CHOICE OF STRATEGIES IN SURVEY SAMPLING

GEETHA RAMACHANDRAN

Thesis submitted to the Indian Statistical Institute
in partial fulfilment of the requirements
for the award of the degree of
Doctor of Philosophy

CALCUTTA

1978

ACKNOWLEDGEMENTS

I am extremely grateful to Dr. T.J.Rao under whose supervision the work reported in this thesis was carried out, for his constant encouragement and the valuable guidance that I have been receiving at every stage of my work. The final version of the thesis owes much to his editing of the manuscript and I am thankful for his advice and the innumerable discussions I have had throughout my research. I also wish to thank Dr. T.J.Rao for his permission to include in the thesis some of our joint work that form chapters 2,3 and 4.

Professor T.V.Hanurav initiated me to work in the field of survey sampling. I am indebted to him for the many valuable discussions and for his encouragement.

I am grateful to Mr. C.H.Sastry, Computer Science Unit, for writing the programmes for some of the illustrations. I wish to thank Mr. C. Narasinga Rao, M.Stat. Computer Science student for his programming of the part of calculations reported in chapter 3.

I wish to thank the authorities of the Indian Statistical Institute for all the facilities extended to carry out my research.

Finally, I thank Mr. Arun Das for his efficient typing and Mr. Apurba Guha for his excellent cyclostyling.

G. Ramachandran

C O N T E N T S

CHAPTER 0

INTRODUCTION

The main problem of sampling from finite populations consists of devising an appropriate procedure for selecting a sample from a given population and developing an appropriate procedure for estimating the population parameter of interest in order to maximize the precision of the estimator subject to certain restrictions on the cost for the survey or alternatively minimize the cost for achieving a given level of precision. During the thirties and forties several contributions were made to develop the theoretical background of sample survey techniques in order to solve this problem. The most significant works of Cochran, Hansen, Hurwitz, Mahalanobis, Neyman, Sukhatme and Yates among others may be mentioned in this context. As well as developing the theoretical techniques, practical techniques for the actual conduct of a survey, data collection and analysis were considered which resulted in large scale sample surveys (Mahalanobis (1944), (1946)).

In most of the sample survey situations information on an auxiliary variate closely related to the study variate is available. Efficient utilisation of this auxiliary information for selection purposes and for estimation gave rise to the varying probability selection method and the theory of ratio and regression estimators and stratification technique (See Neyman (1934), Cochran (1942) and Hansen and Hurwitz (1943)). For reviews on the developments in the theory of sampling from finite populations we refer to Yates (1946), Cochran (1947), Stephen (1948), Seng (1951), Sukhatme((1959),(1966)), Dalenius (1962), Murthy (1963), Vos (1974) and more recently Smith (1976a). Horvitz and Thompson (1952) formulated a systematic theory of sampling from finite populations and defined three classes of estimators. Later in 1955, Godambe proposed a unified theory of sampling from finite populations with a view to discussing the fundamental problems of sampling within this framework. Godambe (1955) established that for any sample design there does not exist a uniformly minimum variance unbiased estimator of the population total in the class of linear unbiased estimators (barring certain exceptions, characterized later by Hanurav (1965)). Since then various criteria such as 'admissibility', 'linear-invariance', 'regularity', 'hyperadmissibility' among others have been putforward for the choice of an optimum estimator. Whenever we have auxiliary information on a characteristic closely related to the study variable it was first shown by Cochran (1946) that this information can be utilised to set up a criterion of optimality of minimum expected variance under suitably defined model. This concept is popularly referred as the 'super-population concept'. In this thesis we work th the general super-population model denoted by $\theta(g)$, ≥ 0 while discussing the choice of an optimum strategy for timating the population total (see chapters 5 and 6). Under stratified set up, when the super-population model $\theta(g)$ is rephrased to apply in a stratified situation, in chapters 2,3 and 7 we discuss the choice between unstratified sampling strategy and stratified strategies with various allocations of sample size to strata.

We give below a brief summary of the author's contributions contained in this thesis.

Following the introductory chapter 0, we give in chapter 1 the basic definitions and explain the concepts that are used in the sequel.

We study in chapter 2 stratified πPS sampling design with various allocations of sample size to strata. The problem of optimum allocation of sample size to strata has been earlier examined in the light of apriori distributions (Hanurav (1965), T.J. Rao (1968)). In this context, under the criterion of minimum expected variance, unstratified πPS sampling strategy was shown to be inferior to stratified πPS sampling strategy with this optimum allocation. It is however not known whether unstratified πPS sampling is still inferior to stratified πPS sampling when one deviates from the optimum allocation. Motivated by this we establish whether for a given allocation stratified πPS sampling should at all be attempted.

In chapter 3 we first derive the optimum allocation of sample size to strata which minimizes the expected variance of the conventional stratified Probability Proportional to Size with Replacement (PPS) sampling strategy. With this optimum allocation we compare the stratified PPS sampling strategy with the unstratified PPS sampling strategy. Next, generalizing Des Raj's (1963) result we also study in the expected variance sense, whether stratified PPS sampling with various non-optimal allocations apart from the X-proportional allocation is likely to be worthwhile. In the last section of the chapter, to achieve considerable gains when stratification is used, a mode of stratification is suggested which is empirically investigated.

Problems considered in chapter 4 have their origin in Royall (1970) wherein the 'ξ-unbiasedness' criterion of a linear estimator of the population total was proposed under a certain 'super-population regression model ξ'. Using the

'ξ-unbiasedness' criterion we first derive in this chapter the class of variance functions in the super-population model for which a given linear estimator is 'optimal' and study its role the construction of 'optimal estimators'. The estimator $\hat{Y}_2$ which is optimal for the variance function $V(x) = x^2$ in our model is found to fare better than some other well known estimators and motivated by this a comparison between any general linear estimator and $\hat{Y}_2$ is presented. A general 'class of Symmetrized Des Raj estimators' of which the estimators suggested by Basu (1970) as alternatives to Murthy's (1957) estimator are particular cases is considered and sufficient conditions are given for selection of the best in this class. A few further generalizations are mentioned.

In chapter 5, an optimal strategy for estimating the population total in the sense of minimum expected variance working with the general super-population model $\theta(g)$ is derived in the class of <u>all</u> p-unbiased strategies with expected sample size fixed.

In chapter 6, we completely characterize the $\theta(g)$-optimal strategies for estimating the population total in the class of equicost p-unbiased strategies. We also show that Cassel *et al.*'s (1976) optimality theorem of the generalized difference estimator is an immediate corollary of our characterization heorem. We discuss the 'Generalized Regression Estimator' or estimating the population total towards the end of the hapter.

We continue our study on p-unbiased strategies that are $\theta(g)$-optimal in chapter 7 and discuss an asymptotically design-unbiased strategy due to Brewer (1977) in the first section. The next two sections of this chapter are devoted to studying stratification and allocation problems working with the model $\theta(g)$ when the population size and the stratum sizes are large. Extending the $\theta(g)$-optimal strategies obtained in chapters 5, 6 and Brewer's asymptotically efficient strategy to stratified set up we study relative efficiencies of unstratified strategy as compared to stratified strategies. The results obtained are illustrated based on live populations some of which were earlier considered by J.N.K. Rao (1969) and Royall (1970).

Numerical examples based on live data are provided to illustrate the results at various stages throughout the thesis. In appendices AI to AVI the complete data for the strations considered are given. A list of references used this thesis is given at the end. The contents of chapter 2 published (J. Roy. Statist. Soc., Ser. B, 36, 1974). The sential contents of chapter 3 were presented at the International mposium on Recent Trends of Research in Statistics, Calcutta, cember '74. An extended version of the same was presented at e 41st session of the International Statistical Institute eld at New Delhi, '77.

CHAPTER 1

CONCEPTS AND DEFINITIONS

1.1 Preliminaries

A collection of units 1,2,...,N is a 'finite population' where N is a known finite number. The units of the finite population are distinguishable. We denote this population by

$$U = \{ 1,2,\ldots,N \}. \tag{1.1.1}$$

N is called the 'size of the population' and a list of units as in (1.1.1) a 'sampling frame'.

$$\text{A sample space } S = \{s\} \tag{1.1.2}$$

is the collection of all samples s from U where

either

> $S = \{s\}$ is the collection of all finite ordered sequences s of units from U in which case
>
> $$s = (i_1, i_2, \ldots, i_n) \tag{1.1.3}$$
>
> is a finite ordered sequence of (not necessarily distinct) units from U

or

> $S = \{s\}$ is the collection of all nonempty subsets s of U in which case s is a nonempty subset of U . (1.1.4)

The number of units in a sample s is called the sample size of s and is denoted by $n(s)$. Thus, if s is a sequence of units from U, $n(s)$ denotes the length of the sequence s and if s is a subset of U, $n(s)$ is the number of units in s.

We consider a real valued variable $\mathcal{Y}$ (study variable $\mathcal{Y}$) defined over U and taking value y_i on unit i, $1 \leq i \leq N$. Let $\underset{\sim}{Y}$ denote the vector

$$\underset{\sim}{Y} = (y_1, y_2, \ldots, y_N) \qquad (1.1.5)$$

The y_i's are unknown apriori and our problem in general is to estimate real valued functions of $\underset{\sim}{Y}$, the parameter of interest which is assumed to be a point in R_N (the N-dimensional Euclidean space), called parametric functions, on the basis of the observations y_i for $i \in s$ where s is a sample drawn with given probability p_s from the totality of all possible samples S. We call the function p on S defining the probabilities p_s for all samples in S, such that for every $s \in S$, $p_s \geq 0$ and $\sum_{s \in S} p_s = 1$, a 'sampling design' (or simply a 'design' or 'sampling plan') and denote it by

$$d = (S, p). \qquad (1.1.6)$$

However, it is never possible in practice to form all possible samples, and choose one from them at random with the probabilities prescribed by a sampling design p. But Hanurav (1962) demonstrated that every sampling design p can be implemented by some practically feasible sampling procedure, such as drawing units one after another, at random with varying probabilities.

Considering the problem of estimation, any real valued function t defined over a design $d = (S, p)$ such that for samples $s \in S$, the function $t_s(\underset{\sim}{Y})$ depends only on the values of $\mathcal{Y}$ for the units in the sample is called an estimator. We denote $t_s(\underset{\sim}{Y})$ by t_s for $s \in S$ when no confusion can occur. An estimator t when used for estimating a parametric function T is called an estimator of T.

An estimator t of T is called an 'unbiased estimator' if

$$E(p : t) = \sum_{s \in S} t_s \, p_s = T \quad \text{for all } \underset{\sim}{Y} \text{ in } R_N. \qquad (1.1.7)$$

When t is an unbiased estimator of T say that t is unbiased for T or t is p-unbiased for T (i.e., t is unbiased for T with respect to p).

An estimator that is not unbiased is called a 'biased estimator' and its bias is denoted by

$$B(t) = E(t) - T . \qquad (1.1.8)$$

The sampling mean square error (or mean square error) of an estimator t of T is

$$M(p : t) = E(t - T)^2 \tag{1.1.9}$$

When t is unbiased for T, then $M(p : t)$ is the variance (sampling variance) of t, in which case

$$\text{Var.}(p : t) = \sum_{s \in S} t_s^2 \, p_s - T^2. \tag{1.1.10}$$

We use E(t), M(t), Var.(t) to denote $E(p : t)$, $M(p : t)$, Var.$(p : t)$ respectively when from the context it is clear as to what the design p is.

A design $d = (S, p)$ together with an estimator t of a parametric function T defined over d is called a 'sampling strategy' (or a 'strategy') for the estimation of T. Thus, a strategy is denoted by

$$H = (p : t). \tag{1.1.11}$$

A strategy $H = (p : t)$ when used to estimate a parametric function T is said to be unbiased for T if t is p-unbiased for T. If t is not p-unbiased for T the strategy H is said to be biased for estimating T.

The mean square error M(H) or the variance Var.(H) of a strategy for estimating T are defined to be the mean square error or the variance of the corresponding estimator.

Given a design $d = (S, p)$ and unbiased estimators t_1 $_2$ of the same parametric function T, both defined over is said to be 'better than' t_2 if

$$Var.(p : t_1) \leq Var(p : t_2), \quad \text{for all } \underset{\sim}{Y} \qquad (1.1.12)$$

strict inequality holding at least once.

For a given design d and a class L_1 of estimators of parametric function T, all defined over d, when every member L_1 is unbiased for T, a member t_1 is called the 'best' L_1 if it is better than every other member of L_1.

Thus, an unbiased strategy H_1 is said to be 'better than' other unbiased strategy H_2 for estimating T, if

$$Var.(H_1) \leq Var.(H_2), \text{ for all } \underset{\sim}{Y} \qquad (1.1.13)$$

with strict inequality holding at least once.

Given a class L_2 of unbiased strategies for estimating T, a member H_1 in the class L_2 is called 'best' if it is better than every other member of L_2.

With respect to a design $d = (S, p)$ a parametric function T is <u>estimable</u> if there exists an estimator t defined over d, that is unbiased for T.

Given a design $d = (S, p)$, let

$$\pi_i = \sum_{s \ni i} p_s \ , \quad 1 \leq i \leq N \tag{1.1.14}$$

and

$$\pi_{ij} = \sum_{s \ni i,j} p_s \ , \quad 1 \leq i \neq j \leq N \tag{1.1.15}$$

where in (1.1.14) the summation in the right hand side is over all samples that contain the unit i and in (1.1.15) the sum is over all samples that contain the units i and j. The π_i's and π_{ij}'s are called the first and second order (joint) inclusion probabilities with respect to the design p.

The particular parametric function $T = \sum_{i=1}^{N} y_i$ is called the population total. We use Y to denote the population total $\sum_{i=1}^{N} y_i$. It was proved by Hanurav (1965) that a necessary and sufficient condition for the estimability of the population total Y is that $\pi_i > 0$ for $i = 1,2,\ldots,N$ and that a necessary and sufficient condition for the estimability of the variance of an unbiased estimator t of Y is that $\pi_{ij} > 0$ for $1 \leq i \neq j \leq N$.

Given a design $d = (S, p)$, Godambe (1955) defined the 'general homogeneous linear estimator' t for estimating the population total Y by

$$t'_s = \sum_{\epsilon} \beta \quad y_i \tag{1.1.16}$$

for every s in S where $\sum_{i \varepsilon s}$ indicates that the summation is taken over all the units contained in the sample s. The coefficients β_{si}'s attached to the variate values depend on the sample as well as the units to which they are attached, but not on the variate values y_i's. Godambe (1955) then proved

Theorem 1.1.1 (Godambe (1955)) : With respect to any design d, there does not exist a best unbiased estimator of the population total Y in the class of general homogeneous linear unbiased estimators of Y.

Hanurav (1966) pointed out some exceptions to this and gave some non-trivial designs where a best estimator exists.

Godambe's negative result mentioned in Theorem 1.1.1 necessitated the weeding out of 'bad' estimators and the criterion of admissibility was introduced in this connection.

Definition 1.1.2 : With respect to a given design $d = (S,p)$, an estimator t_1 belonging to a class L_3 of unbiased estimators of Y is said to be 'admissible' in L_3 if there does not exist any other estimator in L_3 which is better than t_1 ; i.e., given any $t_2(\neq t_1) \varepsilon L_3$, there exists at least one point Y^* in R_N such that

$$\text{Var.}(p : t_1) < \text{Var.}(p : t_2) \qquad (1.1.17)$$

where both the variances in (1.1.17) are evaluated at $\underset{\sim}{Y} = \underset{\sim}{Y}^*$.

The criterion 'admissibility' though helps in eliminating ad estimators, does not help much in restricting to optimum estimators. The results of Basu (1958) and Roy and Chakravarty (1960) in this connection establish inadmissibility of certain customary estimators. Godambe and Joshi (1965) and Joshi (1965 (1965b) considered admissibility removing the restriction of linearity and later relaxing unbiasedness. Other criteria considered in the literature are 'linear invariance' and 'regular estimators' by Roy and Chakravarty (1960), 'hyper-admissibility' by Hanurav (1965) and 'necessary bestness' by Prabhu Ajgaonkar (1965). However, these criteria are not found atisfactory since they refer to particular cases.

1.2 Optimum utilisation of auxiliary information.

For estimating the population total Y, if uniform minimization of the variance of a strategy is the criterion of optimality, we have seen that there does not exist one that is optimum (Godambe (1955), Godambe and Joshi (1965)). However, auxiliary information on a variate $\mathcal{X}$ closely related to the study variable y, taking values x_i on units i, $1 \leq i \leq N$ is often available for all the population units in practical sample survey situations. This information can be utilised to get an optimum sampling strategy to estimate $Y = \sum_{i=1}^{N} y_i$, the population total. Frequently x_i (the value of $\mathcal{X}$, an essentially positive valued characteristic, on the population unit i) is also referred to as the size of unit i, $1 \leq i \leq N$. The terminology derives from the fact that often x_i is a measure of the size of unit i, as for example when the units are farms and x_i is the acreage of ith farm. Here the value of $\underset{\sim}{Y} = (y_1, y_2, \ldots, y_N)$ is supposed to be the realisation of an N-length random vector whose distribution θ depends on $\underset{\sim}{X} = (x_1, x_2, \ldots, x_N)$ and some unknown parameters. This concept is called Super-population concept (Cochran (1946)). We shall use the notation y_i to denote the random variable as well as the value it takes on unit i, given a finite population U for $1 \leq i \leq N$. The information on $\mathcal{X}$, known beforehand can thus be used to assume a reasonable apriori distribution of $\underset{\sim}{Y}$. We denote the expectation, variance and covariance with respect to the distribution θ by E_θ, $Var._\theta$, $Cov._\theta$.

Definition 1.2.1 : Given a prior distribution θ of $\underset{\sim}{Y}$, an unbiased strategy $H_1 = (p_1 : t_1)$ is said to be 'better (in the expected variance sense) than ' another unbiased strategy $H_2 = (p_2 : t_2)$ for estimating Y if

$$E_{\theta} \text{ Var.}(H_1) \leq E_{\theta} \text{ Var.}(H_2) \text{ holds.} \qquad (1.2.1)$$

ien H_1 is better than H_2 for estimating Y, say that H_2 'worse than' H_1.

When strict inequality holds in (1.2.1) say that H_1 is uperior' to H_2 for estimating Y in the expected variance nse (or H_2 is 'inferior' to H_1).

If equality holds in (1.2.1) say that H_1 and H_2 are 'equivalent' (in the expected variance sense) for estimating Y. We also say in this case that H_1 is 'as good as' H_2 for estimating Y.

Definition 1.2.2 : Given a prior distribution θ of $\underset{\sim}{Y}$, in e class D(H) of all unbiased strategies for estimating Y, say that H_o is " θ -optimum" in D(H) if it is better than any other strategy in D(H). If H_o is θ -optimum for estimating Y in D(H) for every θ in $\textcircled{H}$, a class of prior distributions, say that H_o is " $\textcircled{H}$-optimum" in D(H) for estimating Y.

Analogously " θ -optimality" and " $\textcircled{H}$-optimality" of a strategy H_o for estimating Y in subclasses of D(H) are defined.

It is important to note that the assumption of the existence of a prior distribution θ is relevant only for a proper choice of the strategy for estimating Y and that the ultimate inferen about Y by means of an estimator of Y together with an estimated variance of this estimator of Y depends completely o the basis of the observations. Thus here the optimality of the strategy adopted for estimating Y is the one dependent on the prior distribution θ.

Let $\theta(g)$ be the prior distribution of $\underset{\sim}{Y}$ which is furth specified by the model

$$\left.\begin{aligned} & E_{\theta(g)}\ (y_i \mid x_i) = a x_i\ , \\ & Var._{\theta(g)}\ (y_i \mid x_i) = \sigma^2\, x_i^g\ , \\ & Cov._{\theta(g)}\ (y_i, y_j \mid x_i, x_j) = 0,\quad 1 \leq i \neq j \leq N \end{aligned}\right\} (1.2.2$$

where a and σ^2 are unknown parameters of the prior distributio $\theta(g)$. In practice, the value of g is found to be non-negative and lying between 1 and 2 more often. This has been borne out in empirical studies by Mahalanobis (1944), Smith (1938). This model (1.2.2) has been further studied by Brewer(1963), Foreman and Brewer (1971) and more recently by Brewer et al. (1977).

In what follows unless otherwise mentioned, a design $d = (S, p)$ refers to the case $S = \{s\}$, the collection of all non-empty subsets of $U = \{1, 2, \ldots, N\}$. Recall that for every $s \in S$, $n(s)$, the sample size of s is the number of units in s

Definition 1.2.3 : A design d = (S, p) is said to be a 'fixed sample size (=n) design' if for every $s \in S$, $p_s > 0 \Rightarrow n(s) = n$. A strategy H = (p : t) is said to be a fixed sample size strategy if p is a fixed sample size design.

Given a design d = (S, p) the 'Horvitz-Thompson estimator' (HT estimator) $\hat{Y}_{HT}$ of the population total Y is given by

$$\hat{Y}_{HT} = \sum_{i \in s} \frac{y_i}{\pi_i} \tag{1.2.3}$$

for every $s \in S$ where π_i is the inclusion probability of unit i with respect to p.

Observe that $\hat{Y}_{HT}$ is a homogeneous linear p-unbiased estimator of Y where p is such that $\pi_i > 0$ for $i = 1, 2, \ldots, N$.

Given a design d = (S, p) a 'cost function' which is reasonable is given by

$$c_s = c_o + c_1 n(s) \tag{1.2.4}$$

for every s in S where c_s is the cost for a sample s. This consists of the overhead cost c_o and $c_1 n(s)$, the cost of collecting the data (the values of the study variable $\mathcal{Y}$ on units in s) which is assumed to be proportional to the size s (i.e. n(s)) where c_1 stands for the cost of collection data on a single unit. The cost of a strategy $H = (p : t)$ the 'expected cost of the design p' and is given by

$$c(H) = c_o + c_1 \, E(n(s)) \tag{1.2.5}$$

ere $E(n(s)) = \sum_{s \in S} n(s) \, p_s$ is the expected sample size of e design $d = (S, p)$.

Here
$$E(n(s)) = \sum_{s \in S} \sum_{i \in s} p_s = \sum_{i=1}^{N} \sum_{s \ni i} p_s = \sum_{i=1}^{N} \pi_i .$$

Observe that under this set up, two strategies are 'equally costly' if, and only if, they have the same expected sample size. In this thesis, while discussing the choice of an optimum strategy for estimating Y given a prior distribution θ of $\underset{\sim}{Y}$ we consider the class of strategies with $E(n(s)) = n$, a given number (i.e., the class of strategies with expected sample size fixed).

Definition 1.2.4 : If a design $d = (S, p)$ with $\pi_i > 0$ for $i = 1, 2, \ldots, N$ is such that the first order inclusion probabilities with respect to the design p are proportional

x_i where x_i is the value of an auxiliary variate $\mathcal{X}$ on ith unit of U $(x_1, x_2, \ldots, x_N$ all known), i.e.,

$$\pi_i \propto x_i \quad \text{for } i = 1, 2, \ldots, N$$

that the design p is a 'πPS sampling design' (or simply, PS design').

Observe that $\pi_i = n \frac{x_i}{X}$ where n is the expected sample of the design p and $X = \sum_{i=1}^{N} x_i$ and that for a πPS ign to be available, the expected sample size $E(n(s)) = n$ uld be such that

$$n \leq \frac{X}{\max_i \{x_i\}} .$$

Consider now $\textcircled{H} = \{\theta(g)\}$ where $\theta(g)$ is such that the del (1.2.2) holds.

Godambe (1955) proved the following

orem 1.2.5 (Godambe (1955)) : If n is an integer ($n < \frac{X}{\max_i \{x_i\}}$) any strategy $H_o = (p_o : \hat{Y}_{HT})$ where the design p_o is such that

$$\left.\begin{array}{ll}(a) & \pi_i = n\frac{x_i}{X} \quad \text{for } i = 1,2,\ldots,N \quad \text{where} \\ & X = \sum_{i=1}^{N} x_i \\ (b) & n(s) = n \quad \text{for all } s \text{ with } p_s > 0\end{array}\right\} \quad (1.2.6$$

$\theta(2)$-optimum in the class of all linear, fixed sample size, unbiased strategies of the population total Y.

Godambe and Joshi (1965) proved that the strategy $= (p_o : \hat{Y}_{HT})$ where p_o is such that (1.2.6) holds is $\theta(2)$-optimum in the class of all, fixed sample size (=n), unbiased strategies of the population total Y.

