_{(2)}} = 11.7, \quad \frac{S_{(3)}^2}{c_{(3)}} = 2, \quad \frac{S_{(4)}^2}{c_{(4)}} = 1, \quad \frac{B}{c'} = 10,$$

$$\frac{B + W_{(1)}S_{(1)}^2}{c' + W_{(1)}c_{(1)}} = 10.8, \quad \frac{B + W_{(1)}S_{(1)}^2 + W_{(2)}S_{(2)}^2}{c' + W_{(1)}c_{(1)} + W_{(2)}c_{(2)}} = 11.$$

By intuition, we may push $\nu_{(1)} = \nu_{(2)} = 1$ since $\frac{S_{(2)}^2}{c_{(2)}} \geq \frac{B}{c'}$. In this case our intuitional approach does give the optimal solution and this is because

$$\frac{S_{(2)}^2}{c_{(2)}} \; (= 11.7) \geq \frac{B + W_{(1)}S_{(1)}^2 + W_{(2)}S_{(2)}^2}{c' + W_{(1)}c_{(1)} + W_{(2)}c_{(2)}} \; (= 11)$$

Suppose $c_{(2)}$ was \$14 instead of \$12. We see that

$$\frac{S_{(2)}^2}{c_{(2)}} \; (= 10) \geq \frac{B}{c'} \; (= 10).$$

By intuition we push $\nu_{(1)} = 1$ and $\nu_{(2)} = 1$ and re-allocate

$\nu_{(3)} = .433$, $\nu_{(4)} = .306$ and $n' = 120$. The optimal variance would be $V_1 = 1.416$. Using our algorithm we see that

$$\frac{S^2_{(2)}}{c_{(2)}} \; (= 10) \qquad \frac{B + W_{(1)}S^2_{(1)} + W_{(2)}S^2_{(2)}}{c' + W_{(1)}c_{(1)} + W_{(2)}c_{(2)}} \; (= 10.6).$$

The optimal values are $\nu^*_{(1)} = 1$, $\nu^*_{(2)} = .962$, $\nu^*_{(3)} = .430$, $\nu^*_{(4)} = .304$ and $n' = 121$. Our optimal variance is $V^*_1 = 1.400$. Clearly, $V^*_1 < V_1$.

2.4 ILLUSTRATION THROUGH CAUCHY INEQUALITY

For simplicity let us take Double Sampling for stratification with three stratums. Using Rao's (1973) formula, we get, after omitting the term S^2/N,

$$\dot{V} = \frac{B^2}{n'} + \frac{W_1 S_1^2}{n'\nu_1} + \frac{W_2 S_2^2}{n'\nu_2} + \frac{W_3 S_3^2}{n'\nu_3} \tag{2.4.1}$$

$$C^* = n'(c' + W_1 c_1 \nu_1 + W_2 c_2 \nu_2 + W_3 c_3 \nu_3) \tag{2.4.2}$$

where $0 < \nu_1, \nu_2, \nu_3 \leq 1$. Minimizing (2.4.1) subject to (2.4.2) is the same as minimizing

$$\phi = VC^*. \tag{2.4.3}$$

Suppose

$$\frac{B}{c} < \frac{S^2_{(2)}}{c_{(2)}} < \frac{B + W_{(1)}S^2_{(1)} + W_{(2)}S^2_{(2)}}{c' + W_{(1)}c_{(1)} + W_{(2)}c_{(2)}} . \tag{2.4.4}$$

Our intuitional approach gives optimal solution as

$$\nu_{(1)} = \nu_{(2)} = 1 \text{ and } \nu_{(3)} = \frac{S_{(3)}}{\sqrt{c_{(3)}}}\sqrt{\frac{c' + W_{(1)}c_{(1)} + W_{(2)}c_{(2)}}{B+W_{(1)}c_{(1)} + W_{(2)}S^2_{(2)}}}$$

and the minimum ϕ would be

$$\phi_1 = \left\{\alpha^2 + W_{(2)}S^2_{(2)} + W_{(3)}S_{(3)}\sqrt{c_3}\sqrt{\frac{\alpha^2 + W_{(2)}S^2_{(2)}}{\beta^2 + W_{(2)}c_{(2)}}}\right\} \times$$

$$\left\{\beta^2 + W_{(2)}c_{(2)} + W_{(3)}S_{(3)}\sqrt{c_3}\sqrt{\frac{\beta^2 + W_{(2)}c_{(2)}}{\alpha^2 + W_{(2)}S^2_{(2)}}}\right\} \tag{2.4.5}$$

where

$$\alpha^2 = B + W_{(1)}S^2_{(1)}$$

and

$$\beta^2 = c' + c_{(1)}W_{(1)}.$$

Now using our algorithm, we get

$$\nu_{(1)} = 1, \; \nu_{(2)} = \frac{S_{(2)}}{\sqrt{c_{(2)}}} \sqrt{\frac{c' + W_{(1)}c_{(1)}}{B + W_{(1)}S^2_{(1)}}},$$

$$\nu_{(3)} = \frac{S_{(3)}}{\sqrt{c_{(3)}}} \sqrt{\frac{c' + W_{(1)}c_{(1)}}{B + W_{(1)}S^2_{(1)}}}$$

and the minimum ϕ after simplifications would be

$$\phi_2 = (\alpha\beta + W_{(2)}S_{(2)}\sqrt{c_{(2)}} + W_{(3)}S_{(3)}\sqrt{c_{(3)}}\;)^2. \tag{2.4.6}$$

From (2.4.5) and (2.4.6) and using Cauchy inequality, we conclude that

$$\phi_1 \geq \phi_2.$$

CHAPTER III

DOUBLE SAMPLING FOR TWO-STAGE SAMPLING

The situation considered here is characterized by a deficient knowledge of the composition of the clusters in the population. It is assumed, however, that it is possible to select a preliminary sample of small units or elements, and determine by some inexpensive method to which cluster a particular unit of the initial sample belongs. The first sampling phase is thus used to group the units of the population in order to form a provisional sampling frame for the selection of the final sample in the form of a two-stage sample. This method of sampling, of course, does not differ in principle from Double sampling for stratification; in both cases the preliminary sample is used to determine a group identification not known in advance.

The following symbols all refer to the population with L primary units. For the primary unit h: $W_h = \frac{N_h}{N}$, $w_h = \frac{n'_h}{n'}$, $\bar{y}_h$ is the mean based on the n'_h units of s(n') and s'^2_h is the unbiased estimator of S^2_h based on the n'_h units of s(n') and k is the number of (pre-assigned) primary units subsampled from the L' primary units s(L') by pps sampling with replacement, probability proportional to $\frac{n'_h}{n'}$. Let s(k) denote the set of primary units in the sample.

THEOREM 1

The estimator

$$\bar{y}_{d2} = \frac{1}{k} \sum_{h=1}^{k} \bar{y}_h \tag{3.1}$$

is unbiased for $\bar{Y}$ with variance

$$V(\bar{y}_{d2}) = \left(\frac{1}{n'} - \frac{1}{N}\right)S^2 + \frac{1}{k}\left(S^2 - \sum_{h=1}^{L} W_h S_h^2\right) + \frac{1}{kn'}\left\{\sum_{h=1}^{L} \Pi_h \frac{S_h^2}{\nu_h} - S^2\right\} \tag{3.2}$$

where

$$\Pi_h = 1 - \binom{N - N_h}{n'} \Big/ \binom{N}{n'}$$

and 'd2' denotes double sampling for two-stage sampling.

Proof: For a given $\underset{\sim}{n}' = (n_1', n_2', \ldots, n_{L'}')$, s(k) is a pps (probability proportional sampling) sample from s(L'). Let E_2 and V_2 respectively denote the expectation and variance operators for a given $\underset{\sim}{n}'$ and s(n'). Similarly, $s(n_h)$ is drawn from $s(n_h')$ by srs independently, each time, primary unit h is drawn in sample. Consequently, we denote E_3 and V_3 as the expectation and variance operators for a given $\underset{\sim}{n}'$ and s(k). Let E_1 and V_1 be the unconditional expectation and variance operators for s(n'). We have

$$E(\bar{y}_{d2}) = E_1 E_2 E_3(\bar{y}_{d2})$$

$$E_3(\bar{y}_h) = \bar{y}_h'$$

$$E_2 E_3(\bar{y}_{d2}) = E_2\left(\frac{1}{k} \sum_{h=1}^{k} \bar{y}_h'\right) = \sum_{h=1}^{L'} \frac{n_h'}{n'} \bar{y}_h' = \bar{y}' .$$

Similarly

$$V(\bar{y}_{d2}) = V_1E_2E_3(\bar{y}_{d2}) + E_1V_2E_3(\bar{y}_{d2}) + E_1E_2V_3(\bar{y}_{d2})$$

$$= V_1E_2 \frac{1}{k} \sum_{h=1}^{k} \bar{y}'_h + E_1V_2 \frac{1}{k} \sum_{h=1}^{k} \bar{y}'_h$$

$$+ E_1E_2 \frac{1}{k} \sum_{h=1}^{k} (\frac{1}{n_h} - \frac{1}{n'_h}) s'^2_h .$$

Now using the relationship

$$n_h = \nu_h n'_h \tag{3.3}$$

and simplifying, we get

$$V(\bar{y}_{d2}) = V_1(\bar{y}') + E_1\left\{\frac{1}{k} \sum_{h=1}^{L'} w_h(\bar{y}'_h - \bar{y}')^2\right\}$$

$$+ E_1\left\{\frac{1}{k} \sum_{h=1}^{L'} \frac{n'_h}{n'} \frac{s'^2_h}{n'_h} (\frac{1}{\nu_h} - 1)\right\}. \tag{3.4}$$

Now

$$V_1(\bar{y}') = (\frac{1}{n'} - \frac{1}{N})S^2 \tag{3.5}$$

where S^2 is the population variance.

$$E_1\left\{\frac{1}{k} \sum_{h=1}^{L'} w_h(\bar{y}'_h - \bar{y})^2\right\} = E_1\left\{\frac{1}{k} \sum_{h=1}^{L'} w_h[(\bar{y}'_h - \bar{Y}) - (\bar{y}' - \bar{Y})]^2\right\}$$

$$= E_1\left[\frac{1}{k} \sum_{h=1}^{L'} w_h(\bar{y}'_h - \bar{Y}_h)^2\right] + E_1\left[\frac{1}{k} \sum_{h=1}^{L'} w_h(\bar{Y}_h - \bar{Y})^2\right] - E_1\left[\frac{1}{k} (\bar{y}' - \bar{Y})^2\right]$$

$$= \frac{1}{kn'} \sum_{h=1}^{L} S^2_h(\Pi_h - \frac{n'}{N}) + \frac{1}{k} \sum_{h=1}^{L} W_h(\bar{Y}_h - \bar{Y})^2 - \frac{1}{k} V_1(\bar{y}'); \tag{3.6}$$

$$E_1 \frac{1}{k} \sum_{h=1}^{L'} \frac{n'_h}{n'} \frac{S'^2_h}{n'_h} (\frac{1}{\nu_h} - 1) = \frac{1}{kn'} \sum_{h=1}^{L} {}_hS^2_h(\frac{1}{\nu_h} - 1). \tag{3.7}$$