In this thesis, we consider the general model $\theta(g)$ and study the problem of comparison of sampling strategies and choice of an optimum strategy under $\theta(g)$ for the estimation of the population total Y.

It is a common practice among survey practitioners to use auxiliary information either at the design stage or at the estimation stage or both for increased precision. Examples of designs that use the auxiliary information are stratified sampling designs, varying probability sampling designs such as PPS and πPS designs, systematic sampling designs etc. Many estimators incorporating the auxiliary information are available the literature as for example ratio estimator, regression imator, the HT-estimator etc. In the sequel we study, ng stratified sampling designs, allocation of sample size strata and comparison of strategies under the model $\theta(g)$.

CHAPTER 2

ALLOCATION OF SAMPLE SIZE AND RELATIVE EFFICIENCES OF STRATIFIED AND UNSTRATIFIED πPS SAMPLING DESIGNS

2.0 Summary

When a population of size N is divided into k strata of sizes N_i, $i = 1,2,\ldots,k$ defined by k non-overlapping ranges of values of $\mathcal{X}$, the auxiliary variate closely related to the variate $\mathcal{Y}$ under study, the problem of optimum allocation of sample size to the strata for the estimation of the population total Y, has been examined in the light of apriori distributions by Hanurav (1965) and T.J. Rao (1968). In this context, under the criterion of expected variance, the sampling strategy consisting of unstratified πPS sampling design together with the Horvitz-Thompson estimator was shown to be worse than the strategy consisting of stratified πPS sampling design with the corresponding HT-estimator with this optimum allocation. In this chapter, we study whether stratified πPS sampling strategy with various non-optimal allocations is likely to be worthwhile and whether it should at all be attempted. For populations commonly met in practice, we derive sufficient conditions for unstratified πPS sampling to be preferable to non-optimal stratified πPS sampling. Illustrative examples based on live data are provided to enlighten the results.

$$\hat{Y}_S = \sum_{i=1}^{k} \hat{Y}_i$$

where $\hat{Y}_i$ is an unbiased estimator of Y_i based on a design $p_{(i)}$ used to draw a sample from the N_i units of the ith stratum. If further, sampling is done independently in each stratum for estimating Y_i, we have

$$\text{Var.(stratified sampling design : } \hat{Y}) = \sum_{i=1}^{k} \text{Var.}(p_{(i)} : \hat{Y}_i) .$$

We consider in this chapter the stratified sampling design composed of a πPS sampling design with fixed sample size n_i to draw a sample of n_i units from N_i units in the ith stratum for $i = 1,2,\ldots,k$. As an estimator of Y_i, we consider the HT-estimator introduced above based on this πPS sampling design with fixed sample size n_i in the ith stratum for $i = 1,2,\ldots,k$. We describe below in detail our notations and strategies that are studied in this chapter.

Consider a finite population of size N. Values of $\mathcal{X}$ (an essentially positive auxiliary character closely related to the character $\mathcal{Y}$ under study) are available for all units, and the population is divided into k strata of sizes N_i, $i = 1,2,\ldots,k$, defined by k non-overlapping ranges of values of $\mathcal{X}$, i.e.,

$$\hat{Y}_S = \sum_{i=1}^{k} \hat{Y}_i$$

where $\hat{Y}_i$ is an unbiased estimator of Y_i based on a design $p_{(i)}$ used to draw a sample from the N_i units of the ith stratum. If further, sampling is done independently in each stratum for estimating Y_i, we have

$$\text{Var.(stratified sampling design} : \hat{Y}) = \sum_{i=1}^{k} \text{Var.}(p_{(i)} : \hat{Y}_i) .$$

We consider in this chapter the stratified sampling design composed of a πPS sampling design with fixed sample size n_i to draw a sample of n_i units from N_i units in the ith stratum for $i = 1,2,\ldots,k$. As an estimator of Y_i, we consider the HT-estimator introduced above based on this πPS sampling design with fixed sample size n_i in the ith stratum for $i = 1,2,\ldots,k$. We describe below in detail our notations and strategies that are studied in this chapter.

Consider a finite population of size N. Values of $\mathcal{X}$ (an essentially positive auxiliary character closely related to the character $\mathcal{Y}$ under study) are available for all units, and the population is divided into k strata of sizes N_i, $i = 1,2,\ldots,k$, defined by k non-overlapping ranges of values of $\mathcal{X}$, i.e.,

$N = \sum_{i=1}^{k} N_i$. For convenience we label the strata, and the units within the strata, in ascending order of $\mathcal{X}$, so that if x_{ij}, y_{ij} are the values of $\mathcal{X}$, $\mathcal{Y}$ respectively for the jth unit of the ith stratum, then

$$0 < x_{11} \leq x_{12} \leq \cdots \leq x_{1N_1} \leq x_{21} \leq \cdots\cdots \leq x_{kN_k} .$$

Let a πPS design with fixed sample size ($=n_i$) be used to draw a sample of size n_i from the ith stratum for $i = 1,2,\ldots,k$ (Hanurav (1967), Vijayan (1968) and others). The sample sizes n_i's are such that $\sum_{i=1}^{k} n_i = n$, the total sample size. Let sampling be carried out independently within each stratum. Let $\pi_{(i)j}$ denote the probability of inclusion of the jth unit of the ith stratum in the sample, given by $\pi_{(i)j} = n_i \, x_{ij}/X_i$, where $X_i = \sum_{j=1}^{N_i} x_{ij}$ is the total of the $\mathcal{X}$-values in the ith stratum. We shall use the notation $\sum_i$ to denote $\sum_{i=1}^{k}$, summation over all strata and for a given i, $\sum_j$ to denote $\sum_{j=1}^{N_i}$, summation over all population units in the ith stratum for $i = 1,2,\ldots,k$. As an estimator of the population total $Y = \sum_i \sum_j y_{ij}$, consider the Horvitz-Thompson estimator introduced earlier

$$\hat{Y}_S = \sum_{i=1}^{k} \left(\sum_j{}' \, (y_{ij}/\pi_{(i)j} \right)$$

$$= \sum_{i=1}^{k} \left(\sum_j{}' \, (y_{ij}/(n_i \, x_{ij}/X_i)) \right), \qquad (2.1.1)$$

where for a given i, Σ'_j denotes summation over the sampled units in the ith stratum. Next let a πPS design with fixed sample size n be used to draw a sample of size n from the whole population (unstratified). As an estimator of the population total Y based on this πPS design with fixed sample size n consider

$$\hat{Y}_U = \sum_{i=1}^{k} \sum_j{}' \, (y_{ij} \,/\, \pi'_{(i)j}) \,, \qquad (2.1.2)$$

where Σ'_j now denotes summation over those units out of the sampled n that belong to the ith stratum and $\pi'_{(i)j}$ are the probabilities of inclusion of the units given by $\pi'_{(i)j} = n\, x_{ij}/(\sum_i \sum_j x_{ij}) = n\, x_{ij}/X$ where $X = \sum_i X_i$ (We assume that n, the values x_{ij} and the stratification adopted are such that values n_i can be chosen so that none of the $\pi_{(i)j}$ or $\pi'_{(i)j}$ exceeds unity).

Following Cochran (1946) we regard $\underset{\sim}{Y} = (y_{11}, y_{12}, \ldots, y_{kN_k})$ as a realization of an N-length random vector with distribution $\theta(g)$ depending on $\underset{\sim}{X} = (x_{11}, x_{12}, \ldots, x_{kN_k})$ and some unknown parameters.

Given $\underset{\sim}{X}$ we explicitly formulate our model $\theta(g)$ thus

$$\left.\begin{aligned} &E_{\theta(g)}\,(y_{ij}|x_{ij}) = a\,x_{ij} \\ &\text{Var}._{\theta(g)}\,(y_{ij}|x_{ij}) = \sigma^2\,x_{ij}^g \\ &\text{Cov}._{\theta(g)}\,(y_{ij},y_{rh}|x_{ij},x_{rh}) = 0,\ 1 \le i,\ r \le k \\ &\qquad\qquad\qquad\qquad\qquad\qquad j \ne h \text{ if } i=r \end{aligned}\right\}\quad(2.1.3)$$

where $E_{\theta(g)}$, $\text{Var}._{\theta(g)}$ and $\text{Cov}._{\theta(g)}$ denote the expectation, variance and covariance with respect to the prior distribution $\theta(g)$, a and σ^2 are unknown parameters of the distribution $\theta(g)$ of $\underset{\sim}{Y}$. Under this model (2.1.3) with $g \in [1,2]$, we now compare the two strategies, the stratified πPS sampling design together with the estimator $\hat{Y}_S$ and unstratified πPS sampling design together with the estimator $\hat{Y}_U$ in the expected variance sense. Observe that we are comparing two equicost unbiased strategies of Y. Also note that the model (2.1.3) is the same as the model (1.2.2) rephrased to apply in a stratified situation.

Considering the HT-estimator $\hat{Y}_S$ defined in (2.1.1) we have $\hat{Y}_S = \sum\limits_i \hat{Y}_i$ where $\hat{Y}_i$ is an estimator of Y_i, the total of $\mathscr{Y}$-values of the ith stratum. Since sampling is carried out independently in each stratum, using the variance expression given at the beginning of this section we have

$$\text{Var.}(\hat{Y}_S) = \sum_i \text{Var.}\left(\sum_j{}' \frac{y_{ij}}{\pi_{(i)j}} \right)$$

$$= \sum_{i=1}^{k} \left\{ \sum_{j=1}^{N_i} (\pi^{-1}_{(i)j} - 1) y^2_{ij} + \sum_{j \neq h}^{N_i} \sum^{N_i} (\pi_{(i)jh} \, \pi^{-1}_{(i)j} \, \pi^{-1}_{(i)h} - 1) y_{ij} \, y_{i}\right.$$

where $\pi_{(i)jh}$ is the probability of joint inclusion of the jth and hth units of the ith stratum in the sample (a πPS design with fixed sample size n_i is used for drawing a sample in the ith stratum). Further, under the model $\theta(g)$ of (2.1.3) we have

$$E_{\theta(g)} \text{Var.}(\hat{Y}_S) = \sum_{i=1}^{k} \left\{ \sum_{j=1}^{N_i} (\pi^{-1}_{(i)j} - 1) \; (a^2 \, x^2_{ij} + \sigma^2 \, x^g_{ij}) \right.$$

$$\left. + \sum_{j \neq h}^{N_i} \sum^{N_i} (\pi_{(i)jh} \, \pi^{-1}_{(i)j} \, \pi^{-1}_{(i)h} - 1) \; a^2 x_{ij} \, x_{ih} \right\}$$

$$= \sigma^2 \sum_{i=1}^{k} \sum_{j=1}^{N_i} (\pi^{-1}_{(i)j} - 1) \; x^g_{ij} + a^2 \, \text{Var.}(\hat{X}_S),$$

where $\hat{X}_S$ is obtained by replacing y_{ij} in (2.1.1) by x_{ij}. Clearly, $\text{Var.}(\hat{X}_S) = 0$ (since $\hat{X}_S = \sum_{i=1}^{k} \sum_j{}' \frac{x_{ij}}{\pi_{(i)j}} = \sum_{i=1}^{k} X_i = X$), so that

$$E_{\theta(g)} \text{Var.}(\hat{Y}_S) = \sigma^2 \sum_{i=1}^{k} \sum_{j=1}^{N_i} (\pi^{-1}_{(i)j} - 1) \; x^g_{ij} \qquad (2.1.4)$$

where $\pi_{(i)j} = n_i \, x_{ij}/X_i$.

Similarly, considering the corresponding expression for he variance of $\hat{Y}_U$ we have under the model (2.1.3)

$$E_{\theta(g)} \text{Var.}(\hat{Y}_U) = \sigma^2 \sum_{i=1}^{k} \sum_{j=1}^{N_i} (\pi'^{-1}_{(i)j} - 1)\, x^g_{ij} \qquad (2.1.5)$$

where $\pi'_{(i)j} = n\, x_{ij}/X$.

Now consider

$$f(g) = E_{\theta(g)} \left\{ \text{Var.}(\hat{Y}_S) - \text{Var.}(\hat{Y}_U) \right\} / \sigma^2$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{N_i} (\pi^{-1}_{(i)j} - \pi'^{-1}_{(i)j})\, x^g_{ij}$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{N_i} (n_i^{-1} X_i - n^{-1} X)\, x^{g-1}_{ij}$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{N_i} a_i\, x^{g-1}_{ij} \qquad (2.1.6)$$

where $a_i = n_i^{-1} X_i - n^{-1} X$.

The problem of allocation of sample size to strata has been earlier examined in the light of apriori distributions by Hanurav (1965) and T.J.Rao (1968). It was also shown in T.J.Rao (1968) that under the model $\theta(g)$ of (2.1.3) with $g \in [1,2]$, allocation of sample size to strata which minimizes (2.1.4) is given by the '$\theta(g)$-optimum allocation''

$$n_i = n(X_i \sum_{j=1}^{N_i} x_{ij}^{g-1})^{\frac{1}{2}} / (\sum_{i=1}^{k} (X_i \sum_{j=1}^{N_i} x_{ij}^{g-1})^{\frac{1}{2}}). \qquad (2.1.7)$$

From this it is easy to see that when $g = 2$, "$\theta(2)$-optimum allocation" reduces to allocation proportional to the stratum totals of the x_{ij}'s. We next have

Theorem 2.1.1 (T.J.Rao(1968)) : In the sense of expected variance, under $\theta(g)$, for estimating Y unstratified πPS sampling strategy is inferior to stratified πPS sampling strategy with $\theta(g)$-optimum allocation for $g \varepsilon [1,2)$ and when $g = 2$, both the strategies are equivalent.

It is, however, not known under what conditions unstratified πPS sampling strategy is still inferior to stratified πPS sampling strategy when one deviates from the "$\theta(g)$-optimum allocation". With this in mind, we consider whether stratified πPS sampling with various non-optimal allocations is likely to be worthwhile and whether it should at all be attempted.

2.2 Main Results

We first prove in this section, the following

Theorem 2.2.1 : Let $0 < x_{11} \leq x_{12} \leq \ldots \leq x_{1N_1} \leq x_{21} \leq \ldots \leq x_{kN_k}$ and the allocation $\underset{\sim}{n} = (n_1, n_2, \ldots, n_k)$ of sample size be such that $a_i = n_i^{-1} X_i - n^{-1} X$, $i = 1,2,\ldots,k$, are non-decreasing and not all equal where $X_i = \sum_j x_{ij}$. Further, let $a_i \leq 0$ for $i \leq t$ and $a_i > 0$ for $i > t$ and not all x_{ij}'s for $i > t$ are equal to x_{tN_t} . Let $f(g) = \sum_{i=1}^{k} \sum_{j=1}^{N_i} a_i x_{ij}^{g-1}$.

Then

(a) if $f(1) < 0$, there exists a unique g_o in the interval $(1,2]$ such that $f(g) \leq 0$ or > 0 according as $g \leq g_o$ or $> g_o$,

(b) if $f(1) \geq 0$, $f(g) > 0$ for all g in $(1,2]$.

<u>Proof</u> : Let $h(g) = \sum_{i=1}^{k} \sum_{j=1}^{N_i} a_i Z_{ij}^{g-1}$ where $Z_{ij} = x_{ij}/x_{tN_t}$

Note that

$$\frac{dh(g)}{dg} = h'(g) = \sum_{i=1}^{k} \sum_{j=1}^{N_i} a_i Z_{ij}^{g-1} \log Z_{ij}$$

$$= \sum_{i=1}^{t} \sum_{j=1}^{N_i} a_i Z_{ij}^{g-1} \log Z_{ij} + \sum_{i=t+1}^{k} \sum_{j=1}^{N_i} a_i Z_{ij}^{g-1} \log Z_{ij}$$

$$\geq 0 \quad \text{(since all terms are } \geq 0\text{)}$$

with strict inequality when not all x_{ij}'s for $i > t$ are equal to x_{tN_t} . Thus $h(g)$ is increasing with g provided not all x_{ij}'s for $i > t$ are equal to x_{tN_t} .

Next we have

$$f(2) = \sum_{i=1}^{k} \sum_{j=1}^{N_i} (n_i^{-1} X_i - n^{-1} X)\, x_{ij}$$

$$= \sum_{i=1}^{k} X_i^2 \; n_i^{-1} - \{(\sum_i X_i)^2/n\}$$

$$\geq \; 0$$

(by Cauchy-Schwarz inequality), equality occurring when and only when X_i/n_i, for $i = 1,2,\ldots,k$ are all equal.

Now observing that $h(g) = f(g).x_{tN_t}^{1-g}$, it follows that $h(2) \geq 0$: also $h(1) = f(1)$. Hence it follows that when $f(1) < 0$, there exists a unique g_o in the interval $(1,2]$ such that $h(g) \leq 0$ or > 0, and so $f(g) \leq 0$ or > 0 according as $g \leq g_o$ or $> g_o$, and when $f(1) \geq 0$, $h(g) > 0$ and so, $f(g) > 0$ for all g in $(1,2]$.

<u>Corollary 2.2.2</u> : If $f(2) = 0$, then $f(g) = 0$ for all g.

<u>Corollary 2.2.3</u> : When $g = 1$,

$$f(1) = \left(\sum_{i=1}^{k} \frac{N_i X_i}{n_i}\right) - \frac{NX}{n} = \begin{cases} 0 & \text{for } n_i \propto N_i, \text{ i.e. proportional allocation,} \\ 0 & \text{for } n_i \propto X_i, \text{ i.e. allocation proportional to stratum totals,} \\ (k^2/n)\,\mathrm{Cov.}(N_i, X_i) & \text{for equal allocation or for } n_i \propto N_i X_i, \text{ which is} < 0 \text{ when } N_i \text{ decreases as } X_i \text{ increases,} \\ n^{-1}\left[\left\{\sum_{i=1}^{k} \sqrt{N_i X_i}\right\}^2 - NX\right] & \text{for } \theta(1)\text{-optimum allocation which is} < 0 \text{ (cf.(2.1.7)).} \end{cases}$$

When $g = 2$,

$$f(2) = \sum_{i=1}^{k} X_i^2 n_i^{-1} - X^2 n^{-1} = \begin{cases} 0 & \text{for } \theta(2)\text{-optimum allocation,} \\ > 0 & \text{for any other allocation, provided not all } X_i/n_i, \\ & i = 1,2,\ldots,k \text{ are equal.} \end{cases}$$

Remark 2.2.4 : Working with an unstratified population, when the sample size is fixed to be 2 for estimating the population total Y, Vijayan (1966) compares the strategy (πPS : $\hat{Y}_{HT}$) with the strategy (p_{SDR} : $\hat{Y}_{SDR}$) under the model $\theta(g)$ of (1.2.2) for $g \in [1,2]$. The strategy (p_{SDR} : $\hat{Y}_{SDR}$) for fixed sample size 2 is the well known Symmetrized Des Raj strategy (Murthy (1957)) when the sample size is fixed to be 2 wherein p_{SDR} is the design with $p_s = \frac{p_i\, p_j}{1 - p_i} + \frac{p_i\, p_j}{1 - p_j}$ where s is the sample with units i and j and $p_i = x_i/X$ for $i = 1,2,\ldots,N$. The Symmetrized Des Raj estimator $\hat{Y}_{SDR}$ is given by

$$\hat{Y}_{SDR} = \frac{1}{2 - p_i - p_j} \left\{ \frac{y_i}{p_i} (1 - p_j) + \frac{y_j}{p_j} (1 - p_i) \right\}$$

for $s \in S$ with $p_s > 0$ where s is the sample with units i and j. Vijayan (1966) observes that with fixed sample size 2 when $g = 1$, the Symmetrized Des Raj strategy is superior to the (πPS : $\hat{Y}_{HT}$) strategy under $\theta(g)$ of (1.2.2) in the expected variance sense. At $g = 2$, with fixed sample size 2, $(\pi PS : \hat{Y}_{HT})$ strategy is indeed better than the Symmetrized Des Raj strategy under $\theta(g)$ (Godambe(1955), Godambe and Joshi (1965)). We note that the uniqueness of g_o in Vijayan (1966) below which (i.e. for $g < g_o$) the strategy $(p_{SDR} : \hat{Y}_{SDR})$ is superior to $(\pi PS : \hat{Y}_{HT})$ and above which $(g > g_o)$ the strategy $(\pi PS : \hat{Y}_{HT})$ is superior to the Symmetrized Des Raj strategy can be established on similar lines as in Theorem 2.2.1 above.

2.3 Interpretation of the Results

Consider a stratification of the population based on the auxiliary information $\mathcal{X}$ such that

$$0 < x_{11} \leq x_{12} \leq \cdots \leq x_{1N_1} \leq x_{21} \leq \cdots \leq x_{kN_k}$$

(cf. section 2.1). It is then possible to consider an allocation $\underset{\sim}{n} = (n_1, n_2, \ldots, n_k)$ for which X_i/n_i (equivalently a_i's) for $i = 1, 2, \ldots, k$ are non-decreasing and not all equal, where X_i is the total of the $\mathcal{X}$-values of the ith stratum. For example, if the stratification is such that a large number of units with small $\mathcal{X}$-values are grouped in the former strata and a small number of units with large $\mathcal{X}$-values are grouped in the latter strata and X_i's are non-decreasing, then this nature of the stratification might suggest an allocation away from the optimum with n_i's decreasing, thereby implying that X_i/n_i's are non-decreasing and not all equal. Interpretation of the results of section 2.2 would now enable us to study the efficiency of unstratified sampling as compared to stratification with such non-optimal allocations for which a_i's are non-decreasing and not all equal.

Part (a) of Theorem 2.2.1 implies that whenever $(\sum_{i=1}^{k} \frac{N_i X_i}{n_i}) - \frac{NX}{n} < 0$, there exists a value g_0 of g such that the stratified πPS sampling strategy with the given allocation $\underset{\sim}{n}$ is superior (i.e. $f(g) < 0$) or inferior (i.e., $f(g) > 0$) under $\theta(g)$ for estimating Y to unstratified πPS sampling strategy of fixed sample size n according as $g < g_0$ or $g > g_0$. At this value $g = g_0$, stratification is as good as unstratified sampling. Furthermore, whenever $(\sum_{i=1}^{k} \frac{N_i X_i}{n_i}) - \frac{NX}{n} \geq 0$, by part (b) of Theorem 2.2.1, under $\theta(g)$ stratified πPS sampling strategy with the given allocation $\underset{\sim}{n}$ is inferior to unstratified πPS sampling strategy of fixed sample size n for estimating Y for all values of $g \in (1,2]$.

Corollary 2.2.2 implies that when the allocation is "$\theta(2)$-optimum" for all g (i.e. the allocation proportional to the stratum totals X_i's), then stratified πPS sampling strategy and unstratified πPS sampling strategy are equivalent for all $g \in [1,2]$. Provided the required conditions of the theorem are satisfied which reduce to the ordering (non-decreasing of the $\mathcal{H}$-values and the ordering (non-decreasing and not all equal) of the stratum means, Corollary 2.2.3 implies that stratified πPS sampling with allocation proportional to the stratum sizes (i.e., N_i's) is worse than unstratified πPS sampling for $g \in [1,2]$. However, with equal allocation to the strata (or allocation proportional to $N_i X_i$), in practice, we do come across stratified populations with ordered (non-decreasing) $\mathcal{H}$-values for which as N_i decreases, X_i increases so that Cov.(N_i, X_i) is negative and the conditions of the theorem are automatically satisfied, thereby implying that stratified πPS sampling is superior to unstratified πPS sampling for values o g near 1. Moreover, as mentioned for the case of proportional allocation when we have the ordering (non-decreasing) of the $\mathcal{H}$-values and the ordering (non-decreasing and not all equal) of the stratum means, stratified πPS sampling with $\theta(1)$-optimum allocation is superior to unstratified πPS sampling for values of g close to 1.

2.4 Illustrations

We illustrate the results obtained in the above sections by considering the following populations :

Population 1 : We consider real data on crops and grass acreage given by Sampford (1962, page 61) which relates to 35 farms in Orkney. The relevant data is given in the appendix AI. The population was divided into three strata as follows (Sampford, page 72).

Stratum i	Size of the stratum N_i	The total of crops and grass acreage X_i
I: Farms 1-12	12	735
II: Farms 13-24	12	1537
III: Farms 25-35	11	3487

An overall sample of size $n = 9$ is taken for illustration and various feasible allocations (with the restriction that at least two units be selected from each stratum for the estimability of the variance of the estimator) are considered.

We present Table 2.4.1 showing the efficiency of unstratified πPS sampling as compared to stratified πPS sampling for these allocations

(i.e., $E_{\theta(g)}$ Var.($\hat{Y}_S$)/$E_{\theta(g)}$ Var.($\hat{Y}_U$)).

Table 2.4.1

The efficiency of unstratified πPS sampling compared to stratified πPS sampling for all feasible allocations for a total sample size $n = 9$ for $g = 1.0(0.1)1.9$.

	Allocation							
g	(2,3,4)	(2,4,3)	(3,2,4)	(3,4,2)	(3,3,3)	(4,2,3)	(4,3,2)	(2,2,5)
1.0	0.8648	0.9646	0.9612	1.2604	0.9686	1.1092	1.3086	0.9343
1.1	0.8921	1.0118	0.9955	1.3543	1.0219	1.1668	1.4061	0.9510
1.2	0.9208	1.0620	1.0306	1.4540	1.0779	1.2266	1.5089	0.9675
1.3	0.9508	1.1151	1.0662	1.5593	1.1365	1.2882	1.6167	0.9839
1.4	0.9820	1.1708	1.1023	1.6689	1.1972	1.3514	1.7291	1.0002
1.5	1.0138	1.2283	1.1384	1.7821	1.2594	1.4153	1.8444	1.0160
1.6	1.0469	1.2889	1.1748	1.9010	1.3242	1.4809	1.9649	1.0315
1.7	1.0800	1.3500	1.2106	2.0206	1.3889	1.5459	2.0858	1.0464
1.8	1.1141	1.4135	1.2463	2.1445	1.4555	1.6119	2.2107	1.0607
1.9	1.1477	1.4765	1.2809	2.2674	1.5213	1.6764	2.3341	1.0742

In Table 2.4.1 corresponding to the allocations (3,4,2), (4,2,3) and (4,3,2) for which $f(1) = (\sum_{i=1}^{3} N_i X_i n_i^{-1}) - N X n^{-1}$ is positive, stratified πPS sampling is not recommended for all $g \in [1,2]$. If $f(1)$ is negative, which corresponds to the allocations (2,4,3), (3,2,4) and (3,3,3), when the value of g

is not far away from unity, then stratification might be used. Also note that for the allocation (2,3,4) the value of g_0 is between 1.4 and 1.5 and for values of $g \geq 1.6$ the efficiency is nearly 1, which shows that stratification is, as can be expected, better than unstratified sampling since in this case the $\theta(g)$-optimum allocation is <u>very close</u> to (2,3,4) for $g \varepsilon [1,2)$ (cf. Theorem 2.1.1).

For $g = 2$ the optimum allocation (by chance effect of rounding off) reduces to (2,2,5) which does not satisfy the conditions on a_i. In view of the fact that the conditions on a_i are sufficient, but not necessary for the theorem to hold, the efficiency corresponding to this allocation (2,2,5) as well is given in the last column of Table 2.4.1. It is interesting to observe that, for this allocation, the efficiencies for $g > 1.3$ are very close to 1.

<u>Population 2</u> : Population data from the census of India 1971 document was considered (data provided in the appendix AII). It consists of 1961 census population ($\mathcal{X}$) for 142 cities/urban agglomerations of India with population size 100,000 and above according to 1971 population census. The population was arranged in increasing order of the 1961 census figures and was further divided for illustration into 3 strata as follows :

Stratum	Size N_i	Stratum total X_i
1	55	46902
2	62	107797
3	25	257096

A total sample size $n = 15$ was considered (approximately 10 per cent of the population size 142). The admissible allocations that were considered (the allocations that satisfy the sufficient conditions on a_i's in our theorem) are : (4,5,6), (4,6,5), (5,4,6), (5,6,4), (6,4,5), (6,5,4), (3,5,7), (5,3,7), (5,7,3) and (7,3,5). Table 2.4.2 gives the efficiency of unstratified πPS sampling as compared to stratified πPS sampling with these allocations when the overall sample size is $n = 15$ for $g = 1.0(0.1)2.0$.

Table 2.4.2

g	Allocation				
	(4,5,6)	(4,6,5)	(5,4,6)	(5,6,4)	(6,4,5)
1.0	0.7575	0.7550	0.8163	0.8102	0.8531
1.1	0.7871	0.7942	0.8496	0.8673	0.8983
1.2	0.8208	0.8397	0.8867	0.9341	0.9496
1.3	0.8594	0.8933	0.9287	1.0133	1.0088
1.4	0.9046	0.9571	0.9772	1.1086	1.0782
1.5	0.9582	1.0348	1.0342	1.2257	1.1615
1.6	1.0237	1.1315	1.1034	1.3727	1.2642
1.7	1.1063	1.2556	1.1902	1.5633	1.3953
1.8	1.2150	1.4215	1.3042	1.8203	1.5701
1.9	1.3664	1.6556	1.4630	2.1861	1.8167
2.0	1.5951	2.0131	1.7035	2.7485	2.1938

(Table Contd.)

Table 2.4.2 (Contd.)

g	Allocation				
	(6,5,4)	(3,5,7)	(5,3,7)	(5,7,3)	(7,3,5)
1.0	0.8495	0.7753	0.9322	0.9182	0.9953
1.1	0.9089	0.7942	0.9608	1.0012	1.0443
1.2	0.9780	0.8156	0.9914	1.0996	1.0992
1.3	1.0595	0.8399	1.0247	1.2179	1.1619
1.4	1.1570	0.8680	1.0617	1.3621	1.2348
1.5	1.2763	0.9011	1.1038	1.5413	1.3219
1.6	1.4258	0.9410	1.1533	1.7690	1.4290
1.7	1.6192	0.9906	1.2141	2.0670	1.5658
1.8	1.8798	1.0548	1.2927	2.4725	1.7485
1.9	2.2506	1.1432	1.4010	3.0537	2.0073
2.0	2.8208	1.2751	1.5642	3.9528	2.4047

From Table 2.4.2 we have that stratified πPS sampling with allocations (4,5,6), (4,6,5) and (5,4,6) is recommended for values of g upto 1.4. Corresponding to the allocations (5,7,3) and (7,3,5) since g_o lies between 1.0 and 1.1, unless $g = 1$ stratified πPS sampling is not recommended. The allocation (3,5,7) is such that stratified πPS sampling with this allocation of the total sample size is superior to unstratified πPS sampling for values of g upto 1.7 for estimating Y under $\theta(g)$ in the expected variance sense.

<u>Remark 2.4.3</u> : Instead of πPS design with fixed sample size n_i withing each stratum, one can think of using a GπPS(Generalized πPS) sampling design for $g \in [1,2]$ (cf. T.J.Rao (1972))for drawing a sample in the ith stratum for $i = 1,2, \ldots, k$. A GπPS sampling design for drawing a sample in the ith stratum is defined to be the design wherein $\pi_{(i)j}$ is $\propto x_{ij}^{g/2}$ for $j = 1,2,\ldots,N_i$ and $\sum_j' x_{ij}^{1-(g/2)} = c$, a constant for each sample s with $p_s > 0$ from ith stratum, and $g \in [1,2]$. Observe that the constant c is in fact $= n_i X_i/(\sum_j x_{ij}^{g/2})$ where n_i is the expected sample size of the design. Note that when $g = 2$, a GπPS sampling design for drawing a sample in the ith stratum is a πPS sampling design with fixed sample size n_i. Also note that for a GπPS design to be available in the ith stratum, $n_i \leq (\sum_j x_{ij}^{g/2})/x_{ij}^{g/2}$ for every j. The condition $\sum_j' x_{ij}^{1-(g/2)} = c$, a constant is mostly satisfied because of the homogeneity of the $\mathcal{X}$-values within each stratum. Now, suppose the estimator $\hat{Y}_S = \sum_{i=1}^{k} \sum_j' \frac{y_{ij}}{\pi_{(i)j}}$ of the population total $Y = \sum_i \sum_j y_{ij}$ is used with the stratified GπPS design for $g \in [1,2]$ described above (i.e., a GπPS design in the ith stratum for $i = 1,2,\ldots,k$), we have under the model $\theta(g)$ of (2.1.3),

$$E_{\theta(g)}\,[\text{Var.}(\text{stratified } G\pi PS\ .\ \hat{Y}_S)]$$

$$= \sigma^2 \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(\frac{\sum_j x_{ij}^{g/2}}{n_i\, x_{ij}^{g/2}} - 1 \right) x_{ij}^{g}$$

$$= \sigma^2 \left[\sum_{i=1}^{k} \frac{(\sum_j x_{ij}^{g/2})^2}{n_i} - \sum_{i=1}^{k} \sum_{j=1}^{N_i} x_{ij}^{g} \right]. \qquad (2.4.1)$$

On the other hand, let a πPS design with fixed sample size $n\ (= \sum_{i=1}^{k} n_i)$ be used to draw a sample from the whole unstratified population. Considering $\hat{Y}_U$ as an estimator of the population total Y with this πPS design of fixed sample size n we have under $\theta(g)$,

$$E_{\theta(g)}\ \text{Var.}(\pi PS\ \cdot\ \hat{Y}_U) = \sigma^2 \cdot \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(\frac{X}{n\, x_{ij}} - 1 \right) x_{ij}^{g} \,. \qquad (2.4.2)$$

Considering (2.4.1) and (2.4.2) we have

$$E_{\theta(g)} \left\{ \text{Var.}(\text{stratified } G\pi PS : \hat{Y}_S) - \text{Var.}(\pi PS\ .\ \hat{Y}_U) \right\} / \sigma^2$$

$$= \sum_{i=1}^{k} \frac{(\sum_j x_{ij}^{g/2})^2}{n_i} - \frac{X}{n} \sum_{i=1}^{k} \sum_{j=1}^{N_i} x_{ij}^{g-1}$$

$$\leq f(g) \quad \text{for all } g \qquad (2.4.3)$$

(by Cauchy-Schwarz inequality).

Hence, whenever $f(1) < 0$ in which case there is a g_o (by Theorem 2.2.1) below which stratified πPS sampling with the given allocation $\underset{\sim}{n}$ of the sample size to strata is superior to unstratified πPS sampling with sample size $n(=\sum_{i=1}^{k} n_i)$, it automatically follows from (2.4.3) that with a stratified GπPS sampling with the allocation $\underset{\sim}{n}$ of the total expected sample size to strata, one is better off for values of g at least upto this g_o. On the other hand, if $f(1) \geq 0$, while stratified πPS sampling is not recommended, one might expect that with a stratified GπPS sampling, one might still do better for values of g close to unity.

We discuss GπPS sampling designs in detail in Chapters 5, 6 and 7.

CHAPTER 3

STRATIFIED PPS SAMPLING STRATEGIES

3.0 Summary

We first derive in this chapter the optimum allocation of sample size to strata which minimizes the expected variance of the stratified Probability Proportional to Size with Replacement (PPS) sampling strategy. With this optimum allocation we show that stratified PPS sampling strategy is better than the unstratified PPS sampling strategy in the expected variance sense for estimating Y. Next, generalizing Des Raj's result we also study in the expected variance sense, whether stratified PPS sampling with various non-optimal allocations apart from the X-proportional allocation (i.e., allocation proportional to the stratum total of $\mathcal{X}$, the auxiliary character closely related to $\mathcal{Y}$, the study variable) is likely to be worthwhile. In the last section of the chapter, to achieve considerable gains when stratification is used, a mode of stratification is suggested which is empirically investigated. Numerical examples based on live data are provided to illustrate the results at various stages.

Throughout this chapter a sample s denotes a finite ordered sequence of (not necessarily distinct) units of U. Recall that the length of the finite sequence is the 'sample size'. Now, given a design $d = (S, p)$ where $S = \{s\}$, the collection of all finite ordered sequences of units of U, say that p is a fixed sample size (=n) design if $p_s > 0 \Rightarrow n(s)=n$.

3.1 Optimum allocation of sample size to strata

Consider a finite population of size N divided into k strata of sizes N_i, $i = 1,2,\ldots,k$ (i.e., $N = \sum_{i=1}^{k} N_i$). Let y_{ij} and x_{ij} be the values of the characteristic $\mathcal{Y}$ under study and the characteristic $\mathcal{X}$ (essentially positive valued auxiliary information closely related to the characteristic $\mathcal{Y}$) respectively, for the jth unit of the ith stratum. Let a Probability Proportional to Size with Replacement (to be denoted throughout this chapter as PPS) sampling design with fixed sample size n_i be used in the ith stratum for $i = 1,2,\ldots,k$, for selecting the sample. Let sampling be carried out independently in each stratum. The sample in the ith stratum is therefore selected as follows : each unit in the sample is select- using initial probability of selection of the jth unit in the ith stratum given by x_{ij}/X_i, where $X_i = \sum_{j=1}^{N_i} x_{ij}$ after replacing the previous unit drawn. The sample sizes n_i are such that $\sum_{i=1}^{k} n_i = n$. As in chapter 2, we use the notation $\sum_i$ to denote $\sum_{i=1}^{k}$, summation over all strata and for a given i, $\sum_j$ stands for $\sum_{j=1}^{N_i}$, summation over all population units in the ith stratum for $i = 1,2,\ldots,k$. As an estimator of the population total $Y = \sum_i \sum_j y_{ij}$ based on this stratified PPS design, consider the conventional estimator

$$\hat{Y}_{S,PPS} = \sum_{i=1}^{k} \frac{X_i}{n_i} \sum_j{}' \frac{y_{ij}}{x_{ij}}, \tag{3.1.1}$$

where for a given i, $\sum_j{}'$ stands for summation over all the sampled n_i units from the ith stratum.

Following Cochran (1946), we regard $\underset{\sim}{Y} = (y_{11}, y_{12}, \ldots, y_{kN_k})$ as a realization of an N-length random vector with distribution ξ depending on $\underset{\sim}{X} = (x_{11}, x_{12}, \ldots, x_{kN_k})$ and some unknown parameters. We use the notation y_{ij} both as a random variable and the value it takes on the jth unit of the ith stratum given a finite population of size N $(i = 1,2,\ldots,k\,;\ j = 1,2,\ldots,N_i)$. Given $\underset{\sim}{X}$, consider the model ξ such that

$$\left.\begin{aligned} &E_\xi\,(y_{ij}|x_{ij}) = a\,x_{ij} \\ &Var._\xi(y_{ij}|x_{ij}) = \sigma_{ij}^2 \\ &Cov._\xi(y_{ij},\, y_{rh}|x_{ij},\, x_{rh}) = 0,\ 1 \le i,\ r \le k \\ &\qquad\qquad\qquad\qquad j \ne h \ \text{ if } \ i=r \end{aligned}\right\} \tag{3.1.2}$$

en $\sigma_{ij}^2 = \sigma^2 x_{ij}^g$ for $g \in [1,2]$, the model ξ of (3.1.2) is e same as the model $\Theta(g)$ of (2.1.3). We first have the llowing

heorem 3.1.1 : Under the super-population model ξ, the llocation of sample size to the strata that minimizes the pected variance of the strategy (stratified PPS , $\hat{Y}_{S,PPS}$) is ven by

$$n_i = n \sqrt{\sum_j \sigma_{ij}^2 \left(X_i x_{ij}^{-1} - 1\right)} \Bigg/ \sum_{i=1}^{k} \sqrt{\sum_j \sigma_{ij}^2 \left(X_i x_{ij}^{-1} - 1\right)}$$

roof : We have

$$\text{Var.(stratified PPS: } \hat{Y}_{S,PPS}) = E_\xi \left(\sum_{i=1}^{k} \frac{X_i}{n_i} \sum_{j=1}^{N_i} \frac{y_{ij}^2}{x_{ij}} - \sum_{i=1}^{k} \frac{Y_i^2}{n_i} \right)$$

where $Y_i = \sum_j y_{ij}$

since sampling is independent in each stratum

$$= \sum_{i=1}^{k} \frac{X_i}{n_i} \sum_{j=1}^{N_i} \frac{a^2 x_{ij}^2 + \sigma_{ij}^2}{x_{ij}} - \sum_{i=1}^{k} \frac{a^2 X_i^2 + \sum_{j=1}^{N_i} \sigma_{ij}^2}{n_i}$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(\frac{X_i}{n_i x_{ij}} - \frac{1}{n_i} \right) \sigma_{ij}^2$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{N_i} \frac{\sigma_{ij}^2}{n_i} \left(X_i x_{ij}^{-1} - 1\right) . \qquad (3.1.3)$$

inimizing (3.1.3) subject to the condition $\sum_{i=1}^{k} n_i = n$ by troducing Lagrangian multiplier λ, we next have

$$-\sum_{j=1}^{N_i} (\sigma_{ij}^2 \ (X_i \ x_{ij}^{-1} - 1)/n_i^2) + \lambda = 0 \quad \text{for} \quad i = 1,2, \ldots, k$$

ch gives

$$n_i = n \sqrt{\sum_j \sigma_{ij}^2 (X_i \ x_{ij}^{-1} - 1)} \quad \sum_{i=1}^{k} \sqrt{\sum_j \sigma_{ij}^2 \ (X_i \ x_{ij}^{-1} - 1)}$$

r $i = 1,2,\ldots,k$ (3.1.4)

the allocation of total sample size n to strata that inimizes under ξ, the expected variance of the strategy stratified PPS : $\hat{Y}_{S,PPS}$).

efinition 3.1.2 : The allocation of sample size to strata ven in (3.1.4) is called " ξ-optimum allocation".

heorem 3.1.3 : In the sense of expected variance, under ξ, stratified PPS sampling strategy is worse than stratified S sampling strategy with ξ-optimum allocation for estimating

oof : Let $\hat{Y}_{U,PPS} = \frac{X}{n} \sum_{i=1}^{k} \sum_j{}' \frac{y_{ij}}{x_{ij}}$ be the estimator of the pulation total $Y = \sum_{i=1}^{k} Y_i$ based on a Probability Proportional Size with Replacement design (PPS) to select a sample of ize $n(=\sum_{i=1}^{k} n_i)$ from the whole (unstratified) population. Thus, in this design each unit in the sample of size n is selected with the initial probabilities of selection now being given by x_{ij}/X, $1 \le i \le k$, $j = 1,2,\ldots,N_i$ where $X = \sum_{i=1}^{k} X_i$, after replacing the unit selected at the previous draw. $\sum_j'$ for a given i in $\hat{Y}_U$ stands for summation over those units out of the sampled n units that belong to the ith stratum.

Now,

$$E_\xi \text{ Var.}(PPS: \hat{Y}_{U,PPS}) = E_\xi \left\{ \frac{1}{n} \left(\sum_{i=1}^{k} \sum_{j=1}^{N_i} \frac{y_{ij}^2}{x_{ij}} \cdot X - Y^2 \right) \right\}$$

$$= \frac{1}{n} \left\{ \left(\sum_{i=1}^{k} \sum_{j=1}^{N_i} \frac{\sigma_{ij}^2 + a^2 x_{ij}^2}{x_{ij}} \cdot X \right) - \left(a^2 X^2 + \sum_i \sum_j \sigma_{ij}^2 \right) \right\}$$

$$= \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{N_i} \sigma_{ij}^2 \left(X x_{ij}^{-1} - 1 \right) \qquad (3.1.5)$$

When the allocation is $n_i = n X_i/X$ for $i = 1,2,\ldots,k$ i.e., X-proportional allocation we have from (3.1.3) that

$$E_\xi \text{ Var.}(\text{stratified PPS} \cdot \hat{Y}_{S,PPS}^{X\text{-prop.}}) = \sum_{i=1}^{k} \sum_{j=1}^{N_i} \frac{\sigma_{ij}^2}{n} \left(X x_{ij}^{-1} - X X_i^{-1} \right). \qquad (3.1.6)$$

Furthermore, with the ξ-optimum allocation given in (3.1.4)

$$E_\xi \text{ Var.(stratified PPS : } \hat{Y}^{opt.}_{S,PPS}) = (\sum_{i=1}^{k} \sqrt{b_i}\,)^2/n \qquad (3.1.7)$$

where $b_i = \sum_{j=1}^{N_i} \sigma^2_{ij} \, (X_i \, x^{-1}_{ij} - 1)$.

Comparing (3.1.5) and (3.1.6)

$$E_\xi \text{ Var.(PPS : } \hat{Y}_{U,PPS}) - E_\xi \text{ Var.(stratified PPS : } \hat{Y}^{X\text{-prop.}}_{S,PPS})$$

$$= \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{N_i} \sigma^2_{ij} \, (X \, X_i^{-1} - 1) \qquad (3.1.8)$$

$$\geq 0$$

and from (3.1.6) and (3.1.7)

$$E_\xi \text{ Var.(stratified PPS : } \hat{Y}^{X\text{-prop.}}_{S,PPS}) - E_\xi \text{ Var.(stratified PPS : } \hat{Y}^{opt.}_{S,PPS}$$

$$= \frac{1}{n}\{(\sum_{i=1}^{k} (b_i \, X \, X_i^{-1}) - (\sum_{i=1}^{k} \sqrt{b_i}\,)^2\}$$

$$= \frac{1}{n} \{ (\sum_{i=1}^{k} b_i \, X_i^{-1}) \, (\sum_{i=1}^{k} X_i) - (\sum_{i=1}^{k} \sqrt{b_i}\,)^2 \} \quad (3.1.9)$$

$$\geq 0,$$

by Cauchy-Schwarz inequality.

From (3.1.8) and (3.1.9) theorem follows.

3.2 Comparison between Stratified and Unstratified PPS Sampling Strategies

Des Raj (1963) has compared unstratified PPS sampling strategy with stratified PPS sampling strategy when X-proportional allocation is used in terms of exact variance. He observes that

$$\text{Var.}(\text{stratified PPS} : \hat{Y}_{S,PPS}) = \sum_{i=1}^{k} \frac{X_i}{n_i} \sum_{j=1}^{N_i} x_{ij}\left(\frac{y_{ij}}{x_{ij}} - R_i\right)^2, \quad (3.2.1)$$

where $R_i = Y_i/X_i$. When $n_i = n\, X_i/X$ for $i = 1,2,\ldots,k$,

$$\text{Var.}(\text{stratified PPS} : \hat{Y}^{X\text{-prop.}}_{S,PPS}) = \frac{X}{n} \sum_{i=1}^{k} \sum_{j=1}^{N_i} x_{ij}\left(\frac{y_{ij}}{x_{ij}} - R_i\right)^2 \quad (3.2.2)$$

Comparing (3.2.2) with the situation in which no stratification is made, Des Raj observes that

$$\text{Var.}(\text{PPS}:\hat{Y}_{U,PPS}) - \text{Var.}(\text{stratified PPS}:\hat{Y}^{X\text{-prop.}}_{S,PPS}) = \frac{X}{n} \sum_{i=1}^{k} X_i (R_i - R)^2 \quad (3.2.3)$$

$$\geq 0$$

where $R = Y/X$.

Noticing that this is a comparison between unstratified PPS sampling and stratified PPS sampling with a non-optimal allocation, we now study in the expected variance sense, whether stratified PPS sampling with various non-optimal allocations is likely to be worthwhile (cf. chapter 2, Section 2.1).

Considering the model (3.1.2) we have

$$E_\xi \text{ Var.(stratified PPS : } \hat{Y}_{S,PPS}) - E_\xi \text{ Var.(PPS : } \hat{Y}_{U,PPS})$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{N_i} \frac{\sigma_{ij}^2}{x_{ij}} \left(\frac{X_i}{n_i} - \frac{X}{n}\right) + \sum_{i=1}^{k} \sum_{j=1}^{N_i} \sigma_{ij}^2 \left(\frac{1}{n} - \frac{1}{n_i}\right)$$

$$\leq \sum_{i=1}^{k} \sum_{j=1}^{N_i} \frac{\sigma_{ij}^2}{x_{ij}} \left(\frac{X_i}{n_i} - \frac{X}{n}\right). \tag{3.2.4}$$

When $\sigma_{ij}^2 = \sigma^2 x_{ij}^g$, $g \in [1,2]$ the model ξ of (3.1.2) is the model $\theta(g)$ of (2.1.3) and now (3.2.4) reduces to

$$E_{\theta(g)} \left\{ \text{Var.(stratified PPS : } \hat{Y}_{S,PPS}) - \text{Var.(PPS : } \hat{Y}_{U,PPS}) \right\} / \sigma^2$$

$$\leq \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(\frac{X_i}{n_i} - \frac{X}{n}\right) x_{ij}^{g-1} \tag{3.2.5}$$

Observe that the right hand side of (3.2.5) is the $f(g)$ of (2.1.6).