Also

$$(1 - N^{-1})S^2 = \sum_{h=1}^{L} W_h(\bar{Y}_h - \bar{Y})^2 + \sum_{h=1}^{L} (W_h - N^{-1})S^2_h. \tag{3.8}$$

(3.4), (3.5), (3.6), (3.7) and (3.8) lead to (3.2)

Berg (1972) proposed the selection of n_h (fixed number), of units from $s(n'_h)$ by srs with replacement. His variance formula reduces to

$$V(\bar{y}^*_{d2}) = \left(\frac{1}{n'} - \frac{1}{N}\right)S^2 + \frac{\left(S^2 - \sum_{h=1}^{L} W_h S_h^2\right)}{k} + \frac{1}{k}\sum_{h=1}^{L} W_h S_h^2$$

$$+ \frac{1}{kn'}\left\{\sum_{h=1}^{L} \Pi_h S_h^2\left(1 - \frac{1}{n_h}\right) - S^2\right\}. \qquad (3.9)$$

When n' is large it is reasonable that

$$\Pi_h \simeq 1 \quad \text{and} \quad n_h \simeq n' W_h \nu_h \qquad (3.10)$$

and from (3.2), (3.9) and (3.10), we get

$$V(\bar{y}^*_{d2}) - V(\bar{y}_{d2}) = \frac{1}{kn'}\sum_{h=1}^{L} S_h^2\left(1 - \frac{1}{n_h}\right)$$

$$> 0$$

which clearly shows (as may be expected) the loss in efficiency is due to sampling with replacement.

We turn now to the optimal determination of n', k and $\underline{\nu} = (\nu_1, \nu_2, \ldots, \nu_L)$.

The cost function is taken as

$$C = n'c' + kc_t + \sum_{h=1}^{k} n_h c_h \qquad (3.11)$$

where c' is the unit cost of enumerating preliminary sampling, c_t being the unit travel cost between primary units and c_h is the unit cost of enumerating the secondary units. Since C in (3.11) is random, we again use the expected cost;

$$C^* = n'c' + kc_t + kn' \sum_{h=1}^{L} W_h c_h \nu_h . \tag{3.12}$$

Rao (1973) has given a simple procedure for finding the optimal values of n' and $\underline{\nu}$ which minimize

$$\bar{V} = \sum_{h=1}^{L} \frac{a_h^2}{n' W_h \nu_h}$$

subject to

$$C^* = n'c' + \sum_{h=1}^{L} (n' c_h \nu_h W_h) \quad \text{and} \quad 0 < \nu_h \leq 1,$$

where a_h are known constants. To find the optimal n' and $\underline{\nu}$, he first determined the optimal $\underline{\nu}$ for a given n' and then the optimal n'. His procedure may be extended to our case for determining the optimal n', k and $\underline{\nu}$ which minimize (3.2) subject to (3.12) and $0 < \nu_h \leq 1$.

To find the optimal n', k and $\underline{\nu}$ we determine first the optimal k and $\underline{\nu}$ for a given n', and then the optimal n'. Using the Cauchy inequality in conjunction with (3.2) and (3.12), it immediately follows that the optimal k and $\underline{\nu}$ for a given n' are given by

$$n'\nu_h = \sqrt{\frac{\Pi_h S_h^2 c_t}{W_h c_h (B - \frac{S^2}{n'})}} \qquad (h = 1, 2, \ldots, L) \tag{3.13}$$

$$k = c_t^{-1} \left[c^* - c'n' - \frac{\sum_{h=1}^{L} \sqrt{c_h \Pi_h W_h S_h^2}}{\sqrt{(B - S^2/n')}} \right] \tag{3.14}$$

provided $n'\nu_h \leq n'$ for all h; that is,

$$n' \geq \frac{S^2 + \sqrt{S^4 + 4B\,\dfrac{\Pi_{(1)}S^2_{(1)}c_t}{W_{(1)}c_{(1)}}}}{B} = m_1' \text{ (say)}$$

where (1) denotes the group with the largest value of $\dfrac{\Pi_h S_h^2}{W_h c_h}$.

The minimum value of (3.2) for $n' \geq m_1'$, after substituting the optimal k and $\underline{\nu}$, is given by

$$V_1(\bar{y}_{d2}) = \left(\frac{1}{n'} - \frac{1}{N}\right)S^2 + \frac{\left\{\sqrt{\left(B - \dfrac{S^2}{n'}\right)c_t} + \sum_{h=1}^{L}\sqrt{c_h \pi_h W_h S_h^2}\right\}^2}{(C^* - n'c')} \tag{3.15}$$

so that the minimum occurs at the value $m_1 = m_1'$. Note that $\nu_{(1)} = 1$ when $n' = m_1$.

We consider next the values of $n' \leq m_1$. Since $\nu_{(1)} \geq 1$ for these values, we set $\nu_{(1)} = 1$ and re-allocate the remaining ν_h and k again by Cauchy's inequality. This gives

$$n'\nu_{(h)} = \sqrt{\frac{\Pi_{(h)}S^2_{(h)}[c_t + n'W_{(1)}c_{(1)}]}{W_{(h)}c_{(h)}\left[B - \dfrac{S^2 - \Pi_{(1)}S^2_{(1)}}{n'}\right]}} \qquad (h = 2, 3, \ldots, L) \tag{3.16}$$

$$k = [c_t + n'W_{(1)}c_{(1)}]^{-1}\left[C^* - c'n' - \sum_{h=2}^{L}\sqrt{\frac{c_{(h)}W_{(h)}\Pi_{(h)}S^2_{(h)}[c_t + n'W_{(1)}c_{(1)}]}{\left[B - \dfrac{(S^2 - \Pi_{(1)}S^2_{(1)})}{n'}\right]}}\right] \tag{3.17}$$