<u>Remark 3.2.1</u> : Consider the " $\theta(g)$-optimum allocation" given by $n_i = n(X_i \sum_{j=1}^{N_i} x_{ij}^{g-1})^{\frac{1}{2}} / \sum_{i=1}^{k} (X_i \sum_{j=1}^{N_i} x_{ij}^{g-1})^{\frac{1}{2}}$ (cf. (2.1.7)). With the $\theta(g)$-optimum allocation since under the model (2.1.3) $f(g) < 0$ for $g \in [1,2)$ and equal to zero for $g = 2$, in which

case the $\theta(2)$-optimum allocation is $n_i = n\ X_i/X$, for $i = 1,2,\ldots,k$, due to (3.2.5) we have that unstratified PPS sampling is worse than stratified PPS sampling for estimating Y under $\theta(g)$ for all $g \in [1,2]$ with the $\theta(g)$-optimum allocation.

Remark 3.2.2 : Observe that though we derived optimum allocation of sample size to strata for the stratified PPS sampling strategy working with the model ξ , our comparison studies of stratified and unstratified PPS sampling strategies for estimating Y are when the model (2.1.3) holds which is of more practical interest.

The ξ-optimum allocation of (3.1.4) when $\sigma_{ij}^2 = \sigma^2 x_{ij}^g$ for $g \in [1,2]$ in the model (3.1.2) reduces to "ξ_g-optimum allocation" given by

$$n_i = n \frac{\sqrt{(X_i \sum_j x_{ij}^{g-1}) - \sum_j x_{ij}^g}}{\sum_{i=1}^{k} \sqrt{(X_i \sum_j x_{ij}^{g-1}) - \sum_j x_{ij}^g}} \qquad (3.2.6)$$

Thus, both the "$\theta(g)$-optimum allocation" of (2.1.7) and "ξ_g-optimum allocation" of (3.2.6) are such that unstratified PPS sampling strategy is worse than stratified PPS sampling strategy with these allocations for estimating Y, under the model $\theta(g)$ of (2.1.3) for all $g \in [1,2]$ in the expected variance sense. The proof of Theorem 3.1.3 shows that unstratified PPS sampling is worse than stratified PPS sampling with X-proportional allocation under the model $\theta(g)$ of (2.1.3) for all $g \in [1,2]$. Hence in all these situations, stratified PPS sampling is better than unstratified PPS sampling in the expected variance sense, thus generalizing Des Raj's (1963) result. Using the inequality (3.2.5) we now compare unstratified PPS sampling and stratified PPS sampling with various non-optimal allocations for estimating Y under $\theta(g)$ in the expected variance sense for $g \in [1,2]$.

Let us consider (3.2.5). Let the allocation $\underset{\sim}{n} = (n_1, n_2, \ldots, n_k)$ of the total sample size to strata be such that the sufficient conditions of Theorem 2.2.1 are satisfied. By part (a) of Theorem 2.2.1, whenever $f(1) < 0$, there exists a value g_o of g in $(1,2]$ such that $f(g) \lesseqgtr 0$ according as $g \lesseqgtr g_o$ and so with all these allocations that satisfy the sufficient conditions of Theorem 2.2.1, when $f(1) < 0$ stratified PPS sampling is superior to unstratified PPS sampling for values of g at least upto this g_o. When $f(1) \geq 0$, by part (b) of Theorem 2.2.1, even though $f(g) > 0$ for all g in $(1,2]$, stratified PPS sampling might still be superior to unstratified PPS sampling with these allocations for values of g close to unity.

3.3 Illustrations

Population 1 : We illustrate the above results using live data on crops and grass acreage given by Sampford (1962, page 61). The relevant data is given in the appendix AI. As in section 2.4 of chapter 2, there are 3 strata and the stratum sizes are $N_1 = 12$, $N_2 = 12$ and $N_3 = 11$. An overall sample size $n = 9$ is taken for illustration and various feasible allocations with at least 2 units from each stratum that satisfy the sufficient conditions of Theorem 2.2.1 are considered. Table 3.3.1 shows the efficiency of unstratified PPS sampling as compared to stratified PPS sampling for these allocations (i.e., $E_{\theta(g)}$ Var.(stratified PPS : $\hat{Y}_{S,PPS}$)/ $E_{\theta(g)}$Var.(PPS : $\hat{Y}_{U,PPS}$)) for $g = 1.0(0.1)\ 1.9$. In Table 3.3.1, entries in brackets denote the efficiency of unstratified πPS sampling strategy as compared to stratified πPS sampling strategy for comparison with Table 2.4.1.

Table 3.3.1

The efficiency of unstratified PPS sampling as compared to stratified PPS sampling for all feasible allocations for a total sample size n = 9 for g = 1.0(0.1) 1.9.

	Allocation						
g	(2,3,4)	(2,4,3)	(3,2,4)	(3,4,2)	(3,3,3)	(4,2,3)	(4,3,2)
1.0	0.8455 (0.8648)	0.9143 (0.9646)	0.9131 (0.9612)	1.1195 (1.2604)	0.9172 (0.9686)	1.0157 (1.1092)	1.1533 (1.3086)
1.1	0.8664 (0.8921)	0.9478 (1.0118)	0.9376 (0.9955)	1.1818 (1.3543)	0.9547 (1.0219)	1.0546 (1.1668)	1.2174 (1.4061)
1.2	0.8877 (0.9208)	0.9823 (1.0620)	0.9620 (1.0306)	1.2458 (1.4540)	0.9930 (1.0779)	1.0938 (1.2266)	1.2829 (1.5089)
1.3	0.9092 (0.9508)	1.0172 (1.1151)	0.9861 (1.0662)	1.3109 (1.5593)	1.0317 (1.1365)	1.1328 (1.2882)	1.3493 (1.6167)
1.4	0.9308 (0.9820)	1.0533 (1.1708)	1.0095 (1.1023)	1.3769 (1.6689)	1.0705 (1.1972)	1.1713 (1.3514)	1.4162 (1.7291)
1.5	0.9522 (1.0138)	1.0889 (1.2283)	1.0321 (1.1384)	1.4424 (1.7821)	1.1088 (1.2594)	1.2089 (1.4153)	1.4824 (1.8444)
1.6	0.9736 (1.0469)	1.1252 (1.2889)	1.0542 (1.1748)	1.5090 (1.9010)	1.1473 (1.3242)	1.2461 (1.4809)	1.5492 (1.9649)
1.7	0.9943 (1.0800)	1.1604 (1.3500)	1.0750 (1.2106)	1.5734 (2.0206)	1.1844 (1.3889)	1.2815 (1.5459)	1.6138 (2.0858)
1.8	1.0148 (1.1141)	1.1957 (1.4135)	1.0959 (1.2463)	1.6379 (2.1445)	1.2211 (1.4555)	1.3161 (1.6119)	1.6780 (2.2107)
1.9	1.0342 (1.1477)	1.2295 (1.4765)	1.1136 (1.2809)	1.6994 (2.2674)	1.2561 (1.5213)	1.3486 (1.6764)	1.7391 (2.3341)

From Table 3.3.1 we see that corresponding to the allocations (2,4,3), (3,2,4) and (3,3,3) stratified PPS sampling strategy is superior to unstratified PPS sampling strategy for values of g upto 1.2, 1.3 and 1.2 respectively. Observe that with these allocations stratified πPS strategy is superior to unstratified πPS strategy for values of $g = 1.0$, g upto 1.1 and $g = 1.0$ respectively. The " $\theta(g)$-optimum allocation" which is very close to (2,3,4) for all values of g is such that the efficiency of unstratified PPS strategy as compared to stratified PPS strategy is less than one for values of g upto 1.7 and when $g = 1.8$ and 1.9, the efficiencies are 1.0148 and 1.0342 respectively which are very close to one.

Population 2. 142 cities and urban agglomerations in India with population 100,000 and above according to the 1971 census figures was considered from the census of India 1971 document. The 1961 census population figures (number of persons) was considered to be $\mathcal{X}$. The relevant data is presented in the appendix AII. The population was arranged in increasing order of the 1961 census figures and was further divided for illustration into 3 strata with stratum sizes $N_1 = 55$, $N_2 = 62$ and $N_3 = 25$ with stratum totals $X_1 = 46902$, $X_2 = 107797$ and $X_3 = 257096$. A total sample size fixed as $n = 15$ was considered (approximately 10 per cent of the population size 142). The admissible allocations considered (the allocations that satisfy the sufficient conditions on a_i's in our Theorem 2.2.1) were : (4,5,6), (4,6,5), (5,4,6), (5,6,4), (6,4,5), (6,5,4), (3,5,7), (5,3,7), (5,7,3) and (7,3,5). The second columns in Table 3.3.2 to Table 3.3.11 give the efficiency of unstratified PPS sampling as compared to stratified PPS sampling with these allocations for $g = 1.0(0.1)\ 2.0$ (i.e., $E_{\theta(g)}$ Var.(stratified PPS : $\hat{Y}_{S,PPS}$)/ $E_{\theta(g)}$ Var.(PPS : $\hat{Y}_{U,PPS}$)). The first columns in Table 3.3.2 to Table 3.3.11 give the efficiency of unstratified πPS sampling strategy as compared to stratified πPS sampling strategy for these allocations for $g = 1.0(0.1)\ 2.0$ (i.e., $E_{\theta(g)}$ Var.(stratified πPS : $\hat{Y}_S$)/$E_{\theta(g)}$ Var.(πPS : $\hat{Y}_U$)) for comparison. The third columns in Table 3.3.2 to Table 3.3.11 present the values $[f(g)] \times 10^{-8}$ for $g = 1.0(0.1)2.0$ for these allocations, i.e., $[E_{\theta(g)} \{\text{Var. (stratified } \pi\text{PS} : \hat{Y}_S) - \text{Var.}(\pi\text{PS} : \hat{Y}_U)\}/\sigma^2] \times 10^{-8}$. The fourth columns in Table 3.3.2 to Table 3.3.11 present the values $[E_{\theta(g)}\{\text{Var.(stratified PPS} : \hat{Y}_{S,PPS}) - \text{Var.(PPS} : \hat{Y}_{U,PPS})\}/\sigma^2] \times 10^{-8}$, for $g = 1.0(0.1)2.0$, for these allocations. Observe that columns 3 and 4 of Table 3.3.11 enable us to study the inequality (3.2.5).

Table 3.3.2

	Allocation (7,3,5)			
g	Column 1	Column 2	Column 3	Column 4
1.0	.9953	.9785	-.0002	-.0008
1.1	1.0443	1.0196	.0032	.0016
1.2	1.0992	1.0631	.0147	.0109
1.3	1.1619	1.1093	.0498	.0401
1.4	1.2348	1.1579	.1497	.1251
1.5	1.3219	1.2089	.4249	.3611
1.6	1.4290	1.2618	1.1700	1.0010
1.7	1.5658	1.3159	3.1660	2.7130
1.8	1.7485	1.3702	8.5000	7.2650
1.9	2.0073	1.4237	22.7600	19.3500
2.0	2.4047	1.4750	61.0200	51.4900

Table 3.3.3

	Allocation (5,4,6)			
g	Column 1	Column 2	Column 3	Column 4
1.0	.8163	.8212	-.0064	-.0069
1.1	.8496	.8519	-.0108	-.0120
1.2	.8867	.8849	-.0168	-.0198
1.3	.9287	.9203	-.0219	-.0293
1.4	.9772	.9581	-.0145	-.0332
1.5	1.0342	.9983	.0452	-.0029
1.6	1.1034	1.0405	.2818	.1550
1.7	1.1902	1.0843	1.0640	.7239
1.8	1.3042	1.1288	3.4540	2.5270
1.9	1.4630	1.1731	10.4600	7.9050
2.0	1.7035	1.2160	30.5600	23.4200

Table 3.3.4

	Allocation (5,6,4)			
g	Column 1	Column 2	Column 3	Column 4
1.0	.8102	.8125	-.0066	-.0073
1.1	.8673	.8628	-.0095	-.0111
1.2	.9341	.9192	-.0098	-.0139
1.3	1.0133	.9821	.0041	-.0066
1.4	1.1086	1.0518	.0692	.0411
1.5	1.2257	1.1285	.2979	.2222
1.6	1.3727	1.2118	1.0160	.8100
1.7	1.5633	1.3008	3.1520	2.5840
1.8	1.8203	1.3941	9.3160	7.7340
1.9	2.1861	1.4897	26.8000	22.3600
2.0	2.7485	1.5848	75.9600	63.3900

Table 3.3.5

	Allocation (6,4,5)			
g	Column 1	Column 2	Column 3	Column 4
1.0	.8531	.8525	-.0051	-.0057
1.1	.8983	.8925	-.0073	-.0087
1.2	.9496	.9360	-.0075	-.0110
1.3	1.0088	.9831	.0027	-.0062
1.4	1.0782	1.0339	.0498	.0269
1.5	1.1615	1.0884	.2131	.1529
1.6	1.2642	1.1462	.7202	.5593
1.7	1.3953	1.2067	2.2120	1.7750
1.8	1.5701	1.2688	6.4740	5.2740
1.9	1.8167	1.3311	18.4600	15.1200
2.0	2.1938	1.3920	51.8600	42.4900

Table 3.3.6

g	Alocation (6,5,4)			
	Column 1	Column 2	Column 3	Column 4
1.0	.8495	.8473	-.0052	-.0059
1.1	.9089	.8991	-.0065	-.0082
1.2	.9780	.9566	-.0033	-.0075
1.3	1.0595	1.0202	.0183	.0074
1.4	1.1570	1.0902	.1001	.0714
1.5	1.2763	1.1666	.3647	.2880
1.6	1.4258	1.2490	1.1610	.9523
1.7	1.6192	1.3366	3.4650	2.8920
1.8	1.8798	1.4280	9.9910	8.3980
1.9	2.2506	1.5210	28.2600	23.7900
2.0	2.8208	1.6132	79.1000	66.4800

Table 3.3.7

g	Allocation (3,5,7)			
	Column 1	Column 2	Column 3	Column 4
1.0	.7753	.7856	-.0078	-.0083
1.1	.7942	.8048	-.0148	-.0159
1.2	.8156	.8259	-.0274	-.0299
1.3	.8399	.8492	-.0492	-.0554
1.4	.8680	.8746	-.0841	-.0993
1.5	.9011	.9021	-.1305	-.1692
1.6	.9410	.9316	-.1609	-.2614
1.7	.9906	.9628	-.0528	-.3196
1.8	1.0548	.9950	.6227	-.0982
1.9	1.1432	1.0275	3.2360	1.2570
2.0	1.2751	1.0595	11.9500	6.4490

Table 3.3.8

	Allocation (5,3,7)			
g	Column 1	Column 2	Column 3	Column 4
1.0	.9322	.9248	-.0024	-.0029
1.1	.9608	.9499	-.0028	-.0041
1.2	.9914	.9755	-.0013	-.0042
1.3	1.0247	1.0016	.0076	.0006
1.4	1.0617	1.0280	.0393	.0221
1.5	1.1038	1.0544	.1370	.0939
1.6	1.1533	1.0805	.4180	.3079
1.7	1.2141	1.1060	1.1980	.9104
1.8	1.2927	1.1303	3.3240	2.5570
1.9	1.4010	1.1530	9.0620	6.9840
2.0	1.5642	1.1733	24.5100	18.7900

Table 3.3.9

	Allocation (5,7,3)			
g	Column 1	Column 2	Column 3	Column 4
1.0	.9182	.9049	-.0028	-.0037
1.1	1.0012	.9749	.0001	-.0020
1.2	1.0996	1.0540	.0148	.0093
1.3	1.2179	1.1429	.0670	.0524
1.4	1.3621	1.2422	.2309	.1918
1.5	1.5413	1.3520	.7146	.6085
1.6	1.7690	1.4720	2.0960	1.8050
1.7	2.0670	1.6010	5.9700	5.1620
1.8	2.4725	1.7368	16.7200	14.4600
1.9	3.0537	1.8765	46.4100	40.0300
2.0	3.9528	2.0162	128.3000	110.2000

Table 3.3.10

g	Allocation (4,5,6)			
	Column 1	Column 2	Column 3	Column 4
1.0	.7575	.7690	-.0085	-.0089
1.1	.7871	.7975	-.0153	-.0165
1.2	.8208	.8288	-.0266	-.0294
1.3	.8594	.8631	-.0432	-.0503
1.4	.9046	.9006	-.0608	-.0787
1.5	.9582	.9412	-.0551	-.1016
1.6	1.0237	.9847	.0647	-.0585
1.7	1.1063	1.0306	.5950	.2626
1.8	1.2150	1.0781	2.4410	1.5320
1.9	1.3664	1.1261	8.2790	5.7570
2.0	1.5951	1.1733	25.8500	18.7900

Table 3.3.11

g	Allocation (4,6,5)			
	Column 1	Column 2	Column 3	Column 4
1.0	.7550	.7655	-.0085	-.0091
1.1	.7942	.8018	-.0148	-.0161
1.2	.8397	.8425	-.0238	-.0271
1.3	.8933	.8878	-.0328	-.0412
1.4	.9571	.9381	-.0273	-.0490
1.5	1.0348	.9933	.0459	-.0116
1.6	1.1315	1.0532	.3585	.2035
1.7	1.2556	1.1172	1.4300	1.0070
1.8	1.4215	1.1842	4.7860	3.6150
1.9	1.6556	1.2527	14.8100	11.5400
2.0	2.0131	1.3208	44.0100	34.7800

From the above Tables 3.3.2 to 3.3.11 we see that stratified PPS sampling with the allocation (5.4.6) is recommended for values of g upto 1.5. Since with the allocation (7,3,5) of sample size to strata stratified PPS sampling strategy is inferior to unstratified PPS sampling strategy for $g \geq 1.1$, unless $g = 1.0$, stratified PPS sampling with the allocation (7,3,5) is not recommended for estimating the population total Y. If $g \geq 1.4$, stratified PPS sampling with the allocation (5,6,4) is inferior to unstratified PPS sampling. Stratified PPS sampling with the allocations (5,3,7) and (6,5,4) of sample size to strata is recommended as compared to unstratified PPS sampling only for values of $g \leq 1.2$. When $g \geq 1.2$, stratified PPS sampling with the allocation (5,7,3) is inferior to unstratified PPS sampling. Therefore stratification with the allocation (5,7,3) is recommended only when $g = 1.0$ or 1.1. Observe from Table 3.3.10 that the efficiency of unstratified PPS sampling strategy as compared to stratified PPS sampling strategy with the allocation (4,5,6) of sample size to strata is less than one for $g \leq 1.6$, very nearly one for $g = 1.7$ and 1.8, equals 1.1261 when $g=1.9$ and equals 1.1733 when $g=2.0$. With the allocation (4,6,5), stratified PPS sampling is recommended for values of $g \leq 1.5$. With the allocation (3,5,7), stratified PPS sampling is superior to unstratified PPS sampling for values of $g \leq 1.8$ and observe that the efficiency of unstratified PPS sampling as compared to stratified PPS sampling with this allocation (3,5,7) is nearly one when $g = 1.9$ and 2.0.

3.4 Mode of Stratification

Considering the inequality (3.2.5) we have

$$E_{\theta(g)} \{ \text{Var.}(\text{stratified PPS} : \hat{Y}_{S,PPS}) - \text{Var.}(\text{PPS} \quad \hat{Y}_{U,PPS}) \} / \sigma^2$$
$$\leq \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(\frac{X_i}{n_i} - \frac{X}{n} \right) x_{ij}^{g-1}. \qquad (3.4.1)$$

Now,

$$\sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(\frac{X_i}{n_i} - \frac{X}{n}\right) x_{ij}^{g-1} = f(g) \text{ for } g \in [1,2] \qquad \text{(cf. (2.1.6) and Theorem 2.2.1)}.$$

We have that $f(2) \geq 0$ for any allocation $\underset{\sim}{n} = (n_1, n_2, \ldots, n_k)$ of sample size to strata. But for allocations $\underset{\sim}{n}$ that satisfy the sufficient conditions of our Theorem 2.2.1 we have that if $f(1) < 0$, there exists a unique g_o in the interval $(1,2]$ such that $f(g) \leq 0$ or > 0 according as $g \leq g_o$ or $g > g_o$, therefore, for a given allocation $\underset{\sim}{n}$ that satisfies the sufficient conditions of Theorem 2.2.1 if $f(1) < 0$, we construct new strata for the given allocation in such a way that $f(1)$ becomes large (and negative) and $f(2)$ becomes smaller (closer to zero since $f(2) \geq 0$ always), to study whether stratified PPS sampling strategy with the given allocation under the new mode of stratification is superior to unstratified PPS sampling strategy for more values of $g \in (1,2]$ compared to the original stratification for estimating Y.

To investigate if there is any gain achieved through the new mode of stratification, we study empirically Sampford's live data on 35 farms in Orkney (Sampford (1962), page 61). The data is provided in the appendix AI. The population was divided into three strata (Sampford, page 72) containing farms 1-12, farms 13-24 and farms 25-35. The overall sample size is $n = 9$ and as an illustration we consider the allocation (3,3,3) that satisfies the sufficient conditions of our Theorem 2.2.1. The original stratification as considered by Sampford((1962),p. 72) is given by

Stratum No.	Stratum size (N_i)	Stratum totals of the crops and grass acreage (X_i)
1	12	735
2	12	1537
3	11	3487

The second column of Table 3.4.1 below gives the efficiency of unstratified PPS sampling as compared to stratified PPS sampling with the allocation (3,3,3) when the original stratification is $N_1 = 12$, $N_2 = 12$ and $N_3 = 11$

(i.e., $E_{\theta(g)}$Var.(stratified PPS : $\hat{Y}_{S,PPS}$)/$E_{\theta(g)}$Var.(PPS : $\hat{Y}_{U,PPS}$)), for $g = 1.0(0.1)2.0$. The first column of Table 3.4.1 gives the efficiency of unstratified πPS sampling as compared to stratified πPS sampling with the allocation (3,3,3) when the original stratification is $N_1 = 12$, $N_2 = 12$ and $N_3 = 11$.

Table 3.4.1

g	$\frac{E_{\theta(g)}\text{Var.(stratified }\pi\text{PS : }\hat{Y}_S)}{E_{\theta(g)}\text{ Var.(}\pi\text{ PS : }\hat{Y}_U)}$	$\frac{E_{\theta(g)}\text{ Var.(stratified PPS : }\hat{Y}_{S,PPS})}{E_{\theta(g)}\text{ Var.(PPS : }\hat{Y}_{U,PPS})}$
1.0	0.9686	0.9172
1.1	1.0219	0.9547
1.2	1.0779	0.9930
1.3	1.1365	1.0317
1.4	1.1972	1.0705
1.5	1.2594	1.1088
1.6	1.3242	1.1473
1.7	1.3889	1.1844
1.8	1.4555	1.2211
1.9	1.5213	1.2561
2.0	1.5869	1.2897

With the same allocation (3,3,3) for the new mode of stratification with $N_1 = 18$, $N_2 = 10$ and $N_3 = 7$, Table 3.4.2 gives the efficiencies for $g = 1.0(0.1)2.0$. It is interesting to observe that with this new mode of stratification stratified PPS sampling is superior to unstratified PPS sampling for all values of g from 1.0 through 2.0 (while for the original stratification g_o is between 1.2 and 1.3).

Table 3.4.2

$N_1 = 18$, $N_2 = 10$, $N_3 = 7$.

g	$\frac{E_{\theta(g)} \text{ Var.(stratified } \pi PS : \hat{Y}_S)}{E_{\theta(g)} \text{ Var.(} \pi PS : \hat{Y}_U)}$	$\frac{E_{\theta(g)} \text{ Var.(stratified PPS} : \hat{Y}_{S,PPS})}{E_{\theta(g)} \text{ Var.(PPS} : \hat{Y}_{U,PPS})}$
1.0	0.8667	0.8392
1.1	0.8872	0.8533
1.2	0.9087	0.8676
1.3	0.9310	0.8820
1.4	0.9543	0.8964
1.5	0.9781	0.9107
1.6	1.0025	0.9248
1.7	1.0273	0.9385
1.8	1.0523	0.9519
1.9	1.0772	0.9648
2.0	1.1021	0.9771

The following Table 3.4.3 lists all the modes of stratification considered and the corresponding numbers of tables wherein the efficiencies are available for $g = 1.0(0.1)2.0$, with the allocation (3,3,3) of sample size to strata.

Table 3.4.3

Mode of stratification	Table number
$N_1 = 18,\ N_2 = 9,\ N_3 = 8$	Table 3.4.4
$N_1 = 20,\ N_2 = 8,\ N_3 = 7$	Table 3.4.5
$N_1 = 20,\ N_2 = 9,\ N_3 = 6$	Table 3.4.6
$N_1 = 19,\ N_2 = 9,\ N_3 = 7$	Table 3.4.7
$N_1 = 19,\ N_2 = 8,\ N_3 = 8$	Table 3.4.8

Table 3.4.4

$N_1 = 18,\ N_2 = 9,\ N_3 = 8.$

g	$\dfrac{E_{\theta(g)}\,\text{Var.}(\text{stratified } \pi PS : \hat{Y}_S)}{E_{\theta(g)}\,\text{Var.}(\pi PS : \hat{Y}_U)}$	$\dfrac{E_{\theta(g)}\,\text{Var.}(\text{stratified PPS} : \hat{Y}_{S,PPS})}{E_{\theta(g)}\,\text{Var.}(\text{PPS} : \hat{Y}_{U,PPS})}$
1.0	0.8699	0.8417
1.1	0.8943	0.8586
1.2	0.9202	0.8761
1.3	0.9476	0.8941
1.4	0.9764	0.9123
1.5	1.0065	0.9306
1.6	1.0376	0.9490
1.7	1.0696	0.9673
1.8	1.1023	0.9853
1.9	1.1355	1.0030
2.0	1.1690	1.0202

Table 3.4.5

$N_1 = 20$, $N_2 = 8$, $N_3 = 7$.

g	$\frac{E_{Q(g)}\text{ Var.(stratified }\pi\text{PS : }\hat{Y}_S)}{E_{Q(g)}\text{ Var.(}\pi\text{PS : }\hat{Y}_U)}$	$\frac{E_{Q(g)}\text{ Var.(stratified PPS : }Y_{S,PPS})}{E_{Q(g)}\text{ Var.(PPS : }\hat{Y}_{U,PPS})}$
1.0	0.9055	0.8689
1.1	0.9181	0.8766
1.2	0.9316	0.8846
1.3	0.9460	0.8929
1.4	0.9613	0.9014
1.5	0.9773	0.9101
1.6	0.9942	0.9190
1.7	1.0116	0.9279
1.8	1.0296	0.9368
1.9	1.0480	0.9456
2.0	1.0668	0.9543

Table 3.4.6

$N_1 = 20,\ N_2 = 9,\ N_3 = 6.$

g	$\dfrac{E_{\theta(g)}\ \text{Var.}(\text{stratified } \pi PS : \hat{Y}_S)}{E_{\theta(g)}\ \text{Var.}(\pi PS : \hat{Y}_U)}$	$\dfrac{E_{\theta(g)}\ \text{Var.}(\text{stratified PPS} : \hat{Y}_{S,PPS})}{E_{\theta(g)}\ \text{Var.}(PPS : \hat{Y}_{U,PPS})}$
1.0	0.9089	0.8715
1.1	0.9192	0.8774
1.2	0.9300	0.8834
1.3	0.9414	0.8895
1.4	0.9531	0.8956
1.5	0.9652	0.9016
1.6	0.9775	0.9075
1.7	0.9900	0.9132
1.8	1.0027	0.9188
1.9	1.0153	0.9241
2.0	1.0279	0.9292

Table 3.4.7

$N_1 = 19,\ N_2 = 9,\ N_3 = 7$

g	$\frac{E_{\theta(g)}\text{ Var.(stratified }\pi PS : \hat{Y}_S)}{E_{\theta(g)}\text{ Var.(}\pi PS : \hat{Y}_U)}$	$\frac{E_{\theta(g)}\text{ Var.(stratified PPS : }\hat{Y}_{S,PPS})}{E_{\theta(g)}\text{ Var.(PPS : }\hat{Y}_{U,PPS})}$
1.0	0.8805	0.8498
1.1	0.8969	0.8606
1.2	0.9143	0.8718
1.3	0.9326	0.8831
1.4	0.9518	0.8946
1.5	0.9717	0.9062
1.6	0.9923	0.9177
1.7	1.0135	0.9291
1.8	1.0350	0.9404
1.9	1.0568	0.9513
2.0	1.0787	0.9620

Table 3.4.8

$N_1 = 19,\ N_2 = 8,\ N_3 = 8.$

g	$\dfrac{E_{\theta(g)}\text{Var.(stratified } \pi PS : \hat{Y}_S)}{E_{\theta(g)}\text{ Var.(} \pi PS : \hat{Y}_U)}$	$\dfrac{E_{\theta(g)}\text{Var.(stratified PPS : } \hat{Y}_{S,PPS})}{E_{\theta(g)}\text{ Var.(PPS : } \hat{Y}_{U,PPS})}$
1.0	0.8925	0.8590
1.1	0.9132	0.8730
1.2	0.9355	0.8875
1.3	0.9592	0.9025
1.4	0.9844	0.9180
1.5	1.0109	0.9338
1.6	1.0386	0.9497
1.7	1.0673	0.9657
1.8	1.0969	0.9817
1.9	1.1271	0.9975
2.0	1.1580	1.0131

Observe from Table 3.4.5, Table 3.4.6 and Table 3.4.7 that the gain due to the new mode of stratification for PPS sampling is considerable since stratified PPS sampling with these new modes of stratification is superior to unstratified PPS sampling for <u>all</u> values of g from 1.0 through 2.0. Note that our aim here is to study only the gain achieved through the new mode of stratification for a given allocation $\underset{\sim}{n}$ of sample size to strata that satisfies the sufficient condition of our Theorem 2.2.1, for which $f(1) < 0$. We are not suggesting any optimum mode of stratification (for a given allocation $\underset{\sim}{n}$) that maximizes the gain achieved.

As for stratified πPS sampling while with the original stratification the value of g_0 is between 1.0 and 1.1 (cf. Table 3.4.1) with the mode of stratification in which $N_1 = 18$, $N_2 = 10$ and $N_3 = 7$, stratified πPS sampling with the allocation (3,3,3) is superior to unstratified πPS sampling for values of $g \leq 1.5$ (cf . Table 3.4.2). When the mode of stratification is such that $N_1 = 18$, $N_2 = 9$, $N_3 = 8$ the value of g_0 is between 1.4 and 1.5. Therefore, stratified πPS sampling with the given allocation (3,3,3) is superior to unstratified πPS sampling for values of g upto 1.4 (cf. Table 3.4.4). From Table 3.4.5 it is clear that with the mode of stratification in which $N_1 = 20$, $N_2 = 8$, $N_3 = 7$ stratified πPS sampling is superior to unstratified πPS sampling for values of $g \leq 1.6$.

CHAPTER 4

CHOICE OF ESTIMATORS UNDER CERTAIN LINEAR REGRESSION MODELS

4.0 Summary

In this chapter, the problem of estimating the population total of a characteristic "$\mathcal{Y}$" (the study variable) is studied under a certain 'super-population' regression model when auxiliary information on a closely related characteristic "$\mathcal{X}$" is available. Using the "ξ-unbiasedness" criterion (Royall (1970)), the class of variance functions in the super-population model for which a given linear estimator is "optimal" is derived and its rôle is studied in the construction of "optimal" estimators. The estimator $\hat{Y}_2$ which is optimal for the variance function $V(x)=x^2$ in our model is found to fare better than some other well known estimators and motivated by this a comparison between any general linear estimator and $\hat{Y}_2$ is presented. A general 'class of Symmetrized Des Raj estimators' of which the estimators suggested by Basu (1970) as alternatives to Murthy's (1957) estimator are particular cases, is considered and sufficient conditions are given for selection of the best in this class. A few further generalizations are mentioned. Towards the end of the chapter, numerical examples are provided to illustrate the results obtained. An appendix at the end of the chapter explains some calculations needed for the illustrations.

4.1 Introduction

Consider a finite population of N identifiable units

$$U = \{1, 2, \ldots, N\}. \tag{4.1-1}$$

As before, let $\mathcal{Y}$ and $\mathcal{X}$ be real positive valued characteristics taking values y_i and x_i respectively on unit i, $1 \leq i \leq N$ where the x_i's are known and y_i's are fixed but unknown. $\mathcal{Y}$ is the study variable and $\mathcal{X}$ is the auxiliary information available closely related to $\mathcal{Y}$. We are interested in estimating the population total $\sum_{i=1}^{N} y_i = Y$. We consider in this chapter only those sampling plans for which the expected sample size is fixed, say n, i.e., given a design $d = (S, p)$, $E(n(s)) = \sum_{s \in S} n(s)\, p_s = n$. Recall that a general homogeneous linear estimator $\hat{Y}$ of the population total Y is given by $\hat{Y} = \sum_{i \in s} \beta_{si}\, y_i$ for every s in S where β_{si} is a function of the sample s and the units in s.

The problem of estimating Y will be studied under the following model ξ: $y_1, y_2, \ldots, y_N$ are assumed to be realized values of independent random variables with joint distribution ξ. We shall use y_i to denote both the random variable and the value it takes on unit i given a finite population of size N; y_i has mean βx_i and variance $\sigma_i^2 = \sigma^2 V(x_i)$. The function V is known with $V(x) > 0$ for $x > 0$. The constants β and σ^2 are unknown (cf. (3.1.2)).

Definition 4.1.1 : Given the prior distribution ξ, a strategy $(p : \hat{Y})$ is said to be better than another strategy $(p' : \hat{Y}')$ ($\hat{Y}, \hat{Y}'$ are estimators of Y) if

$$E_\xi \; M(p : \hat{Y}) \leq E_\xi \; M(p' : \hat{Y}') \qquad (4.1.2)$$

where $M(p : \hat{Y}) = \sum_{s \epsilon S} (\hat{Y} - Y)^2 p_s$ is the sampling mean square error of $\hat{Y}$. E_ξ denotes expected value with respect to the probability distribution ξ.

In (4.1.2) if strict inequality holds, say that the strategy $(p : \hat{Y})$ is superior to $(p' : \hat{Y}')$ ($(p' : \hat{Y}')$ is inferior to $(p : \hat{Y})$). If equality holds in (4.1.2) say that the strategies $(p : \hat{Y})$ and $(p' : \hat{Y}')$ are equivalent for estimating Y.

We next observe that any estimator $\hat{Y}$ of Y can be uniquely expressed in the form

$$\hat{Y} = \sum_{i \epsilon s} y_i + \hat{\beta} \; (X - \sum_{i \epsilon s} x_i) \qquad (4.1.3)$$

for $s \epsilon S$ with $p_s > 0$ where $\hat{\beta}$ is defined by this expression and will be referred to as the " implied estimator for β ". Clearly $\hat{\beta}$ does not depend on the unobserved y_i's (i.e., the y-values for $i \not\epsilon s$) by definition of an estimator. The following shows that, of two estimators $\hat{Y}'$ and $\hat{Y}''$ of Y, the one implied estimator for β is better, is the better estimator

4.1.2 (Royall (1970)) : For any sampling plan p, if estima- $\hat{Y}'$ and $\hat{Y}''$ of Y have implied estimators $\hat{\beta}'$ and $\hat{\beta}''$ which satisfy

$$E_{\xi}\,(\hat{\beta}' - \beta)^2 \leq E_{\xi}(\hat{\beta}'' - \beta)^2, \tag{4.1.4}$$

h s in S with $p_s > 0$, then

$$E_{\xi}\, M(p : \hat{Y}') \leq E_{\xi}\, M(p : \hat{Y}''). \tag{4.1.5}$$

some s_o with $p_{s_o} > 0$ the inequality in (4.1.4) is strict, rict inequality holds in (4.1.5).

The basic approach in this chapter, following Royall (1970) ved from recognition of the fact that after the sample is , the population total can be written as

$$Y = \sum_{i\varepsilon s} y_i + (Y - \sum_{i\varepsilon s} y_i)$$

e first sum is known and the second must be estimated sample. Thus, regardless of how s was selected, the an estimator

$$\hat{Y} = \sum_{i\varepsilon s} y_i + (\hat{Y} - \sum_{i\varepsilon s} y_i),$$

for every s in S with $p_s > 0$ is the error in $(\hat{Y} - \sum_{i \in s} y_i)$ as an estimator of the sum $\sum_{i \in \bar{s}} y_i$ of the unobserved y-values.

Under the model ξ, after the sample is selected the inference problem can be considered to be simply the classical one of predicting the sum of the unobserved random variables $y_i (i \notin s)$ and the sample s should be one which permits a good predictor to be constructed.

Recall that given a sampling plan p, an estimator $\hat{Y}$ of Y is said to be p-unbiased for Y, if whatever the numbers $y_1, y_2, \ldots, y_N$,

$$\sum_{s \in S} \hat{Y} \; p_s = Y . \tag{4.1.6}$$

Under the model ξ, we have the following alternative definition of "unbiasedness".

Definition 4.1.3 (Royall (1970)) : For a given sampling plan p, an estimator $\hat{Y}$ is said to be ξ-unbiased for Y if $E_\xi(\hat{Y} - Y) = 0$ for every s such that $p_s > 0$. Equivalently $\hat{Y}$ is ξ-unbiased for Y if, and only if, $E_\xi(\hat{\beta} - \beta) = 0$ for every s such that $p_s > 0$ (since $E_\xi(\hat{Y} - Y) = (X - \sum_{i \in s} x_i)\, E_\xi(\hat{\beta} - \beta)$ and all the x_i's are positive).

Thus, $\hat{Y}$ is ξ-unbiased for Y if, and only if, the lied estimator $\hat{\beta}$ is ξ-unbiased for β, To complete the introduction on "ξ-unbiasedness", we observe that an estimator can be p-unbiased but not ξ-unbiased for Y and vice versa. For example, consider the simple random sampling plan p in which

$$p_s = \begin{cases} \dfrac{1}{\binom{N}{n}} & \text{if} \quad n(s) = n \\ 0 & \text{otherwise,} \end{cases}$$

then under the present model ξ, N times the sample mean is a p-unbiased estimator for Y, but since

$$E_\xi \left((N \sum_{i \in s} y_i/n) - Y \right) = \beta((N \sum_{i \in s} x_i/n) - X),$$

it is not ξ-unbiased for Y. On the other hand, the classical ratio estimator $\hat{Y}_{ratio} = (\sum_{i \in s} y_i / \sum_{i \in s} x_i) \sum_{i=1}^{N} x_i$ is not p-unbiased, but it is easily seen to be ξ-unbiased for Y.

4.2 A class of Variance Functions

For any sampling plan p, we first derive in this section the "optimum linear ξ-unbiased estimator for Y for a given (x) of our model ξ".

Consider the most general linear estimator of Y given by

$$\hat{Y} = \sum_{i \in s} \beta_{si}\, y_i . \tag{4.2.1}$$

Observe that $\hat{Y}$ is ξ-unbiased for estimating Y if, and only if, $\sum_{i\epsilon s} \beta_{si} x_i = X$ for each s in S with $p_s > 0$ where $X = \sum_{i=1}^{N} x_i$. For any sampling plan p, in view of Lemma 4.1.2, to minimize $E_\xi M(p : \hat{Y})$, it is enough to minimize $E_\xi(\hat{\beta} - \beta)^2$ for each s in S such that $p_s > 0$. Thus minimization of $E_\xi(\hat{\beta} - \beta)^2$ for each $s \epsilon S \ni p_s > 0$, subject to $\hat{Y}$ being ξ-unbiased for Y leads to the 'best linear ξ-unbiased estimator' for any sampling plan p, that has minimum expected (with respect to ξ) mean square error for a given $V(x)$ of our model.

Writing
$$\hat{Y} = \sum_{i\epsilon s} \beta_{si}\, y_i$$
$$= \sum_{i\epsilon s} y_i + \hat{\beta}\,(X - \sum_{i\epsilon s} x_i)$$

where
$$\hat{\beta} = \sum_{i\epsilon s} (\beta_{si} - 1)\, y_i \,/\, (X - \sum_{i\epsilon s} x_i) \qquad (4.2.2)$$

we have

$$E_\xi(\hat{\beta} - \beta)^2 = \sigma^2 \sum_{i\epsilon s} (\beta_{si} - 1)^2\, V(x_i)/(X - \sum_{i\epsilon s} x_i)^2. \qquad (4.2.3)$$

Introducing Lagrangian multiplier λ, minimization of (4.2.3) subject to $\sum_{i\epsilon s} \beta_{si} x_i = X$ gives

$$\beta_{si} = \{\lambda\, x_i\, (X - \sum_{i\epsilon s} x_i)^2 / 2\,\sigma^2\, V(x_i)\} + 1,$$

and consequently

$$\lambda = 1 / \left(\sum_{i\varepsilon s} (x_i^2 (X - \sum_{i\varepsilon s} x_i)/2\sigma^2 V(x_i) \right).$$

Hence the optimum linear ξ-unbiased estimator for a given $V(x)$ in our model ξ is

$$\hat{Y}^*_{\xi\text{-opt.}} = \sum_{i\varepsilon s} y_i + \frac{\sum_{i\varepsilon s} (x_i y_i / V(x_i))}{\sum_{i\varepsilon s} (x_i^2 / V(x_i))} (X - \sum_{i\varepsilon s} x_i) \qquad (4.2.4)$$

for every s in S such that $p_s > 0$ (cf. Royall (1970)). Observe that for a given $V(x)$ in our model ξ, $\hat{Y}^*_{\xi\text{-opt.}}$ is optimum for any sampling plan.

Now, when $\beta_{si} > 1$, $V(x_i) = c_s\, x_i/(\beta_{si} - 1)$, where c_s is a constant (> 0) which depends on all the units in the sample, when substituted in (4.2.4) gives $\hat{Y} = \sum_{i\varepsilon s} \beta_{si}\, y_i$. Thus $\hat{Y}$ of (4.2.1) when $\beta_{si} > 1$ for every $i \varepsilon s$, $s \varepsilon S$ with $p_s > 0$ is optimum with $V(x_i) = c_s\, x_i/(\beta_{si} - 1)$. In order that $V(x_i)$ be a function of x_i alone, we should have $c_s/(\beta_{si} - 1) = f(x_i)$, a function of x_i. This, then leads to the class of ξ-unbiased estimators for which

$$\beta_{si} = 1 + \frac{(X - \sum_{i\varepsilon s} x_i)/f(x_i)}{\sum_{\varepsilon} (x_i/f(x_i))} . \qquad (4.2.5)$$

When $f(x_i) = x_i^{-}, 1, x_i$

we have

$$\beta_{si} = 1 + \frac{(X - \sum_{i\varepsilon s} x_i)\, x_i}{\sum_{i\varepsilon s} x_i^2},$$

$$\beta_{si} = 1 + \frac{(X - \sum_{i\varepsilon s} x_i)}{\sum_{i\varepsilon s} x_i},$$

$$\beta_{si} = 1 + \frac{(X - \sum_{i\varepsilon s} x_i)}{x_i \; n(s)}$$

respectively and thus the estimators

$$\hat{Y}_0 = \sum_{i\varepsilon s} y_i + \frac{\sum_{i\varepsilon s} x_i y_i}{\sum_{i\varepsilon s} x_i^2} (X - \sum_{i\varepsilon s} x_i),$$

$$\hat{Y}_1 = \sum_{i\varepsilon s} y_i + \frac{\sum_{i\varepsilon s} y_i}{\sum_{i\varepsilon s} x_i} (X - \sum_{i\varepsilon s} x_i)$$

and $$\hat{Y}_2 = \sum_{i\varepsilon s} y_i + \frac{1}{n(s)} \sum_{i\varepsilon s} \frac{y_i}{x_i} (X - \sum_{i\varepsilon s} x_i)$$

of the population total Y which are optimum with corresponding variance functions $V(x_i) = 1, x_i, x_i^2$ respectively considered by Royall (1970) are obtained as special cases of the above general formulation. It can be easily verified that the choice $f(x_i) = x_i/(X-x_i)$ when the sampling plan p is such that $p_s > 0$ for $s \varepsilon S \Rightarrow n(s) = 2$ yields

$$\beta_{si} = (1 + p_i - \sum_{i \varepsilon s} p_i) / (2 - \sum_{i \varepsilon s} p_i) p_i$$

where $p_i = x_i/X$ for $i = 1,2,\ldots,N$, which corresponds to the well known Symmetrized Des Raj estimator $\hat{Y}_{SDR}$ when $n(s) = 2$. Thus the Symmetrized Des Raj estimator with a fixed sample size (= 2) plan is ξ-optimum when the variance function in our model ξ is given by $V(x_i) = x_i^2/(X - x_i)$. When $f(x_i) = n(s)\ x_i/(X - (n(s)\ x_i))$, we have

$$\beta_{si} = 1 + \frac{X - (n(s)\ x_i)}{n(s)x_i}$$

which corresponds to the estimator $X \sum_{i \varepsilon s} (y_i/(n(s)\ x_i))$. Thus, the estimator

$$\hat{Y}_{R_{n(s)}} = X \sum_{i \varepsilon s} \frac{y_i}{n(s)x_i}$$

is ξ-optimum with $V(x_i) = n(s)\ x_i^2/(X - (n(s)x_i))$. Observe that $\hat{Y}_{R_{n(s)}} = \hat{Y}_{R_n} = \hat{Y}_{HT}$ if the design is πPS with fixed sample size (= n). Furthermore, when β_{si} is a function of i alone, e the choice of c_s does not play any significant rôle in the truction of optimum $\hat{Y}$, we can as well consider the models h $V(x_i) = x_i/(\beta_i - 1)$ in such cases.

Comparison Between Estimators

In the previous section 4.2 we derived the best linear nbiased estimator for a given $V(x)$ of our model ξ that is ptimum. In this section 4.3 we first motivate certain questions t arise naturally in the context of ξ-unbiasedness and then wer them using an approach similar to Royall (1970). Our results this section compare any general linear ξ-unbiased estimator $\sum_{i \varepsilon s} \beta_{si}\, y_i$ ($s \varepsilon S$ with $p_s > 0$) with the ξ-unbiased estimator which is optimum when the variance function is $V(x) = x^2$ in model ξ. The motivation for comparing any general linear iased estimator $\hat{Y}$ with $\hat{Y}_2$ is as follows:

It has been shown (Godambe (1955), Godambe and Joshi (1965)) for sampling plans with fixed sample size (= n), under the l ξ with $V(x) = x^2$

$$E_\xi \; M(\pi PS : \hat{Y}_{R_n}) \leq E_\xi \; M(p : \hat{Y}) \tag{4.3.1}$$

$\hat{Y}$ that is p-unbiased for Y. But when p-unbiasedness is not ed, the optimum ξ-unbiased estimator $\hat{Y}_2$ with $V(x) = x^2$ r model is such that

$$E_{\xi} \ M(\pi PS : \hat{Y}_2) \leq E_{\xi} \ M(\pi PS : \hat{Y}_{R_{n(s)}}).$$

In particular if the sampling plan is πPS with fixed sample size (=n) then

$$E_{\xi} \ M(\pi PS : \hat{Y}_2) \leq E_{\xi} \ M(\pi PS : \hat{Y}_{R_n}) \qquad (4.3.2)$$

where $\hat{Y}_{R_n} = \frac{X}{n} \sum_{i \varepsilon s} \frac{y_i}{x_i}$.

Royall (1970) proved that the relation (4.3.2) holds good not only for πPS sampling plans with fixed sample size (= n) and $V(x) = x^2$, but for any sampling plan with fixed sample size (= n) and a wide class of variance functions which are such that $V(x)/x^2$ is a nonincreasing function. Working with p-unbiasedness, under the particular model ξ with $V(x) = x^g$, $g \varepsilon [1,2]$ we have from Vijayan (1966) that there exists a $g_o \varepsilon [1,2)$ such that for $g < g_o$, when $n(s) = n = 2$ for every $s \varepsilon S$ with $p_s > 0$

$$E_{\xi} \ M(p_{SDR} : \hat{Y}_{SDR}) \leq E_{\xi} \ M(\pi PS : \hat{Y}_{R_n}) \qquad (4.3.3)$$

where p_{SDR} is the sampling design given by

$s \varepsilon S$, $n(s) = 2 \Rightarrow p_s = \frac{p_i \, p_j}{1 - p_i} + \frac{p_j \, p_i}{1 - p_j}$ where $p_i = x_i/X$ for $i = 1,2,\ldots,N$; $X = \sum_{i=1}^{N} x_i$ and $i,j \varepsilon s$, and $\hat{Y}_{SDR}$ is the

Symmetrized Des Raj estimator given by

$$\hat{Y}_{SDR} = \frac{1}{2 - p_i - p_j} \left\{ \frac{y_i}{p_i} (1 - p_j) + \frac{y_j}{p_j} (1 - p_i) \right\} \qquad (4.3.4)$$

for $i, j \in s$ with $n(s) = 2$.