where (h) denotes the group with the h^{th} largest value of

$(\Pi_h S_h^2/W_h c_h)$ and provided $n'\nu_{(h)} \leq n'$ for all h; that is

$$n' \geq \left\{ [S^2 - \Pi_{(1)}S_{(1)}^2 + \Pi_{(2)}S_{(2)}^2 W_{(1)}c_{(1)}/W_{(2)}c_{(2)}] + \frac{\sqrt{[S^2 - \Pi_{(1)}S_{(1)}^2 + \Pi_{(2)}S_{(2)}^2 W_{(1)}c_{(1)}/W_{(2)}c_{(2)}]^2 + 4B\,\frac{\Pi_{(2)}S_{(2)}^2}{W_{(2)}c_{(2)}}}}{2B} \right\}$$

$$= m'_{(2)} \text{ (say).} \tag{3.18}$$

The minimum value of (3.2) for $n' \geq m_2'$ after substituting the optimal $\underline{\nu}$ and k from (3.17) and (3.18) is given by

$$V_2(\bar{y}_{d2}) = \left(\frac{1}{n'} - \frac{1}{N}\right)S^2 + \frac{\left\{\left[B - \frac{(S^2 - \Pi_{(1)}S_{(1)}^2)}{n'}\right]\left[c_t + \frac{W_{(1)}c_{(1)}}{n'}\right] + \sum_{h=2}^{L} \sqrt{c_{(h)}W_{(h)}\Pi_{(h)}S_{(h)}^2}\right\}^2}{(C^* - n'c')} . \tag{3.19}$$

We examine the derivative $\frac{dV_2(\bar{y}_{d2})}{dn'}$ over the range $m_2' \leq n' \leq m_1$ to find the optimal n'. The derivative vanishes at $n' = \tilde{m}_2$ (say), and the derivative is $\lesseqgtr 0$ for $n' \lesseqgtr m_2$. Consequently, if $\tilde{m}_{(2)} > m_1$, $V_2(\bar{y}_{d2})$ monotonically decreases as n' increases so that the minimum occurs at the value $m_2 = m_1$. If, however, $\tilde{m}_2$ lies in (m_2', m_1), the true optimum will be given by $m_2 = \tilde{m}_2$ and the procedure may be terminated here since, in practice, $\bar{V}(n')$ will have a unique minimum.

The general procedure is now clear. If $m_2 \neq \tilde{m}_2$, we set $\nu_{(1)} = \nu_{(2)} = 1$ and re-allocate the remaining ν_j. All in all, L steps will be involved if the derivative is non-vanishing over the (L-1) ranges $n' \geq m_1'$, $m_h' \leq n' \leq m_{h-1}'$, $h = 2,\ldots,(L-1)$ where

$$m_h' = \frac{\left\{\left[S^2 - \sum_{j=1}^{h-1} \Pi_{(j)} S_{(j)}^2 + \frac{\Pi_{(h)} S_{(h)}^2}{W_{(h)} c_{(h)}} \sum_{j=1}^{h-1} W_{(j)} c_{(j)}\right] + \sqrt{\left(S^2 - \sum_{j=1}^{h-1} \Pi_{(j)} S_{(j)}^2 + \frac{\Pi_{(h)} S_{(h)}^2}{W_{(h)} c_{(h)}} \sum_{j=1}^{h-1} W_{(j)} c_{(j)}\right)^2 + 4B \frac{\Pi_{(h)} S_{(h)}^2}{W_{(h)} c_{(h)}}}\right\}}{2B} \tag{3.20}$$

(j) denotes the group with the j^{th} largest value of $\frac{\Pi_h S_{(h)}^2}{W_h c_h}$.

The derivative of $V_h(\bar{y}_{d2})$ in the range

$$m_h' \leq n' \leq m_{h-1}'$$

vanishes at

$$n' = \tilde{m}_h \quad \text{(say)}.$$

Denoting the optimal n' in the range

$$m_h' \leq n' \leq m_{h-1}'$$

by m_h, we compare the values $V_h(\bar{y}_{d2})$ to find the true optimal n' and then the corresponding $\underline{\nu}$ and k where

$$V_h(\bar{y}_{d2}) = \left(\frac{1}{n'} - \frac{1}{N}\right)S^2 + (PQ + R)^2 \tag{3.21}$$

and where

$$P = \left[B - \left(S^2 - \sum_{j=1}^{h-1} \pi_{(j)} S_{(j)}^2\right)/n'\right]^{\frac{1}{2}}$$

$$Q = \left[c_t + n' \sum_{j=1}^{h-1} W_{(j)} c_{(j)}\right]^{\frac{1}{2}}$$

and

$$R = \sum_{j=h}^{L} \sqrt{c_{(j)} W_{(j)} \Pi_{(j)} S^2_{(j)}} \; .$$

If the optimum occurs in the range $m'_h \leq n' \leq m'_{h-1}$, the optimal $\underline{\nu}$ and k is given by

$$\left.\begin{aligned} &\nu_{(1)} = \nu_{(2)} = \ldots = \nu_{(h-1)} = 1 \\ &n'\nu_{(j)} = \sqrt{\frac{\Pi_{(j)} S^2_{(j)}}{W_{(j)} c_{(j)}} \frac{Q}{P}} \qquad j = h,\ldots,L \end{aligned}\right\} \qquad (3.22)$$

$$k = (c^* - c'n' - \frac{RQ}{P})/Q. \qquad 3.23)$$

The determination of optimal n' might involve scanning of the minimum variance in the neighbourhood of n' obtained from (3.21).

APPENDIX

The following conditions are used in computing the algorithm given in Chapter II for the optimal determination of λ_{hi} and ν_h. Though we analyzed both boundary as well as non-boundary conditions, we do not present the boundary conditions here in order to simplify our discussion. Using Thompson's Lemma, we distinguish the following cases.

1. $\ell_{h(1)} < m_h$, $h = 1, 2, \ldots, L$

$$m_{(1)} < n'$$

$$\frac{\partial g}{\partial n'} = 0, \quad \frac{\partial g}{\partial m_h} = 0, \qquad h = 1, 2, \ldots, L$$

$$\frac{\partial g}{\partial \ell_{hi}} = 0, \quad i = 1, 2, \ldots, k_h; \ h = 1, 2, \ldots, L.$$

The conditions are

$$\frac{S^2_{h(1)}}{c_{h(1)}} < \frac{A_h}{c_h}, \quad h = 1, 2, \ldots, L$$

$$\frac{A_{(1)}}{c_{(1)}} < \frac{B}{c'} .$$

Optimal Solution:

$$\nu_h = \sqrt{\frac{A_h}{c_h} \frac{c'}{B}}, \quad h = 1, 2, \ldots, L$$

$$\lambda_{hi} = \sqrt{\frac{S^2_{hi}}{c_{hi}} \frac{c_h}{A_h}}; \quad i = 1, 2, \ldots, k_h; \ h = 1, 2, \ldots, L.$$

2. $\ell_{h(1)} < m_h$, $h = 1, 2, \ldots, L$

$$m_{(1)} = n', \; m_{(2)} < n'$$

$$\frac{\partial g}{\partial n'} + \frac{\partial g}{\partial m_{(1)}} = 0, \quad \frac{\partial g}{\partial m_{(1)}} < 0$$

$$\frac{\partial g}{\partial m_{(h)}} = 0, \quad h = 2, 3, \ldots, L$$

$$\frac{\partial g}{\partial \ell_{h(i)}} = 0, \quad i = 1, 2, \ldots, k_h; \; h = 1, 2, \ldots, L.$$

The conditions are

$$\frac{S^2_{h(1)}}{c_{h(1)}} < \frac{A_h}{c_h}, \quad h = 1, 2, \ldots, L.$$

$$\frac{A_{(1)}}{c_{(1)}} > \frac{B + W_{(1)}A_{(1)}}{c' + W_{(1)}c_{(1)}}, \quad \frac{A_{(2)}}{c_{(2)}} < \frac{B + W_{(1)}A_{(1)}}{c' + W_{(1)}c_{(1)}},$$

or equivalently,

$$\frac{A_{(1)}}{c_{(1)}} > \frac{B}{c'}$$

$$\frac{A_{(2)}}{c_{(2)}} < \frac{B + W_{(1)}A_{(1)}}{c' + W_{(1)}c_{(1)}}.$$

Optimal Solution:

$$\nu_{(1)} = 1$$

$$\nu_{(h)} = \sqrt{\frac{A_h}{c_h} \frac{(c' + (W_{(1)}c_{(1)})}{(B + W_{(1)}A_{(1)})}}, \quad h = 2, 3, \ldots, L$$

$$\lambda_{hi} = \sqrt{\frac{S^2_{hi}}{c_{hi}} \frac{c_h}{A_h}}, \quad i = 1, 2, \ldots, k_h; \; h = 1, 2, \ldots, L.$$

3. $\ell_{h(1)} < m_h$, $h = 1, 2, \ldots, L$

$$m_{(r)} = n', \quad r = 2, 3, \ldots, L$$

$$m_{(h)} < n', \quad h = r+1, \ldots, L$$

$$\frac{\partial g}{\partial \ell_{hi}} = 0, \quad i = 1, 2, \ldots, k_h; \; h = 1, 2, \ldots, L$$

$$\frac{\partial g}{\partial m_{(r)}} < 0, \quad r = 2, 3, \ldots, L$$

$$\frac{\partial g}{\partial m_{(h)}} = 0, \quad h = 3, 4, \ldots, L; \; h \neq 4$$

$$\frac{\partial g}{\partial n'} + \sum_{j=1}^{r} \frac{\partial g}{\partial m_{(j)}} = 0, \quad r = 2, 3, \ldots, L.$$

The conditions are

$$\frac{S^2_{h(1)}}{c_{h(1)}} < \frac{A_h}{c_h}, \quad h = 1, 2, \ldots, L$$

$$\frac{A_{(r)}}{c_{(r)}} > \frac{B + W_{(1)}A_{(1)} + \ldots + W_{(r)}A_{(r)}}{c' + W_{(1)}c_{(1)} + \ldots + W_{(r)}c_{(r)}}, \quad r = 2, 3, \ldots, L$$

$$\frac{A_{(h)}}{c_{(h)}} < \frac{B + W_{(1)}A_{(1)} + \ldots + W_{(r)}A_{(r)}}{c' + W_{(1)}c_{(1)} + \ldots + W_{(r)}c_{(r)}}, \quad h = r+1, \ldots, L$$

Optimal Solution:

$$\lambda_{hi} = \sqrt{\frac{S^2_{hi}}{c_{hi}} \frac{c_h}{A_h}}, \quad i = 1, 2, \ldots, k_h; \; h = 1, 2, \ldots, L.$$

$$\nu_{(r)} = 1, \quad r = 1, 2, \ldots, L$$

$$\nu_{(h)} = \frac{S^2_{(h)}}{c_{(h)}} \frac{c' + W_{(1)}c_{(1)} + \ldots + W_{(r)}c_{(r)}}{B + W_{(1)}A_{(1)} + \ldots + W_{(r)}A_{(r)}}, \quad h = r+1, \ldots, L.$$