Let us now look at the well known Rao-Hartley-Cochran strategy (Rao et. al. (1962)) defined as follows :

Let p_t be the probability of drawing the t-th unit in the first draw from the whole population of size N. For example, if we are sampling with probability proportional to size, namely x_t, $p_t = x_t / \Sigma x_t$. The sampling procedure (to be denoted by p_{RHC}) consists of the following two stages :

(a) split the population at random into n groups of sizes $N_1, N_2, \ldots, N_n$ where $N_1 + N_2 + \ldots + N_n = N$.

(b) Draw a sample of size one with probabilities proportional to p_t from each of these n groups independently.

If group i denotes the group containing i-th unit the actual probability that it will be selected is $p_i / (\sum_{\text{group } i} p_t)$.

The estimator of the population total Y to be used with the above sampling plan p_{RHC} is

$$\hat{Y}_{RHC} = \sum_{i=1}^{n} \frac{y_i}{p_i / (\sum_{\text{group } i} p_t)} \qquad (4.3.5)$$

where the suffixes $1,2,\ldots,n$ denote the n units selected from the n groups separately.

The estimator $\hat{Y}_{RHC}$ of (4.3.5) is p-unbiased for the population total Y when the p_{RHC} sampling plan is used. Further, with $V(x) = x^g$, $g \geq 0$ in our model ξ, for sampling plans with fixed sample size (= n), we have from J. N. K. Rao (1966)

$$E_\xi \; M(p_{RHC} : \hat{Y}_{RHC}) < E_\xi \; M(\pi PS : \hat{Y}_{R_n})$$

for $g < 1$ and that

$$E_\xi \; M(p_{RHC} : \hat{Y}_{RHC}) \geq E_\xi \; M(\pi PS : \hat{Y}_{R_n}) \qquad (4.3.6)$$

for $g \geq 1$ working under p-unbiasedness set up. Now, comparing (4.3.3) and (4.3.6) we have that there exists a $g_o \in [1,2)$ such that for $g \in [1, g_o)$ when $n(s) = n = 2$ for every $s \in S$ with $p_s > 0$

$$E_\xi \; M(p_{SDR} : \hat{Y}_{SDR}) \leq E_\xi \; M(p_{RHC} : \hat{Y}_{RHC}) \, . \qquad (4.3.7)$$

Considering the inequalities (4.3.3) and (4.3.7) since $\hat{Y}_{SDR}$ and $\hat{Y}_{RHC}$ are ξ-unbiased, it is natural to ask if $\hat{Y}_2$ is better than $\hat{Y}_{SDR}$ and $\hat{Y}_{RHC}$ as well when p-unbiasedness is not demanded for the same wide class of variance functions as considered by Royall (1970). This, then motivates the comparison between $\hat{Y}_2$ and any general <u>linear</u> ξ-unbiased estimator for sampling plans with fixed sample size (= n) which is given below.

<u>Theorem 4.3.1</u> : Let $\hat{Y} = \sum_{i \in s} \beta_{si} y_i$ for $s \in S$ with $p_s > 0$ be any ξ-unbiased estimator with $\beta_{si} \geq 1$ for every $i \in s$ with not all β_{si}'s equal to one. Then for any sampling plan p with fixed sample size $(= n)$

$$E_\xi \; M(p : \hat{Y}) \geq E_\xi \; M(p : \hat{Y}_2)$$

if, for $i, j \in s$ and $s \in S$ with $p_s > 0$

$$a_i \leq a_j \Rightarrow \frac{V(x_i)}{x_i^2} \leq \frac{V(x_j)}{x_j^2}$$

where for $s \in S$ and $i \in s$, $a_i = n\, x_i(\beta_{si} - 1)/(X - \sum_{i \in s} x_i)$.

<u>Proof</u> : We have

$$E_\zeta \; (\hat{\beta} - \beta)^2 - E_\zeta \; (\hat{\beta}_2 - \beta)^2$$

$$= \frac{\sigma^2}{n^2} \sum_{i \in s} [(a_i^2 - 1) \frac{V(x_i)}{x_i^2}] \cdot \qquad (4.3.8)$$

ce $\quad a_i^2 \leq a_j^2 \Rightarrow \frac{V(x_i)}{x_i^2} \leq \frac{V(x_j)}{x_j^2}$

ecause $\beta_{si} \geq 1$ and $a_i \leq a_j \Rightarrow \frac{V(x_i)}{x_i^2} \leq \frac{V(x_j)}{x_j^2}$),

also $\sum_{i \in s} (a_i^2 - 1) = \sum_{i \in s} (a_i - \bar{a})^2$ where $\bar{a} = \sum_{i \in} a_i/n = 1$

$$\geq 0 \quad \text{for each} \quad s \in S, \qquad (4.3.9)$$

proof of Theorem 4.3.1 is completed by the following lemma.

Lemma 4.3.2 (Royall (1970)): If $0 \leq b_1 \leq \cdots \leq b_m$ and $c_1 \leq c_2 \leq \cdots \leq c_m$ satisfy $c_1 + c_2 + \cdots + c_m \geq 0$, then $b_1c_1 + \cdots + b_mc_m \geq 0$.

Corollary 4.3.3:

If $x_i(\beta_{si} - 1) \leq x_j(\beta_{sj} - 1)$ implies $x_i \geq x_j$,

$i, j \in s$, $s \in S$ with $p_s > 0$ then

$$E_\xi \; M(p : \hat{Y}_2) \leq E_\xi \; M(p : \hat{Y}),$$

when $V(x)/x^2$ is a nonincreasing function where p is any plan with fixed sample size $(= n)$.

Corollary 4.3.4:

If $x_i(\beta_{si} - 1) \leq x_j(\beta_{sj} - 1)$ implies $x_i \leq x_j$,

$i, j \in s$, $s \in S$ with $p_s > 0$ then

$$E_\xi \; M(p : \hat{Y}_2) \leq E_\xi \; M(p : \hat{Y}),$$

if $V(x)/x^2$ is a non-decreasing function where p is any plan with fixed sample size $(= n)$.

Remark 4.3.5: Observe that Theorem 4.3.1 above gives a condition on the coefficients of a linear ξ-unbiased estimator $\hat{Y}$ of Y (based on the model ξ) which is sufficient to insure that $\hat{Y}$ has iformly (over all samples of fixed sample size n) greater ξ-variance" than $\hat{Y}_2$. Under any fixed sample size (= n) plan for all $s \in S$ (such that $p_s > 0$ only if $n(s) = n$) we have own that the condition $a_i \leq a_j \Rightarrow \frac{V(x_i)}{x_i^2} \leq \frac{V(x_j)}{x_j^2}$ for $i, j \in s$ sufficient to guarantee that $\hat{Y}_2$ is better than $\hat{Y}$ (where $_{si} \geq 1 \;\forall\; i \in s$ with not all β_{si}'s equal to one). We note at it is not possible to specify conditions on the coefficient a linear ξ-unbiased estimator $\hat{Y}$ such that the relation $M(p : \hat{Y}) \leq E_\xi \; M(p : \hat{Y}_2)$ holds where p is any fixed sample ize (= n) plan. This is so because our Lemma 4.3.2 requires hat the b_i's are non-negative. For example, it is easy to rify that the estimators $\hat{Y}_{R_n}$ (when used with any fixed sample e (= n) plan) and $\hat{Y}_{SDR}$ (when used with a fixed sample size 2) plan) are such that the relation $x_i(\beta_{si} - 1) \leq x_j(\beta_{sj} - 1)$ r $i, j \in s$, $s \in S$ with $p_s > 0 \Rightarrow x_i \geq x_j$. Therefore, in ew of Corollary 4.3.3, if $V(x) = x^g$, $g \in [0,2]$ in our model though $\hat{Y}_2$ is <u>better than</u> both $\hat{Y}_{R_n}$ and $\hat{Y}_{SDR}$ (since x^{g-2} a nonincreasing function for $g \in [0,2]$) we are not able to sert that $\hat{Y}_{R_n}$ ($\hat{Y}_{SDR}$) is better than $\hat{Y}_2$ if $V(x) = x^g$, $g > 2$ $(x)/x^2$ is now a non-decreasing function). Towards the end of chapter we compare the performances of the estimators

$\hat{Y}_1$, $\hat{Y}_2$, $\hat{Y}_{R_n}$, $\hat{Y}_{SDR}$, $\hat{Y}_{RHC}$ for different samples from a small population when $V(x) = x^g$ in our model ξ for values of g ranging from 0 to 3.0.

Remark 4.3.6 : Considering (4.3.8), we have that

$$E_\xi(\hat{\beta}-\beta)^2 - E_\xi(\hat{\beta}_2-\beta)^2 = \frac{\sigma^2}{n^2} \sum_{i\epsilon s} (a_i+1)(a_i-1) \frac{V(x_i)}{x_i^2} \text{ .} \qquad (4.3.10)$$

When $a_i = 1 \ \forall \ i \ \epsilon \ s$, $s \ \epsilon \ S$ with $p_s > 0$ we have that $\hat{\beta}$ reduces to $\hat{\beta}_2$ and the right hand side (r.h.s) of (4.3.10) is zero. Furthermore, when $V(x_i) \ (a_i+1)/x_i^2 = c$, a constant for $i \ \epsilon \ s$, the r.h.s. of (4.3.10) is zero. This gives $a_i = (cx_i^2 - V(x_i))/V(x_i)$, where $c = 2n/(\sum_{i\epsilon s} (x_i^2/V(x_i)))$, since $\sum_{i\epsilon s} a_i = n$. This then implies that for $s \ \epsilon \ S$ and $i \quad s$,

$$\frac{n \ x_i(\beta_{si} - 1)}{(X - \sum_{i\epsilon s} x_i)} = \frac{c \ x_i^2 - V(x_i)}{V(x_i)}$$

which gives

$$\begin{aligned}\hat{Y} &= \sum_{i\epsilon s} \beta_{si} \ y_i \\ &= \sum_{i\epsilon s} y_i + (X - \sum_{i\epsilon s} x_i) \frac{(\sum_{i\epsilon s} (\beta_{si}-1)y_i)}{(X - \sum_{i\epsilon s} x_i)} \\ &= \sum_{i\epsilon s} y_i + (X - \sum_{i\epsilon s} x_i) \left\{ \frac{2 \sum_{i\epsilon s}(x_i y_i/V(x_i))}{\sum_{i\epsilon s} (x_i^2/V(x_i))} - \frac{1}{n} \sum_{i\epsilon s} \frac{y_i}{x_i} \right\} \qquad (4.3.11)\end{aligned}$$

for which the corresponding implied estimator $\hat{\beta} = 2\,\hat{\beta}^*_{\xi\text{-opt.}} - \hat{\beta}_2$ where $\hat{\beta}^*_{\xi\text{-opt.}}$ is the optimum implied estimator for β under the given model ξ with variance function $V(x)$. Here s is such that $n(s) = n \;\forall\; s \in S$ with $p_s > 0$. Thus, from (4.3.10) we infer that $E_\xi(\hat{\beta} - \beta)^2 = E_\xi(\hat{\beta}_2 - \beta)^2$ where for every s such that $p_s > 0$ ($\Rightarrow n(s) = n$), $\hat{\beta} = \hat{\beta}_2$ or $\hat{\beta} = 2\,\hat{\beta}^*_{\xi\text{-opt.}} - \hat{\beta}_2 = \hat{\beta}_2 + 2(\hat{\beta}^*_{\xi\text{-opt.}} - \hat{\beta}_2)$. In general, for any ξ-unbiased $\hat{\beta}'$ different from the optimum $\hat{\beta}^*_{\xi\text{-opt.}}$ noticing that $\hat{\beta}^*_{\xi\text{-opt.}} - \hat{\beta}'$ is a zero function, one can in fact consider $\hat{\beta} = \hat{\beta}' + \lambda(\hat{\beta}^*_{\xi\text{-opt.}} - \hat{\beta}')$ and check by substituting the values $\hat{\beta}^*_{\xi\text{-opt.}}$ and $\hat{\beta}'$ that $E_\xi(\hat{\beta} - \beta)^2 = E_\xi(\hat{\beta}' - \beta)^2$ for every s such that $p_s > 0$ ($\Rightarrow n(s) = n$), if, and only if, $\lambda = 0$ or 2.

4.4 Choice of Optimal Sampling Plan

We have discussed the choice of optimum estimator for any sampling plan for a given $V(x)$ of our model ξ in the previous sections. We also studied the estimator $\hat{Y}_2$ (which is optimal for $V(x) = x^2$) under a wide class of variance functions of the model ξ. In this section, we discuss the choice of optimal sampling plans.

Let S_n denote the collection of all samples s in S

with $n(s) = n$ and let P_n denote the collection of all sampling plans p with fixed sample size n. Let s^* be any sample for which

$$\max_{s \in S_n} \{ \sum_{i \in s} x_i \} = \sum_{s^*} x_i \tag{4.4.1}$$

and let $p^*_{\xi\text{-opt.}}$ be the sampling plan which selects s^* with certainty, i.e., $p^*_{(\xi\text{-opt.})s^*} = 1$.

Then, following Royall (1970) we have for any p in P_n and any $V(x)$ for which both $V(x)$ and $x^2/V(x)$ are non-decreasing

$$E_\xi M(p^*_{\xi\text{-opt.}} : \hat{Y}_2) \leq E_\xi M(p : \hat{Y}_2). \tag{4.4.2}$$

[For any plan p in P_n, $E_\xi M(p : \hat{Y}_2) = \sum_{s \in S_n} E_\xi (\hat{Y}_2 - Y)^2 p_s$ $= \sum_{s \in S_n} \{ E_\xi (\hat{\beta}_2 - \beta)^2 (X - \sum_{i \in s} x_i)^2 + \sigma^2 \sum_{i \in \bar{s}} V(x_i) \} p_s$ and for $p^*_{\xi\text{-opt.}} \in P_n$ when both $V(x)$ and $x^2/V(x)$ are nondecreasing $E_\xi M(p : \hat{Y}_2)$ is minimized. Here $\sum_{i \in \bar{s}}$ denotes summation over the units not in s].

Remark 4.4.1 : Thus, in view of Corollary 4.3.3 and the inequality (4.4.2), if the coefficients of a linear ξ-unbiased estimator $\hat{Y} = \sum_{i \in s} \beta_{si} y_i$ (for $s \in S$ with $p_s > 0$) are such that

$x_i(\beta_{si} - 1) \leq x_j(\beta_{sj} - 1) \Rightarrow x_i \geq x_j$ $(i, j \in s$ for $s \in S$ with $p_s > 0)$, then when $V(x)$ is non-decreasing and $V(x)/x^2$ is non-increasing, for all these wide class of variance functions the estimator $\hat{Y}_2$ with the purposive sampling plan $p^*_{\xi\text{-opt.}}$ fares better than the ξ-unbiased estimator $\hat{Y}$ with any fixed sample size (= n) plan for any β, σ^2 of our model ξ.

4.5 Choice from a class of Symmetrized Des Raj Estimators

In this section we consider some classical estimators of Y that depend on the order of selection of units in the sample (Des Raj (1956)) and show how our ξ-unbiasedness criterion working with a sample as a subset of units of U helps obtain an "optimum estimator". Therefore, we make it clear when we deal with samples that are finite ordered sequences of units from U. We use $\mathcal{X}$ to denote the size measure (auxiliary information). In literature, Probability Proportional to Size Without Replacement design (PPSWOR) of fixed sample size (= n) corresponds to the selection procedure in which the n units in the sample are obtained as follows : the first unit in the sample is obtained by selecting a unit i from the N population units with selection probabilities $p_i = x_i/X$ for $i = 1, 2, \ldots, N$; p_i's are also referred to as the initial probabilities of selection. Having obtained the first unit of the sample, the second unit is obtained choosing one from the remaining (N-1) units using selection obabilities that are proportional to the sizes (the x_i 's) of e remaining units. The procedure is thus a without replacement ocedure, at each stage of selection the selection probabilities units being proportional to size and terminates when n units e obtained. Thus, this procedure gives rise to an ordered mple of n units, all distinct. Therefore, PPSWOR design with xed sample size $n = (S,p)$ in our notation where $= \{s = (i_1, i_2, \ldots, i_n)\}$ is the collection of all ordered sequences of distinct units from U, with length n and for $s \in S$, $(i_1, i_2, \ldots, i_n)$ is the probability of obtaining the particular quence (sample) $s = (i_1, i_2, \ldots, i_n)$ using the selection procedure scribed above.

For the estimation of the population total $Y = \sum_{i=1}^{N} y_i$, ing the PPSWOR design with fixed sample size (= n), Des Raj(1956) ggested an estimator based on the order in which the units are lected in the sample as follows :

Let $y_{i_1}, y_{i_2}, \ldots, y_{i_n}$ and $p_{i_1}, p_{i_2}, \ldots, p_{i_n}$ be respectively $\mathcal{Y}$-values of the units in the sample $s = (i_1, i_2, \ldots, i_n)$ and ir initial probabilities of selection in the PPSWOR design of ed sample size n. Then Des Raj's estimator is given by

$$\hat{Y}_D = \sum_{k=1}^{n} c_k t_k \qquad (4.5.1)$$

here

$$= y_{i_1}+y_{i_2}+\dots+y_{i_{k-1}}+\frac{y_{i_k}}{p_{i_k}}(1 - p_{i_1} - p_{i_2} - \dots - p_{i_{k-1}}) \qquad (4.5.2)$$

nd c_k's are constants such that $\sum_{k=1}^{n} c_k = 1$.

ordering this estimator, Murthy (1957) considered

$$\hat{Y}_{SDR} = \Sigma \hat{Y}_D \,(p_{s,(i_1,i_2,\dots,i_n)} / \Sigma \; p_{s,(i_1,i_2,\dots,i_n)}) \qquad (4.5.3)$$

here the summation in (4.5.3) is over all $n!$ permutations of $i_1,i_2,\dots,i_n$).

Murthy (1957) proved that the strategy (p_{SDR} : $\hat{Y}_{SDR}$) is etter than the strategy (PPSWOR : $\hat{Y}_D$) in the exact variance sense, here p_{SDR} is the design with fixed sample size (= n) in which he sample s as a set with elements $i_1,i_2,\dots,i_n$ is chosen with robability $\Sigma\, p_{s,(i_1,i_2,\dots,i_n)}$, where Σ is as in (4.5.3), ummation over all $n!$ permutations of $(i_1,i_2,\dots,i_n)$.

Basu (1970) considers

$$t_n = \sum_{i_j \in s} y_{i_j} + \frac{y_{i_n}}{p_{i_n}} (1 - \sum_{i_j \in s} p_{i_j}) \qquad (4.5.4)$$

(4.5.2) and writes it as $S_1 + S_2$ where S_2 is an estimate of $- \sum_{i_j \in s} y_{i_j}$. It is clear that S_2 would be an exact estimate of

$\cdot \sum_{i_j \in s} y_{i_j}$ if, and only if, $y_{i_n}/x_{i_n} =$

$Y - \sum_{i_j \in s} y_{i_j})/(X - \sum_{i_j \in s} x_{i_j})$ = weighted average of unobserved tios, weights being the sizes of the corresponding units. Basu an suggests that it is 'natural' for the surveyor to estimate $- \sum_{i_j \in s} y_{i_j})/(X - \sum_{i_j \in s} x_{i_j})$ by some sort of average of observed tios such as $(\sum_{i_j \in s} y_{i_j})/(\sum_{i_j \in s} x_{i_j})$ or $n^{-1} \sum_{i_j \in s} (y_{i_j}/x_{i_j})$. se then lead to the estimators

$$\hat{Y}_{B_1} = \sum_{i \in s} y_{i_j} + \frac{(\sum_{i_j \in s} y_{i_j})}{(\sum_{i_j \in s} x_{i_j})} (X - \sum_{i_j \in s} x_{i_j})$$

$$= \sum_{i \in s} y_i + \frac{\sum_{i \in s} y_i}{\sum_{i \in s} x_i} (X - \sum_{i \in s} x_i) \tag{4.5.5}$$

$$\hat{Y}_{B_2} = \sum_{i_j \in s} y_{i_j} + \frac{1}{n} \sum_{i_j \in s} \frac{y_{i_j}}{x_{i_j}} (X - \sum_{i_j \in s} x_{i_j})$$

$$= \sum_{i \in s} y_i + \frac{1}{n} \sum_{i \in s} \frac{y_i}{x_i} (X - \sum_{i \in s} x_i) . \tag{4.5.6}$$

erve that $\hat{Y}_{B_1}$ and $\hat{Y}_{B_2}$ are the same as $\hat{Y}_1$ and $\hat{Y}_2$ of the ious sections.Basu recommends them because of their simple ure. But observe that $\hat{Y}_{B_1}$ and $\hat{Y}_{B_2}$ are to be used with a

PPSWOR plan and hence they are p-biased.

On the other hand, considering

$$\hat{Y}^{+} = \Sigma \; w_{i_1 i_2 \ldots i_n} \, t_n \qquad (4.5.7)$$

where t_n is the estimator of (4.5.2) (based on the order $(i_1, i_2, \ldots, i_n)$ of units) and $w_{i_1 i_2 \ldots i_n}$ is the weight attached to t_n such that $\Sigma \; w_{i_1 i_2 \ldots i_n} = 1$, summation being taken over all $n!$ permutations of $(i_1, i_2, \ldots, i_n)$, it is easily seen that the choices $w_{i_1 i_2 \ldots i_n} = p_{i_n} / (n-1)! \sum_{i_j \varepsilon s} p_{i_j}$ and $1/(n!)$ lead to $\hat{Y}_{B_1}$ and $\hat{Y}_{B_2}$ respectively while $w_{i_1 i_2 \ldots i_n} = p_{s,(i_1, i_2, \ldots, i_n)} / \Sigma \; p_{s,(i_1, i_2, \ldots, i_n)}$ gives $\hat{Y}_{SDR}$ of (4.5.3) when $c_1 = c_2 = \ldots = c_{n-1} = 0$ and $c_n = 1$. Thus, for various choices of $w_{i_1 i_2 \ldots i_n}$ in (4.5.7) we have a 'class of Symmetrized Des Raj estimators' and the problem now is to choose the "best estimator" in this class. The sampling designs with which the estimator $\hat{Y}^{+}$ of (4.5.7) is to be used are designs $d = (S, p)$ where $S = \{s\}$ is the collection of all non-empty subsets of units of U and p_s for $s \varepsilon S$ is > 0 only if $n(s) = n$.

If p-unbiasedness is not demanded, since $\hat{Y}^{+}$ is ξ-unbiased [since $E_{\xi}(\hat{Y}^{+})$

$$\Sigma w_{i_1 i_2 \cdots i_n}\Big(\sum_{i_j \varepsilon s} \beta x_{i_j} + \frac{\beta x_{i_n}}{p_{i_n}}\Big(1 - \sum_{i_j \varepsilon s} p_{i_j}\Big)\Big)$$

$$\beta(\Sigma w_{i_1 i_2 \cdots i_n})\, X \text{ for every } s \in S \text{ with } p_s > 0\,],$$

sing the theory developed in the previous sections of this hapter, we have

$$\hat{Y}^{+} = \sum_{i\varepsilon s} \beta_{si}\, y_i\,, \qquad (4.5.8)$$

ith

$$\beta_{si} = 1 + [(\Sigma' w_{i_1 i_2 \cdots i_{n-1} i})(1 - \sum_{i\varepsilon s} p_i)/p_i] \qquad (4.5.9)$$

r $s \in S$ with $p_s > 0$. Observe that the sample s in (4.5.8) the set consisting of the units $i_1, i_2, \ldots, i_n$. The summation in (4.5.9) is over all $(n-1)!$ permutations of $i_1, i_2, \ldots, i_{n-1}$), the first (n-1) units in the sample s with the st unit being i when the sample s was considered as a finite ered sequence $(i_1, i_2, \ldots, i_{n-1}, i)$ of distinct units from U.