4. $\ell_{s_1}(1) = m_{s_1}$, $\ell_{s_2}(1) = m_{s_2}$, ..., $\ell_{s_r}(1) = m_{s_r}$

$1 \leq s_1, s_2, \ldots, s_r \leq L$, $s_1 \neq s_2 \neq \ldots \neq s_r$

$\ell_{h(1)} < m_h$, $h = 1, 2, \ldots, L$, $h \neq s_1, s_2, \ldots, s_r$

$\ell_{si(2)} < m_{si}$, $i = 1, 2, \ldots, r$

$m_{(1)} < n'$

$$\frac{\partial g}{\partial \ell_{si(1)}} < 0, \quad i = 1, 2, \ldots, r$$

$$\frac{\partial g}{\partial \ell_{si(j)}} = 0, \quad j = 2, 3, \ldots, k_{si}; \; i = 1, 2, \ldots, r.$$

$$\frac{\partial g}{\partial \ell_{h(j)}} = 0, \quad j = 1, 2, \ldots, k_h; \; h = 1, 2, \ldots, L; \; h \neq s_1, s_2, \ldots, s_r$$

$$\frac{\partial g}{\partial m_{si}} + \frac{\partial g}{\partial \ell_{si(1)}} = 0, \quad i = 1, 2, \ldots, r$$

$$\frac{\partial g}{\partial m_h} = 0, \quad h = 1, 2, \ldots, L, \; h \neq s_1, s_2, \ldots, s_r$$

$$\frac{\partial g}{\partial n'} = 0.$$

The conditions are

$$\frac{S^2_{si(1)}}{c_{si(1)}} > \frac{W_{si}A_{si} + W_{si(1)}S^2_{si(1)}}{W_{si}c_{si} + W_{si(1)}c_{si(1)}}, \quad i = 1, 2, \ldots, r$$

or equivalently,

$$\frac{S^2_{si(2)}}{c_{si(2)}} > \frac{W_{si}A_{si} + W_{si(1)}S^2_{si(1)}}{W_{si}c_{si} + W_{si(1)}c_{si(1)}}, \quad i = 1, 2, \ldots, r$$

$$\frac{S^2_{h(1)}}{c_{h(1)}} < \frac{A_h}{c_h}, \quad h = 1, 2, \ldots, L; \; h \neq s_1, s_2, \ldots, s_r$$

$$\frac{W_{si}A_{si} + W_{si(1)}S^2_{si(1)}}{W_{si}c_{si} + W_{si(1)}c_{si(1)}} < \frac{B}{c'}, \quad i = 1, 2, \ldots, r$$

$$\frac{A_h}{c_h} < \frac{B}{c'}, \quad h = 1, 2, \ldots, L; \; h \neq s_1, s_2, \ldots, s_r$$

Optimal Solution:

$$\lambda_{si(1)} = 1, \quad i = 1, 2, \ldots, r$$

$$\lambda_{si(j)} = \sqrt{\frac{S^2_{si(j)}}{c_{si(j)}} \frac{W_{si}c_{si} + W_{si(1)}c_{si(1)}}{W_{si}S^2_{si} + W_{si(1)}S^2_{si(1)}}}, \quad j = 2, \ldots, k_{si}; \; i = 1, 2, \ldots, r$$

$$\lambda_{h(i)} = \sqrt{\frac{S^2_{h(i)}}{c_{h(i)}} \frac{c_h}{A_h}}, \quad k = 1, 2, \ldots, k_h; \; h = 1, 2, \ldots, L; \; h \neq s_1, s_2, \ldots, s_r$$

$$\nu_{si} = \sqrt{\frac{W_{si}A_{si} + W_{si(1)}S^2_{si(1)}}{W_{si}c_{si} + W_{si(1)}c_{si(1)}} \frac{c'}{B}}, \quad i = 1, 2, \ldots, r$$

$$\nu_h = \sqrt{\frac{A_h}{c_h} \frac{c'}{B}}, \quad h = 1, 2, \ldots, L; \; h \neq s_1, s_2, \ldots, s_r.$$

5. $\ell_{si(1)} = \ldots = \ell_{si(q_i)} = m_{si}$

$$1 \leq q_i \leq k_{si}, \quad i = 1, 2, \ldots, r$$

$$1 \leq s_1, s_2, \ldots, s_r \leq L, \; s_1 \neq s_2 \neq \ldots \neq s_r$$

$$\ell_{si(j)} < m_{si}, \quad j = q_{i+1}, \ldots, k_{si}$$

$$\ell_{h(1)} < m_h, \quad h = 1, 2, \ldots, L; \; h \neq s_1, s_2, \ldots, s_r$$

$$m_h < n', \quad h = 1, \ldots, L$$

$$\frac{\partial g}{\partial \ell_{si(k)}} < 0, \quad k = 1, 2, \ldots, q_i, \quad L = 1, \ldots, r$$

$$\frac{\partial g}{\partial \ell_{si(j)}} = 0, \quad j = q_{i+1}, \ldots, k$$

$$\frac{\partial g}{\partial \ell_{h(j)}} = 0, \quad j = 1, 2, \ldots, k_h; \; h = 1, 2, \ldots, L; \; h \neq s_1, s_2, \ldots, s_r$$

$$\frac{\partial g}{\partial m_{si}} + \sum_{j=1}^{q_i} \frac{\partial g}{\partial \ell_{si(q_i)}} = 0, \quad i = 1, 2, \ldots, r$$