Thus, from (4.5.8) and (4.5.9) we have according to our vious notation, for $i \varepsilon s$ and $s \in S$ with $p_s > 0$ ($\Rightarrow n(s)=n$),

$$a_i = n\, \Sigma'\, w_{i_1 i_2 \cdots i_{n-1} i} \qquad (4.5.10)$$

h Σ' is as in (4.5.9).

Therefore, from Theorem 4.3.1 we have

Theorem 4.5.1 : For any sampling plan p with fixed sample size (= n), under the model ξ (with variance function $V(x)$)

$$E_\xi \; M(p : \hat{Y}^+) \geq E_\xi \; M(p : \hat{Y}_{B_2}) \qquad (4.5.11)$$

if for $i, j \in s$ and $s \in S$ with $p_s > 0$

$$a_i \leq a_j \Rightarrow \frac{V(x_i)}{x_i^2} \leq \frac{V(x_j)}{x_j^2} ,$$

where for $i \in s$, $a_i = n \Sigma' w_{i_1 i_2 \ldots i_{n-1} i}$.

Corollary 4.5.2 : It is easy to see that the choice $w_{i_1 i_2 \ldots i_{n-1} i} = 1/(n!)$ will give $a_i = 1 \; \forall \; i \in s$ and in this case $\hat{Y}^+$ reduces to $\hat{Y}_{B_2}$ and equality occurs in (4.5.11).

Corollary 4.5.3 : When $w_{i_1 i_2 \ldots i_{n-1} i} = p_i/(n-1)! \; (\sum_{i \in s} p_i)$, $a_i = n \, p_i / (\sum_{i \in s} p_i)$ and $a_i \leq a_j \Rightarrow x_i \leq x_j$ for $i, j \in s$ with $p_s > 0$. Hence when $V(x)/x^2$ is non-decreasing,

$E_\xi \; M(p : \hat{Y}_{B_1}) \geq E_\xi \; M(p : \hat{Y}_{B_2})$ where p is a plan with fixed sample size (= n).

Corollary 4.5.4 : When $w_{i_1 i_2 \ldots i_{n-1} i} = (1-n \, p_i)/(n! (1- \sum_{i \in s} p_i))$, $a_i = (1 - n \, p_i)/(1 - \sum_{i \in s} p_i)$ and $a_i \leq a_j \Rightarrow x_i \geq x_j$ for $i, j \in s$ with $p_s > 0$. Hence when $V(x)/x^2$ is nonincreasing,

$E_\xi M(p : \hat{Y}_{R_n}) \geq E_\xi \; M(p : \hat{Y}_{B_2})$ where p is a plan with fixed sample size (= n).

Corollary 4.5.5 : For samples of size 2, when

$w_{i_1 i_2} = p_{s,(i_1,i_2)} / (p_{s,(i_1,i_2)} + p_{s,(i_2,i_1)})$ where $p_{s,(i_1,i_2)}$ is the probability of obtaining the ordered sample (i_1, i_2) in the PPSWOR design $(= \frac{p_{i_1} p_{i_2}}{(1 - p_{i_1})})$ we have $w_{i_1 i_2} = (1-p_{i_2})/(2-p_{i_1}-p_{i_2})$ and $w_{i_2 i_1} = (1-p_{i_1})/(2-p_{i_1}-p_{i_2})$. Therefore $a_i = 2(1-p_i)/(2-p_i-p_j)$ for $i \varepsilon s$ and hence $a_i \leq a_j \Rightarrow x_i \geq x_j$ for $i, j \varepsilon s$ with $p_s > 0$. Hence when $V(x)/x^2$ is nonincreasing, $E_\xi \; M(p_{SDR} : \hat{Y}_{SDR})$ $E_\xi \; M(p_{SDR} : \hat{Y}_{B_2})$ when p_{SDR} is the fixed sample size (= 2) design in which the sample s as a set with units i, j is chosen with probability $(\frac{p_i \, p_j}{1 - p_i} + \frac{p_j \, p_i}{1 - p_j})$.

Table 4.5.6 below gives the exact bias and exact mean square error $(\sum_{s \varepsilon S} (\hat{Y} - Y) \, p_s$ and $\sum_{s \varepsilon S} (\hat{Y} - Y)^2 \, p_s$ respectively) for the estimators $\hat{Y}_2$, $\hat{Y}_{R_n}$ and $\hat{Y}_{SDR}$ based on the p_{SDR} design for the three hypothetical populations (populations 1,2 and 3 below) considered by Yates and Grundy (1953) and Sampford's live data (Sampford (1962), p.61) with fixed sample size = 2.

	units	1	2	3	4
pulation 1	$\mathcal{Y}$	5	12	21	32
	$\mathcal{X}$	1	2	3	4
pulation 2	$\mathcal{Y}$	8	14	18	20
	$\mathcal{X}$	1	2	3	4
pulation 3	$\mathcal{Y}$	2	6	9	8
	$\mathcal{X}$	1	2	3	4
pulation 4		See Appendix AI			

Table 4.5.6

Bias and Mean Square Error of $\hat{Y}_2$, $\hat{Y}_{R_n}$ and $\hat{Y}_{SDR}$ with p_{SDR} design for $n(s) = n = 2$.

	Total number of samples	$\hat{Y}_2$	$\hat{Y}_{R_n}$	$\hat{Y}_{SDR}$
pulation 1				
Bias	6	.4366	-.9720	-.0000
$M(\hat{Y})$		30.2559	34.3254	31.2356
pulation 2				
Bias	6	-.4365	.9722	-.0000
$M(\hat{Y})$		30.2559	34.3254	31.2356
pulation 3				
Bias	6	-.0712	.2480	-.0000
$M(\hat{Y})$		7.2098	6.7956	7.0035
pulation 4				
Bias	595	-1.4962	1.6351	-.0000
$M(\hat{Y})$		106621.4460	107345.8630	106938.1290

If p-unbiasedness is demanded, $\hat{Y}_{SDR}$ is the unique p-unbiased estimator in this class (all the three estimators $\hat{Y}_2$, $\hat{Y}_{R_n}$, $\hat{Y}_{SDR}$ belong to the 'class of Symmetrized Des Raj estimators'). Table 4.5.6 shows that though biased, $\hat{Y}_2$ and $\hat{Y}_{R_n}$ for certain populations have smaller mean square error as compared to $\hat{Y}_{SDR}$ under the p_{SDR} design with fixed sample size (= 2). Therefore, no clear cut choice is possible between the estimators. On the other hand, if ξ-unbiasedness is used, based on our Corollaries 4.5.4 and 4.5.5 it follows that $\hat{Y}_{B_2} = \hat{Y}_2$ fares better than the other two estimators for the class of variance functions $V(x) = x^g$ for $g \in [0,2]$ (since x^{g-2} is nonincreasing).

Remark 4.5.7 : Writing t_n of (4.5.4) as $S_1 + S_2$ where S_2 is an estimate of $Y - \sum_{i_j \in s} y_{i_j}$, Basu (1970) suggested $\hat{Y}_{B_1}$ of (4.5.5) and $\hat{Y}_{B_2}$ of (4.5.6) by using (average of observed ratios) $(\sum_{i_j \in s} y_{i_j})/(\sum_{i_j \in s} x_{i_j})$ and $n^{-1} \sum_{i_j \in s} (y_{i_j}/x_{i_j})$ respectively as estimates of $(Y - \sum_{i_j \in s} y_{i_j})/(X - \sum_{i_j \in s} x_{i_j})$. Therefore, along the lines of Basu's recommendations we can as well consider

$$\frac{\sum_{i_j \in s} \lambda_{i_j} y_{i_j}}{\sum_{i_j \in s} \lambda_{i_j} x_{i_j}} \text{ as an estimate of } (Y - \sum_{i_j \in s} y_{i_j})/(X - \sum_{i_j \in s} x_{i_j})$$

hich gives

$$\hat{Y}_B = \sum_{i_j \varepsilon s} y_{i_j} + \frac{\sum_{i_j \varepsilon s} \lambda_{i_j} y_{i_j}}{\sum_{i_j \varepsilon s} \lambda_{i_j} x_{i_j}} (X - \sum_{i_j \varepsilon s} x_{i_j}) \qquad (4.5.12)$$

s an estimator of the population total Y. The λ_{i_j}'s in (4.5.12) re completely arbitrary. Observe that $\hat{Y}_B$ of (4.5.12) is btained if we substitute

$$_{i_1 i_2 \cdots i_n} = \frac{\lambda_{i_n} p_{i_n}}{(n-1)! \sum_{i_j \varepsilon s} \lambda_{i_j} p_{i_j}}$$

in (4.5.7). Further note that hen the weights $\lambda_{i_j} = 1$ for all $i_j \varepsilon s$, $s \varepsilon S$ with $p_s > 0$ have that $\hat{Y}_B = \hat{Y}_{B_1} = \hat{Y}_1$ and when $\lambda_{i_j} = x_{i_j}^{-1}$ for $i_j \varepsilon s$, εS with $p_s > 0$ we have that $\hat{Y}_B = \hat{Y}_{B_2} = \hat{Y}_2$. The

itimator $\hat{Y}_B$ of (4.5.12) with

$$_j = \frac{\sum_{i_j=1}^{N} (V(x_{i_j}))^{1/2}}{n(V(x_{i_j}))^{1/2}} - 1, \quad i_j \varepsilon s$$

has been studied by Brewer(1977) er the super-population model ξ with variance function $V(x)$. Again ing Theorem 4.5.1 we have that under the model ξ with variance nction $V(x)$, if $\lambda_i p_i \leq \lambda_j p_j \Rightarrow \frac{V(x_i)}{x_i^2} \leq \frac{V(x_j)}{x_j^2}$ for $i, j \varepsilon s$, εS with $p_s > 0$ then for any fixed sample size n) plan the relation $E_\xi M(p : \hat{Y}_2) \leq E_\xi M(p : \hat{Y}_B)$ holds.

Further Generalizations

Let $\hat{Y}_\alpha$ denote the estimator which is optimal when $V(x) = x^\alpha$ our model ξ, where α is any nonnegative real number (see .2.4)). If $0 \le r \le t$ and $V(x)/x^r$ is nonincreasing, then yall (1970) proved that $E_\xi \ M(p : \hat{Y}_r) \le E_\xi \ M(p : \hat{Y}_t)$ for ery p. The following theorem gives a comparison between $\hat{Y}_\alpha$ ere $\alpha \le 2$ and $\hat{Y} = \sum_{i \varepsilon s} \beta_{si} y_i$ with $\beta_{si} \ge 1 \ \forall \ i \varepsilon s$ and not l β_{si}'s equal to one.

eorem 4.6.1 : For any sampling plan p for which $p_s > 0$ for ε S only if $n(s) = n$,

$$E_\xi \ M(p : \hat{Y}) \ge E_\xi \ M(p : \hat{Y}_\alpha)$$

(i) $V(x)/x^\alpha$ is nonincreasing and

(ii) $x_i^\alpha (\beta_{si}-1)^2 \ge x_j^\alpha(\beta_{sj}-1)^2 \Rightarrow x_j \ge x_i$ for $i,j \varepsilon s$,

$s \varepsilon S$ with $p_s > 0$.

oof : Observe that

$$(\hat{\beta} - \beta)^2 - E_\xi(\hat{\beta}_\alpha - \beta)^2 = \sigma^2 \sum_{i \varepsilon s} (V(x_i) \ b_i/x_i^\alpha) \quad \forall \ s \varepsilon S \text{ with } p_s>0,$$

re $$b_i = \frac{x_i^\alpha (\beta_{si} - 1)^2}{(X - \sum_{i \varepsilon s} x_i)^2} - \frac{x_i^{2-\alpha}}{(\sum_{i \varepsilon s} x_i^{2-\alpha})^2} .$$

If conditions (i) and (ii) hold then $b_i \geq b_j$ for $i,j \in s$

lies $V(x_i)/x_i^{\alpha} \geq V(x_j)/x_j^{\alpha}$ and $\sum_{i \in s} b_i \geq 0$ by Cauchy-Schwarz quality. Therefore the proof is immediate from Lemma 4.3.2.

It can be checked that $\hat{Y}_{\alpha}$ which is the optimal estimator n $V(x) = x^{\alpha}$ can be obtained from (4.5.7) with

$_{i_2,\ldots i_n} = p_{i_n}^{2-\alpha} / (n-1)! \sum_{i_j \in s} p_{i_j}^{2-\alpha}$. Hence by Theorem 4.6.1 it llows that $\hat{Y}_{\alpha}$ is better than any other $\hat{Y}^{+}$ when the suffici- conditions (i) and (ii) of Theorem 4.6.1 are satisfied.

7 Illustrations

We illustrate the results obtained in the above tions of this chapter by using a hypothetical population of e 4 (Yates and Grundy (1953)) for which the $\mathcal{X}$-characteristic s the values 1,2,3,4 on the units $i = 1,2,3,4$ respectively. les of size 2 are drawn from this population. We assume her that the variance function of our model ξ is of the form

$= x^{g}$. Since for any given sample s with $p_s > 0$,

$$(\hat{Y} - Y)^2 = E_{\xi}(\hat{\beta} - \beta)^2 (X - \sum_{i \in s} x_i)^2 + \sigma^2 \sum_{i \in \bar{s}} V(x_i),$$

comparison purposes the values of $E_{\xi}(\hat{\beta} - \beta)^2/\sigma^2$ are lated in Table 4.7.1 through Table 4.7.7 for each one of the sible samples with fixed sample size $n = 2$ design for the estimators $\hat{Y}_1$, $\hat{Y}_2$, $\hat{Y}_{R_n}$, $\hat{Y}_{SDR}$, $\hat{Y}_{RHC}$, $\hat{Y}_g$ (the optimal estimator with $V(x) = x^g$ in our model ξ given in (4.2.4)) for $g = 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0$.

$E_\xi(\hat{\beta} - \beta)^2/\sigma^2$ when $V(x) = x^g$ in our model ξ for the estimators $\hat{Y}_1, \hat{Y}_2, \hat{Y}_{R_n}, \hat{Y}_{SDR}$, $\hat{Y}_{RHC}$ and $\hat{Y}_g$

Table 4.7.1

$g = 0$

samples	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}_{R_n}$	$\hat{Y}_{SDR}$	$\hat{Y}_{RHC}$	$\hat{Y}_g$
{1,2}	.2222	.3125	.3724	.3356	.3114	.2000
{1,3}	.1250	.2778	.4568	.3377	.2651	.1000
{1,4}	.0800	.2656	.6425	.3700	.2594	.0588
{2,3}	.0800	.0903	.1078	.0953	.0876	.0769
{2,4}	.0555	.0781	.1445	.0931	.0742	.0500
{3,4}	.0408	.0434	.0563	.0455	.0432	.0400

Table 4.7.2

$g = 0.5$

samples	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}_{R_n}$	$\hat{Y}_{SDR}$	$\hat{Y}_{RHC}$	$\hat{Y}_g$
{1,2}	.2682	.3384	.3915	.3586	.3374	.2612
{1,3}	.1707	.2981	.4658	.3532	.2866	.1614
{1,4}	.1200	.2813	.6450	.3800	.2755	.1111
{2,3}	.1258	.1365	.1580	.1425	.1335	.1246
{2,4}	.0948	.1196	.2067	.1384	.1149	.0923
{3,4}	.0762	.0794	.0994	.0824	.0791	.0758

Table 4.7.3

$g = 1.0$

Samples	$\hat{Y}_1 (=\hat{Y}_g)$	$\hat{Y}_2$	$\hat{Y}_{R_n}$	$\hat{Y}_{SDR}$	$\hat{Y}_{RHC}$
{1,2}	.3333	.3750	.4184	.3910	.3749
{1,3}	.2500	.3333	.4815	.3802	.3239
{1,4}	.2000	.3125	.6500	.4000	.3076
{2,3}	.2000	.2083	.2333	.2148	.2053
{2,4}	.1667	.1875	.2969	.2092	.1824
{3,4}	.1428	.1458	.1759	.1499	.1456

Table 4.7.4

$g = 1.5$

Samples	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}_{R_n}$	$\hat{Y}_{SDR}$	$\hat{Y}_{RHC}$	$\hat{Y}_g$
{1,2}	.4254	.4268	.4564	.4369	.4263	.4142
{1,3}	.3873	.3943	.5086	.4269	.3885	.3660
{1,4}	.3600	.3750	.6600	.4400	.3717	.3333
{2,3}	.3210	.3211	.3469	.3269	.3190	.3178
{2,4}	.3008	.3018	.4290	.3227	.2978	.2929
{3,4}	.2693	.2693	.3121	.2739	.2691	.2680

Table 4.7.5

$g = 2.0$

Samples	$\hat{Y}_1$	$\hat{Y}_2 (=\hat{Y}_g)$	$\hat{Y}_{R_n}$	$\hat{Y}_{SDR}$	$\hat{Y}_{RHC}$
{1,2}	.5556	.5000	.5102	.5017	.5001
{1,3}	.6250	.5000	.5556	.5078	.5004
{1,4}	.6800	.5000	.6800	.5200	.5001
{2,3}	.5200	.5000	.5200	.5022	.5008
{2,4}	.5556	.5000	.6250	.5102	.5009
{3,4}	.5102	.5000	.5556	.5029	.5001

Table 4.7.6

$g = 2.5$

samples	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}_{R_n}$	$\hat{Y}_{SDR}$	$\hat{Y}_{RHC}$	$\hat{Y}_g$
{1,2}	.7396	.6036	.5863	.5935	.6042	.5858
{1,3}	1.0368	.6830	.6369	.6479	.6942	.6340
{1,4}	1.3200	.7500	.7200	.6800	.7568	.6667
{2,3}	.8498	.7866	.7862	.7794	.7942	.7786
{2,4}	1.0460	.8536	.9205	.8291	.8679	.8284
{3,4}	.9712	.9330	.9920	.9282	.9341	.9282

Table 4.7.7

$g = 3.0$

samples	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}_{R_n}$	$\hat{Y}_{SDR}$	$\hat{Y}_{RHC}$	$\hat{Y}_g$
{1,2}	1.0000	.7500	.6939	.7232	.7515	.6667
{1,3}	1.7500	1.0000	.7778	.8906	1.0298	.7500
{1,4}	2.6000	1.2500	.8000	1.0000	1.2702	.8000
{2,3}	1.4000	1.2500	1.2000	1.2222	1.2721	1.2000
{2,4}	2.0000	1.5000	1.3750	1.3878	1.5462	1.3333
{3,4}	1.8571	1.7500	1.7778	1.7219	1.7538	1.7146

In Table 4.7.1 through Table 4.7.7 we compare the perfor-[manc]es of each of the estimators $\hat{Y}_1$, $\hat{Y}_2$, $\hat{Y}_{R_n}$, $\hat{Y}_{SDR}$, $\hat{Y}_{RHC}$ and $\hat{Y}_g$ [fo]r different samples from a small population for values of g [fr]om 0 through 3.0. Thus in each case, the basis of the compari-[so]n is $\hat{Y}_g$, the optimal estimator when $V(x) = x^g$ in our model ξ. [In] the case of $\hat{Y}_{R_n}$ and $\hat{Y}_{SDR}$, since for $i,j \in s$,

$a_1 \leq a_j \Rightarrow x_1 \geq x_j$, when $V(x) = x^g$ for $g \in [0,2]$, $\hat{Y}_2$ is better than $\hat{Y}_{R_n}$ and $\hat{Y}_{SDR}$ (since $V(x)/x^2 = x^{g-2}$ for $g \in [0,2]$ is non-increasing) as illustrated by $E_\xi(\hat{\beta}_2 - \beta)^2/\sigma^2$ as compared to $E_\xi(\hat{\beta}_{R_n} - \beta)^2/\sigma^2$ and $E_\xi(\hat{\beta}_{SDR} - \beta)^2/\sigma^2$ in Table 4.7.1, Table 4.7.2, Table 4.7.3, Table 4.7.4, Table 4.7.5 (these tables thus illustrate our Theorem 4.3.1).

Regarding the classical ratio estimator $\hat{Y}_1$, for the particular choice of $V(x) = x^g$ in our model, when the plan is a fixed sample size (n = 2) plan the estimators $\hat{Y}_2$, $\hat{Y}_{R_n}$ and $\hat{Y}_{SDR}$ all satisfy the sufficient condition (ii) of Theorem 4.6.1. Therefore, since $V(x)/x = x^{g-1}$ is nonincreasing for $g \leq 1$, $\hat{Y}_1$ fares better than these three estimators as illuatrated by $E_\xi(\hat{\beta}_1 - \beta)^2/\sigma^2$ as compared to $E_\xi(\hat{\beta}_2 - \beta)^2/\sigma^2$, $E_\xi(\hat{\beta}_{R_n} - \beta)^2/\sigma^2$ and $E_\xi(\hat{\beta}_{SDR} - \beta)^2/\sigma^2$ in Table 4.7.1, Table 4.7.2, Table 4.7.3; this comparison therefore illustrates our Theorem 4.6.1.

Regarding our Theorem 4.3.1 we have already made it clear that it is not possible to specify conditions on the coefficients of a linear ξ-unbiased estimator so that the relation $E_\xi M(p : \hat{Y}) \leq E_\xi M(p : \hat{Y}_2)$ holds (in view of Lemma 4.3.2). As stated in our Remark 4.3.5, though we are not able to assert that $e_{R_n}(\hat{Y}_{SDR})$ is better than $\hat{Y}_2$ when $V(x) = x^g$ for $g > 2$ (since

x^{g-2} is now a nondecreasing function) we have tabulated $E_\xi(\hat{\beta}-\beta)^2/\sigma^2$ for $\hat{Y}_1$, $\hat{Y}_2$, $\hat{Y}_{R_n}$, $\hat{Y}_{SDR}$, $\hat{Y}_{RHC}$ and $\hat{Y}_g$ for $g = 2.5$ and 3.0 for comparison purposes. For the small population with $N = 4$ that we have chosen for our illustration, from Table 4.7.6 and Table 4.7.7 we note that the relation $E_\xi(\hat{\beta}_{R_n} - \beta)^2 \leq E_\xi(\hat{\beta}_2 - \beta)^2$ holds for four samples out of the six possible samples when $g = 2.5$ and for five samples out of the six possible samples for $g = 3.0$ respectively. Regarding the estimator $\hat{Y}_{SDR}$ when a fixed sample size (= 2) design is used, Table 4.7.6 and Table 4.7.7 indicate $\hat{Y}_{SDR}$ performing better than $\hat{Y}_2$ when $g = 2.5$ and 3.0 as judged by $E_\xi(\hat{\beta}_{SDR} - \beta)^2/\sigma^2$ and $E_\xi(\hat{\beta}_2 - \beta)^2/\sigma^2$ for all the six different samples. Though our theory developed in the previous sections of this chapter does not help us conclude this, some more empirical investigation is needed to understand the behaviour of $\hat{Y}_{SDR}$ (as compared to $\hat{Y}_2$) when the variance function of our model ξ is of the form $V(x) = x^g$ for $g > 2$ (so that x^{g-2} is nondecreasing) for fixed sample size (= 2) plans. However, for our population of 4 units with $x_i = 1, 2, 3, 4$ on the four units respectively, corresponding to the sample $\{3, 4\}$, $E_\xi(\hat{\beta}_{SDR} - \beta)^2$ > $E_\xi(\hat{\beta}_2 - \beta)^2$ when the variance function in our model ξ is $V(x) = x^g$ with $g = 2.1$ and 2.2.

Now, considering the estimators $\hat{Y}_1$ and $\hat{Y}_2$ note that $\hat{Y}_1$ such that $a_i \leq a_j \Rightarrow x_i \leq x_j$, $i, j \in s$, $s \in S$ with $p_s > 0$

$\Rightarrow n(s) = n$). Therefore when $V(x)/x^2$ is nondecreasing we have that $E_\xi \ M(p : \hat{Y}_2) \leq E_\xi \ M(p : \hat{Y}_1)$ for any fixed sample size (= n) plan (cf. Corollary 4.3.4 and Corollary 4.5.3). Table 4.7.6 and Table 4.7.7 illustrate this fact (since with $V(x) = x^g$ for $g > 2$, $V(x)/x^2$ is non-decreasing).