$$\frac{\partial g}{\partial m_h} = 0, \quad h = 1, 2, \ldots, L, \; h \neq s_1, s_2, \ldots, s_r$$

$$\frac{\partial g}{\partial n'} = 0.$$

The conditions are

$$\frac{S^2_{si(k)}}{c_{si(k)}} > \frac{W_{si}A_{si} + \sum_{j=1}^{q_i} W_{si(j)}S^2_{si(j)}}{W_{si}c_{si} + \sum_{j=1}^{q_i} W_{si(j)}c_{si(j)}}, \quad k = 1, 2, \ldots, q_i; \quad i = 1, 2, \ldots, n$$

$$\frac{S^2_{si(j)}}{c_{si(j)}} < \frac{W_{si}A_{si} + \sum_{j=1}^{q_i} W_{si(j)}S^2_{si(j)}}{W_{si}A_{si} + \sum_{j=1}^{q_i} W_{si(j)}S^2_{si(j)}}, \quad j = q_{i+1}, \ldots, k_{si}; \quad i = 1, 2, \ldots, r$$

$$\frac{S^2_{h(i)}}{c_{h(i)}} < \frac{A_h}{c_h}, \quad i = 1, 2, 3, \ldots, k_h, \; h \neq s_1, s_2, \ldots, s_r$$

$$\frac{W_{si}A_{si} + \sum_{j=1}^{q_i} W_{si(j)}S^2_{si(j)}}{W_{si}c_{si} + \sum_{j=1}^{q_i} W_{si(j)}c_{si(j)}} < \frac{B}{c'}, \quad i = 1, 2, \ldots, r$$

$$\frac{A_h}{c_h} < \frac{B}{c'}, \quad h = 1, 2, \ldots, L; \ h \neq s_1, s_2, \ldots, s_r$$

Optimal Solution:

$$\lambda_{si(j)} = 1, \quad j = 1, 2, \ldots, q_i; \ i = 1, 2, \ldots, r$$

$$\lambda_{si(k)} = \sqrt{\frac{S^2_{si(k)}}{c_{si(k)}} \frac{W_{si}c_{si} + \sum_{j=1}^{q_i} W_{si(j)}c_{si(j)}}{W_{si}A_{si} + \sum_{j=1}^{q_i} W_{si(j)}S^2_{si(j)}}}, \quad k = 1, 2, \ldots, q_i; \quad i = 1, 2, \ldots, r$$

$$\lambda_{h(i)} = \sqrt{\frac{S^2_{h(i)}}{c_{h(i)}} \frac{c_h}{A_h}}, \quad i = 1, 2, \ldots, k_h; \ h = 1, 2, \ldots, L; \quad h \neq s_1, s_2, \ldots, s_r$$

$$\nu_{si} = \sqrt{\frac{W_{si}A_{si} + \sum_{j=1}^{q_i} W_{si(j)}S^2_{si(j)}}{W_{si}c_{si} + \sum_{j=1}^{q_i} W_{si(j)}c_{si(j)}} \frac{c'}{B}}, \quad i = 1, 2, \ldots, 4$$

$$\nu_h = \sqrt{\frac{A_h}{c_h} \frac{c'}{B}}, \quad h = 1, 2, \ldots, L; \ h \neq s_1, s_2, \ldots, s_r$$

6. $\ell_{si(1)} = \ldots = \ell_{si(q_i)} = m_{si}$

$1 \leq q_i \leq k_{si}, \quad k = 1, 2, \ldots, r$

$1 \leq s_1, s_2, \ldots, s_r \leq L$

$\ell_{si(j)} < m_{si}, \quad j = q+1, \ldots, k_{si}$

$\ell_{h(1)} < m_h, \ h = 1, 2, \ldots, L; \ h \neq s_1, s_2, \ldots, s_r$

$$m_{t_i} = n', \quad i = 1, 2, \ldots, k; \; 1 \leq t_1, t_2, \ldots, t_k \leq L,$$
$$t_1 \neq t_2 \neq \ldots \neq t_k$$

$$m_h < n', \quad h = 1, 2, \ldots, L; \; h \neq t_1, t_2, \ldots, t_k$$

$$\frac{\partial g}{\partial \ell_{si(k)}} < 0, \quad k = 1, 2, \ldots, q_i; \; i = 1, 2, \ldots, r$$

$$\frac{\partial g}{\partial \ell_{si(j)}} = 0, \quad j = q+1, \ldots, k_{si}$$

$$\frac{\partial g}{\partial \ell_{h(j)}} = 0, \quad j = 1, 2, \ldots, k_h; \; h = 1, 2, \ldots, L; \; h \neq s_1, s_2, \ldots, s_r$$

$$\frac{\partial g}{\partial m_{t_i}} + \sum_{j=1}^{q_i} \frac{\partial g}{\partial \ell_{ti(j)}} < 0, \quad i = 1, 2, \ldots, k; \; t_i = s_1, s_2, \ldots, s_r$$

$$\frac{\partial g}{\partial m_{t_i}} < 0, \quad i = 1, 2, \ldots, k; \; t_i \neq s_1, s_2, \ldots, s_r$$

$$\frac{\partial g}{\partial m_h} + \sum_{j=1}^{q_i} \frac{\partial g}{\partial \ell_{h(j)}} = 0, \quad h = 1, 2, \ldots, L; \; h \neq t_1, t_2, \ldots, t_k;$$
$$h = s_1, s_2, \ldots, s_r$$

$$\frac{\partial g}{\partial m_h} = 0, \quad h = 1, 2, \ldots, L; \; h \neq s_1, s_2, \ldots, s_r, t_1, t_2, \ldots, t_k$$

$$\frac{\partial g}{\partial n'} + \sum_{j=1}^{k} \frac{\partial g}{\partial m_{t_j}} + \sum_{i=1}^{x} \sum_{j=1}^{q_i} \frac{\partial g}{\partial \ell_{ui(j)}} = 0$$

where $(u_1, u_2, \ldots, u_x) = (s_1, s_2, \ldots, s_r) \cap (t_1, t_2, \ldots, t_k)$.

The conditions are

$$\frac{S^2_{si(k)}}{c_{si(k)}} > \frac{W_{si}A_{si} + \sum_{j=1}^{q_i} W_{si(j)}S^2_{si(j)}}{W_{si}c_{si} + \sum_{j=1}^{q_i} W_{si(j)}c_{si(j)}}, \quad k = 1, 2, \ldots, q_i; \; i = 1, 2, \ldots, r$$

$$\frac{S^2_{si(j)}}{c_{si(j)}} > \frac{W_{si}A_{si} + \sum_{j=1}^{q_i} W_{si(j)}S^2_{si(j)}}{W_{si}c_{si} + \sum_{j=1}^{q_i} W_{si(j)}c_{si(j)}}, \quad k = 1, 2, \ldots, q_i; \quad i = 1, 2, \ldots, r$$

$$\frac{S^2_{h(i)}}{c_{h(i)}} < \frac{A_h}{c_h}, \quad i = 1, 2, \ldots, k_h; \; h \neq s_1, s_2, \ldots, s_r$$

$$\frac{W_{ti}A_{ti} + \sum_{j=1}^{q_i} W_{ti(j)}S^2_{ti(j)}}{W_{ti}c_{ti} + \sum_{j=1}^{q_i} W_{ti(j)}c_{ti(j)}} > R, \quad t_i = s_1, \ldots, s_r, \; i = 1, 2, \ldots, k$$

where

$$R = \frac{B + \sum_{i=1}^{k} W_{ti}A_{ti} + \sum_{i=1}^{x} \sum_{j=1}^{q_i} W_{ui(j)}S^2_{ui(j)}}{c' + \sum_{i=1}^{k} W_{ti}c_{ti} + \sum_{i=1}^{x} \sum_{j=1}^{q_i} W_{ui(j)}c_{ui(j)}}$$

$$\frac{A_{t_i}}{c_{t_i}} > R, \quad t_i = s_1, s_2, \ldots, s_r; \; i = 1, 2, \ldots, k$$

$$\frac{W_hA_h + \sum_{j=1}^{q_i} W_{h(j)}S^2_{h(j)}}{W_hc_h + \sum_{j=1}^{q_i} W_{h(j)}c_{h(j)}} < R, \quad h = s_1, s_2, \ldots, s_r; \; h \neq t_1, t_2, \ldots, t_k$$

$$\frac{A_h}{c_h} < R, \quad h \neq s_1, s_2, \ldots, s_r, t_1, t_2, \ldots, t_k; \; h = 1, 2, \ldots, L$$

<u>Optimal Solution</u>:

$$\lambda_{si(j)} = 1, \quad j = 1, 2, \ldots, q_i; \; i = 1, 2, \ldots, r$$

$$\lambda_{si(k)} = \sqrt{\frac{S^2_{si(k)}}{c_{si(k)}} \frac{W_{si}c_{si} + \sum_{j=1}^{q_i} W_{si(j)}c_{si(j)}}{W_{si}A_{si} + \sum_{j=1}^{q_i} W_{si(j)}S^2_{si(j)}}}\ , \quad k = 1, 2, \ldots, L \quad i = 1, 2, \ldots, r$$

$$\lambda_{h(j)} = \sqrt{\frac{S^2_{h(i)}}{c_{h(i)}} \frac{c_h}{A_h}}\ , \quad h = 1, 2, \ldots, L;\ h \neq s_1, s_2, \ldots, s_r$$

$$\nu_{t_i} = 1, \quad i = 1, 2, \ldots, k$$

$$\nu_h = \sqrt{\frac{W_hA_h + \sum_{j=1}^{q_i} W_{h(j)}S^2_{h(j)}}{W_hc_h + \sum_{j=1}^{q_i} W_{h(j)}c_{h(j)}} \frac{1}{R}}\ , \quad h = s_1, s_2, \ldots, s_r; \quad h \neq t_1, t_2, \ldots, t_k$$

$$\nu_h = \sqrt{\frac{A_h}{c_h} \frac{1}{R}}\ , \quad h = 1, 2, \ldots, L;\ h \neq s_1, s_2, \ldots, s_r, t_1, t_2, \ldots, t_k.$$

BIBLIOGRAPHY

Cochran, W. G. (1963). Sampling Techniques, 2nd edition, New York: Wiley and Sons.

Hansen, M. H. and Hurwitz, W. N. (1946). The problem of non-response in sample surveys, Journal of the American Statistical Association, 41, 517-529.

Neyman, J. (1938). Contributions to the theory of sampling human populations. Journal of the American Statistical Association, 33, 101-116.

Rao, J. N. K. (1973). On double sampling for stratification and analytical surveys. Biometrika, 60,

Robson, D. S. and King, A. J. (1952). Multiple sampling of attributes. Journal of the American Statistical Association, 47, 203-215.

Singh, B. D. and Singh, D. (1965). Some remarks on double sampling for stratification. Biometrika, 52, 587-590.

Srinath, K. P. (1971). Multiphase sampling in non-response problems. Journal of the American Statistical Association, 66, 583-586.

Sven, Berg (1972). Double sampling for cluster - or two-stage sampling. International Statistical Review, 40, 13-22.

Thompson, W. A., Jr. (1962). The problem of negative estimates of variance components. Annals of Mathematical Statistics, 33, 273-289.