In the Appendix to this chapter we have reported the calculations concerning the estimator $\hat{Y}_{RHC}$ of (4.3.5) for computing $E_\xi(\hat{\beta}_{RHC} - \beta)^2/\sigma^2$ in the case of our population chosen. Note that the estimator $\hat{Y}_{RHC}$ is such that with regard to our population its coefficients do not satisfy the condition $a_i \leq a_j \Rightarrow x_i \geq x_j$ for $i, j \in s$, $s \in S$ with $p_s > 0$. But still, the condition

$$a_i \leq a_j \Rightarrow \frac{V(x_i)}{x_i^2} \leq \frac{V(x_j)}{x_j^2} \quad \text{for } i, j \in s, \ s \in S \ \text{ with } p_s > 0$$

($\Rightarrow n(s) = n$) being a sufficient condition for Theorem 4.3.1 to hold we have tabulated $E_\xi \ (\hat{\beta}_{RHC} - \beta)^2/\sigma^2$ for g ranging from 0 through 3.0 in our Table 4.7.1 through Table 4.7.7. Observe that for values of g up to 2.0 (i.e., $V(x)/x^2 = x^{g-2}$ nonincreasing) $\hat{Y}_{RHC}$ fares better than $\hat{Y}_2$. When $g = 2.0$, $\hat{Y}_2$ being the optimal estimator, it is better than $\hat{Y}_{RHC}$.

We illustrate the inequality (4.4.2), namely

$$E_\xi\ M(p^*_{\xi\text{-opt.}} : \hat{Y}_2) \leq E_\xi\ M(p : \hat{Y}_2)$$

Table 4.7.8. Observing that when $p \in P_n$,

$$M(p : \hat{Y}_2) = \sum_{s \in S_n} \{E_\xi(\hat{\beta}_2 - \beta)^2 (X - \sum_{i \in s} x_i)^2 + \sigma^2 \sum_{i \in \bar{s}} V(x_i)\} p_s ,$$

a purposive sampling plan $p^*_{\xi\text{-opt.}}$ which selects units 3 4 with certainty, since $V(x) = x^g$ is nondecreasing <u>for fixed</u>

$\sigma^2 \sum_{i \in \bar{s}} x_i^g$ is minimum for the sample $s^* = \{3,4\}$ such that

-opt.)s* = 1. Further $V(x)/x^2$ is nonincreasing for $V(x) = x^g$,

$\in [0,2]$. Table 4.7.8 below brings out the fact that the deli-
rate choice of the largest n units is ξ-optimal.

<u>Table 4.7.8</u>

Purposive sampling plan $p^*_{\xi\text{-opt.}}$ compared to other sampling plans.

$E_\xi(\hat{\beta}_2 - \beta)^2 (X - \sum_{i \in s} x_i)^2 / \sigma^2$ for $V(x) = x^g$

mple \ g	0	0.5	1.0	1.5	2.0
2}	15.3125	16.5816	18.3750	20.9132	24.5000
3}	10.0008	10.7316	11.9988	14.1948	18.0000
4}	6.6400	7.0325	7.8125	9.3750	12.5000
3}	2.2575	3.4125	5.2075	8.0275	12.5000
4}	1.2496	1.9136	3.0000	4.8288	8.0000
4}	.3906*	.7146*	1.3122*	2.4237*	4.5000*

Appendix to Chapter 4

Calculation of $E_\xi(\hat{\beta}_{RHC} - \beta)^2/\sigma^2$ with $V(x) = x^g$

The sampling procedure p_{RHC} consists of the following two stages :

(a) Split the population at random into n groups of sizes $N_1, N_2, \ldots, N_n$ where $N_1 + N_2 + \ldots + N_n = N$.

(b) Draw a sample of size one with probabilities proportional to $p_t (= x_t / \sum_{t=1}^{N} x_t$ since we are sampling with probability proportional to size) from each of these n groups independently.
The exact variance of the estimator $\hat{Y}_{RHC} = \sum_{i=1}^{n} \frac{y_i}{(p_i / \sum_{\text{Group } i} p_t)}$

is minimized by choosing $n = N/k$ when N is an integral multiple of n, namely $N = nk$. Therefore, we split our population (Yates and Grundy (1953)) with $N = 4$ into 2 groups in step (a) above.

Considering the overall p-expectation of $\hat{Y}_{RHC}$ when the plan p_{RHC} is used it is easy to note that

$$E_1 E_2 (\hat{Y}_{RHC}) = \sum_s (\hat{Y}^{RHC}) p_s \qquad (*)$$

where E_2 denotes the p-expectation of $\hat{Y}_{RHC}$ over all possible samples for a given split, E_1 denotes expectation over all splits of size n. In the right hand side of (*) we write the over all

$E_{p_{RHC}}$-expectation of $\hat{Y}_{RHC}$ as summation over all possible six samples $s = \{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}$ and $\hat{Y}^{RHC}$ is the statistic defined for all these six possible samples and p_s is the probability of obtaining these six samples using p_{RHC} procedure. Observe that the right hand side of (*) gives us the natural way to transform the theory (required for $\hat{Y}_{RHC}$) to ξ-unbiasedness set up. Thus, using the right hand side of (*) we write $\hat{Y}_{RHC}$ in the form $\sum_{i\varepsilon s} \beta_{si}\, y_i$ $(= \hat{Y}^{RHC})$ and then evaluate $E_\xi(\hat{\beta}_{RHC} - \beta)^2/\sigma^2$ for $V(x) = x^g$. Observe that $\sum_{i\varepsilon s} \beta_{si}\, x_i = X$ for $s \varepsilon S$ with $p_s > 0$ which implies that $\hat{Y}_{RHC}$ is ξ-unbiased.

Samples	The statistic $\hat{Y}^{RHC}$ obtained using the right side of (*)
{1,2}	$(4.4898)y_1 + (2.7551)\ y_2$
{1,3}	$(3.9130)y_1 + (2.0290)\ y_3$
{1,4}	$(3.4667)y_1 + (1.6333)\ y_4$
{2,3}	$(2.2000)y_2 + (1.8667)\ y_3$
{2,4}	$(1.9565)y_2 + (1.5217)\ y_4$
{3,4}	$(1.4966)y_3 + (1.3776)\ y_4$

Samples	Formula for computing $E_\xi(\hat{\beta}_{RHC}-\beta)^2/\sigma^2$ with $V(x)=x^g$
{1,2}	$\frac{1}{49}[(12.1787 \times 1^g) + (3.0804 \times 2^g)]$
{1,3}	$\frac{1}{36}[(8.4856 \times 1^g) + (1.0588 \times 3^g)]$
{1,4}	$\frac{1}{25}[(6.0846 \times 1^g) + (.4011 \times 4^g)]$
{2,3}	$\frac{1}{25}[(1.4400 \times 2^g) + (.7512 \times 3^g)]$
{2,4}	$\frac{1}{16}[(.9149 \times 2^g) + (.2722 \times 4^g)]$
{3,4}	$\frac{1}{9}[(.2466 \times 3^g) + (.1426 \times 4^g)]$

CHAPTER 5

GENERALIZED πPS DESIGNS

5.0 Summary

In this chapter, under the model $\theta(g)$ of (1.2.2), optimal strategy for estimating the population total Y in the sense of minimum expected variance is obtained in the class of <u>all</u> p-unbiased strategies with expected sample size fixed, for <u>all</u> non-negative values of g. An algebraic comparison of two strategies due to T.J. Rao (1972) is shown to be an immediate consequence of our theorem.

5.1 Introduction

Godambe (1955) showed that for any sampling design $d = (S, p)$ there does not exist a best estimator of the population total Y in the class of linear p-unbiased estimators. Hanurav (1965) pointed out some nontrivial exceptions to this where a best estimator exists. Such designs were termed 'unicluster designs' by Hanurav. Godambe's nonexistence theorem was later extended to the class of all p-unbiased estimators (Godambe and Joshi (1965)). However, we have already seen before that whenever auxiliary information on a characteristic $\mathcal{X}$ (positive valued) closely related to the characteristic $\mathcal{Y}$ under study, taking values x_i on the units $i = 1, 2, \ldots, N$, is available, this information can be used to set up a criterion of optimality for estimating Y as shown by Cochran (1946). Here $\underset{\sim}{Y} = (y_1, y_2, \ldots, y_N)$ is considered as a realization of an N-length random vector with distribution depending on $\underset{\sim}{X} = (x_1, x_2, \ldots, x_N)$ and some unknown parameters. Thus, we have a 'super-population model $\theta(g)$' formulated as

$$\left.\begin{aligned} E_{\theta(g)}(y_i \mid x_i) &= a\, x_i, \\ Var._{\theta(g)}(y_i \mid x_i) &= \sigma^2 x_i^g, \\ Cov._{\theta(g)}(y_i, y_j \mid x_i, x_j) &= 0, \quad 1 \le i \ne j \le N \end{aligned}\right\} \qquad (5.1.1)$$

where $E_{\theta(g)}$, $Var._{\theta(g)}$, $Cov._{\theta(g)}$ denote the conditional expectation, variance and covariance given x_i's respectively with respect to the prior distribution $\theta(g)$ of $\underset{\sim}{Y}$. The model (5.1.1) is the same as the model (1.2.2). As before y_i denotes both the random variable and the value it takes on unit i given a finite population U of size N. a and σ^2 are the unknown parameters of the prior distribution $\theta(g)$. Recall that g is non-negative in practice, and is observed to be between 1 and 2 more often.

If $n(s)$ denotes the number of units in a sample s and if the cost of drawing and inspecting a sample s is assumed to be proportional to $n(s)$, expected value of $n(s)$, $E(n(s)) = \sum_{s \in S} n(s) p_s$, for a given design $d = (S, p)$ is fixed would imply that expected cost of the design is fixed. If $H = (p : \hat{Y})$ is a strategy for estimating Y where p is a design with expected cost fixed, we say that the cost of the strategy H is fixed. Given a strategy $H = (p : \hat{Y})$ for estimating Y, the expected sample size of the strategy is defined to be the expected sample size of the corresponding design p. Considering the prior distribution $\theta(g)$ of $\underset{\sim}{Y}$ for $g \geq 0$, further specified by the model (5.1.1), a strategy H that has minimum expected (with respect to $\theta(g)$) sampling variance in the class of all p-unbiased strategies $D(H)$ of the population total Y is said to be $\theta(g)$-optimum in $D(H)$ for estimating Y. Analogously, $\theta(g)$-optimality of a strategy H for estimating Y in subclasses of $D(H)$ is defined.

Godambe (1955) proved that the strategy $H_o = (p_o : \hat{Y}_{HT})$ is $\theta(2)$-optimum in the class of linear, fixed sample size (= n), p-unbiased strategies of the population total Y where p_o is a πPS design with fixed sample size (= n). Here n is such that $\max_i \{n\ x_i\} \leq X$. Later, Godambe and Joshi (1965) proved that the above strategy $H_o = (p_o : \hat{Y}_{HT})$ is $\theta(2)$-optimum in the class of <u>all</u>, fixed sample size (= n), p-unbiased strategies of the population total Y. In practice, there are many sampling procedures that result in the required inclusion probabilities π_i to be proportional to x_i for $i = 1,2,\ldots,N$ (Goodman and Kish (1950), Horvitz and Thompson (1952), Yates and Grundy (1953), Durbin (1953), Grundy (1954), Des Raj (1956a), Hájek (1964), Hanurav (1962b), Rao, Hartley and Cochran (1962), Brewer and Undy (1962), Hartley and Rao (1962), Fellegi (1963), Brewer (1963a), Stuart (1963), J.N.K. Rao((1963), (1965)), Hanurav ((1965), (1967)), Sampford (1967), Vijayan (1968), Sankaranarayanan (1969), Dodds and Fryer (1971), Das and Mohanty (1973), Brewer (1975), Chaudhuri (1975) and others).

Hansen and Hurwitz (1943) demonstrated the profitability of selecting sampling units with probability proportional to size of the units and indicated methods of determining the probability of selection which minimize the variance of the estimator at a fixed cost. Hansen and Hurvitz (1949) also showed that sampling with probability proportional to the square root of size (x_i) is more efficient than sampling with probability proportional to size under certain conditions. Under the model (5.1.1), T.J. Rao (1971) studied the Horvitz-Thompson estimator of the population total Y with designs wherein the inclusion probability π_i is proportional to modified measure of size x_i^{α}, $i = 1, 2, \ldots, N$, for any positive real α.

<u>Definition 5.1.1 (T.J. Rao (1972))</u> : If a design $d = (S, p)$ with $\pi_i > 0$ for $i = 1, 2, \ldots, N$ is such that π_i is proportional to $x_i^{g/2}$, $g \geq 0$, $i = 1, 2, \ldots, N$, and $\sum_{i \in s} x_i^{1-(g/2)} = c$, a constant for every $s \in S$ with $p_s > 0$, say that the design p is a Generalized πPS design and denote it by " GπPS design ".

Observe that for a GπPS design to be available the expected sample size n is such that

$$\max_i \left\{ n\, x_i^{g/2} \right\} \leq \sum_{i=1}^{N} x_i^{g/2} .$$

When $g = 2$, a GπPS design is a πPS design with fixed sample size.

Since $\pi_i = n\, x_i^{g/2} / (\sum_{i=1}^{N} x_i^{g/2})$ for $i = 1,2,\ldots,N$ and $\sum_{i \epsilon s} x_i^{1-(g/2)} = c$, a constant $\forall$ s with $p_s > 0$ in the definition of a GπPS design with expected sample size n, it follows that the constant c is in fact $= n\, X/(\sum_{i=1}^{N} x_i^{g/2})$ where $X = \sum_{i=1}^{N} x_i$.

While estimating the population total Y, T.J. Rao (1971) observes that the strategy consisting of GπPS design and the corresponding HT-estimator is better than the strategy consisting of the πPS design with fixed sample size (= n) and the corresponding HT-estimator in the expected variance sense under $\theta(g)$ for all $g \epsilon [1,2]$. When the expected sample size is fixed to be 2, i.e. n = 2, comparing the strategy H_1-consisting of the GπPS design and the corresponding HT-estimator, with the fixed sample size (= 2) strategy H_2 — consisting of the p_{SDR} design and the Symmetrized Des Raj estimator, T.J. Rao (1972) proves algebraically that the strategy H_1 is better than H_2 in the expected variance sense under $\theta(g)$ for all $g \epsilon [1,2]$. In section 5.2 we establish the $\theta(g)$-optimality of the strategy consisting of the GπPS design and the corresponding HT estimator in the class of all p-unbiased strategies with expected sample size fixed (= n) for all $g \geq 0$, for estimating Y, the population total.

5.2 An Existence Theorem

Given a design $d = (S, p)$ consider the most general type of linear estimator of $Y = \sum_{i=1}^{N} y_i$ defined as

$$\hat{Y} = \sum_{i \varepsilon s} \beta_{si}\, y_i \tag{5.2.1}$$

for $s \varepsilon S$ with $p_s > 0$, where β_{si} is defined in advance for all the logically possible s and for all units $i \varepsilon s$. In a given design $d = (S, p)$ in which the inclusion probabilities π_i's are all non-zero, a necessary and sufficient condition for $\hat{Y}$ to be p-unbiased for Y is

$$\sum_{s \ni i} \beta_{si}\, p_s = 1 \tag{5.2.2}$$

for $i = 1, 2, \ldots, N$; $\sum_{s \ni i}$ standing for summation over all samples $s \varepsilon S$ which contain unit i.

Now, $\quad \text{Var.}(p : \hat{Y}) = \sum_{s \varepsilon S} \hat{Y}^2 p_s - Y^2.$

Hence following Godambe (1955) we have

$$E_{\theta(g)} \text{Var.}(p : \hat{Y}) = \Big(\sum_{s \varepsilon S} (E_{\theta(g)} (\hat{Y}^2))\, p_s \Big) - E_{\theta(g)}(Y^2)$$

$$= \Big[\sum_{s \varepsilon S} \{ E^2_{\theta(g)}(\hat{Y}) + \text{Var.}_{\theta(g)}(\hat{Y}) \}\, p_s \Big] - E_{\theta(g)}(Y^2)$$

$$= \sum_{s \varepsilon S} (E^2_{\theta(g)}(\hat{Y}))\, p_s + \sigma^2 \sum_{i=1}^{N} x_i^g \Big(\sum_{s \ni i} \beta^2_{si} p_s \Big) - E_{\theta(g)}(Y^2) \tag{5.2.3}$$

under our model (5.1.1).

Now subject to the condition

$$\sum_{s\in S} \hat{Y}\ p_s = Y$$

$$\sum_{s\in S} (E^2_{\theta(g)}(\hat{Y}))\ p_s \geq E^2_{\theta(g)}(Y) \qquad (5.2.4.)$$

and subject to the condition

$$\sum_{s\ni i} \beta_{si}\ p_s = 1$$

$$(\sum_{s\ni i} \beta^2_{si}\ p_s) \geq 1/(\sum_{s\ni i} p_s) = 1/\pi_i \qquad (5.2.5)$$

for $i = 1,2,\ldots,N$. From (5.2.3), (5.2.4) and (5.2.5) it follows that for any sampling design p in which the inclusion probabilities π_i's are all non-zero

$$E_{\theta(g)}\ \mathrm{Var.}(p : \hat{Y}) \geq \sigma^2 [\sum_{i=1}^{N} \frac{x_i^g}{\pi_i} - \sum_{i=1}^{N} x_i^g]$$

$$= \sigma^2 \sum_{i=1}^{N} (\frac{1}{\pi_i} - 1)\ x_i^g \qquad (5.2.6)$$

under our model (5.1.1) where π_i is the probability of including the ith unit ($i = 1,2,\ldots,N$) in the sample drawn with the sampling design p. Observe that, in particular, the best linear unbiased estimator (the one with minimum expected variance) with respect to the design p also satisfies (5.2.6).

Remark 5.2.1: Inequality (5.2.6) is true when $Var._{\theta(g)}(y_i \mid x_i) = \sigma_i^2$, $i = 1,2,\ldots,N$ (Godambe (1955)) in our model (5.1.1).

Remark 5.2.2: Inequality (5.2.6) is also true for any p-unbiased estimator $\hat{Y}$ of the population total. In other words, one need not restrict attention to linear estimators alone (Godambe and Joshi (1965)).

Now we consider the class of all p-unbiased strategies with expected sample size fixed (= n say). Minimising the right hand side of (5.2.6) subject to the condition $\sum_{i=1}^{N} \pi_i = n$ we have that π_i should be proportional to $x_i^{g/2}$ for $i = 1,2,\ldots,N$ in order that the minimum be attained. Now consider the strategy F consisting of the $G\pi PS$ design together with the corresponding Horvitz-Thompson estimator of the population total Y.

$$Var.(G\pi PS : \hat{Y}_{HT}) = \sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 1\right) y_i^2 + \sum_{\substack{i \neq j \\ 1}}^{N} \sum \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1\right) y_i y_j.$$

$$E_{\theta(g)} Var.(G\pi PS : \hat{Y}_{HT}) = \sigma^2 \sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 1\right) x_i^g + a^2 \, Var.\left(\sum_{i \in s} \frac{x_i}{\pi_i}\right)$$

where $\sum_{i \in s} \frac{x_i}{\pi_i}$ is the HT-estimator obtained by replacing the y_i's by the x_i's for $i \varepsilon s$, $s \varepsilon S$ with $p_s > 0$.

$$= \sigma^2 \sum_{i=1}^{N} \left(\frac{\sum_{i=1}^{N} x_i^{g/2}}{n\, x_i^{g/2}} - 1 \right) x_i^g + a^2 \text{Var.} \left[\frac{\left(\sum_{i \varepsilon s} x_i^{1-(g/2)} \right) \left(\sum_{i=1}^{N} x_i^{g/2} \right)}{n} \right] \qquad (5.2.7)$$

$$= \sigma^2 \sum_{i=1}^{N} \left(\frac{\sum_{i=1}^{N} x_i^{g/2}}{n\, x_i^{g/2}} - 1 \right) x_i^g \quad \text{since}$$

$$\sum_{i \varepsilon s} x_i^{1-(g/2)} = n\, X / \sum_{i=1}^{N} x_i^{g/2}$$

for every s with $p_s > 0$.

Observe that the lower bound of the inequality (5.2.6) is in fact attained by the strategy $F = (G\pi PS : \hat{Y}_{HT})$. Since the inclusion probabilities π_i's with respect to the $G\pi PS$ design are proportional to $x_i^{g/2}$, we have therefore established the following theorem :

<u>Theorem 5.2.3</u> : For estimating the population total Y, in the class of all p-unbiased strategies with expected sample size fixed (= n), a $\theta(g)$-optimal strategy exists for $g \geq 0$, whatever be a and σ^2 in our model (5.1.1), and that it is given by the strategy consisting of the $G\pi PS$ design together with the corresponding HT-estimator.

<u>Remark 5.2.4</u> : The $\theta(2)$-optimal strategy for estimating Y derived by Godambe and Joshi (1965) is thus given by the strategy H_0 consisting of the πPS design with fixed sample size (= n) and the corresponding HT-estimator. The strategy H_0 is therefore $\theta(2)$-optimum in the class of all p-unbiased strategies of Y with expected sample size fixed (= n) by our Theorem 5.2.3 above. Observe that for any design p with $\pi_i > 0$ for $i = 1, 2, \ldots, N$, $E(n(s)) = \sum_{s \in S} n(s)\, p_s = \sum_{i=1}^{N} \pi_i = n$. Thus, our $\theta(g)$-optimal strategy of Theorem 5.2.3 has been obtained by considering all equicost p-unbiased strategies of Y which is as it should be for realistic comparisons.

<u>Remark 5.2.5</u> : We have shown that the strategy $F = (G\pi PS : \hat{Y}_{HT})$ is $\theta(g)$-optimum, $g \geq 0$, for estimating Y in the class of all p-unbiased strategies of Y with expected sample size fixed(=n). A natural question at this stage is to ask if F is unique. As is clearly seen there may be one or more designs that may result in the required inclusion probabilities (in the definition of a $G\pi$PS design). $G\pi$PS of course refers to any one of these. So, if uniqueness of the strategy F is understood in the above sense, it is natural to discuss the uniqueness and we study this problem in chapter 6.

Remark 5.2.6 : When the expected sample size is fixed to be 2, while estimating Y, T.J. Rao has shown that the strategy consisting of a design with π_i proportional to $x_i^{g/2}$ (for $i = 1,2,\ldots,N$) and satisfying $\sum_{i \varepsilon s} x_i^{1-(g/2)} = c$, a constant for all s with $p_s > 0$ (observe that c is in fact $= n X / \sum_{i=1}^{N} x_i^{g/2}$ where n is the expected sample size of the design; here n is 2) together with the corresponding estimator $\sum_{i \varepsilon s} (y_i/\pi_i)$ is superior to the strategy $(p_{SDR} : \hat{Y}_{SDR})$ with fixed sample size (= 2) under $\theta(g)$ for all values of $g \varepsilon [1,2]$ (T.J. Rao (1972)). It was also shown in T.J. Rao (1971) that for estimating Y, the strategy (GπPS : $\hat{Y}_{HT}$) with expected sample size n is better than the strategy consisting of a πPS design with fixed sample size n and the corresponding HT-estimator under $\theta(g)$ for all $g \varepsilon [1,2]$. These are immediate consequences of our Theorem 5.2.3.

Remark 5.2.7 : That the class of GπPS designs is non-empty was illustrated by an example by T.J. Rao (1971). We give below yet another example of a GπPS design for $g = 1.5$.

Let the population consist of 5 units with the auxiliary information being x_i, $i = 1,2,3,4,5$.

unit	x_i	$x_i^{g/2}$	$x_i^{1-(g/2)}$
1	1	1	1
2	1	1	1
3	16	8	2
4	1	1	1
5	16	8	2

Let n be the expected cost of the design fixed beforehand $= (4 \times 19)/35$.

We then have $\sum_{i \varepsilon s} x_i^{1-(g/2)} = n\,X/(\sum_{i=1}^{N} x_i^{g/2}) = 4$.

We now need to construct a design for which $\sum_{i \varepsilon s} x_i^{1-(g/2)} = 4 \;\forall\; s$ with $p_s > 0$ and $\pi_i \propto x_i^{g/2}$ for $i = 1,2,3,4,5$. Consider the following design

Sample(s) (units)	Probabilities p_s
{2,3,4}	1/35
{2,4,5}	1/35
{3,5}	29/35
{1,2,3}	2/35
{1,4,5}	2/35

wherein we have $\sum_{i\varepsilon s} x_i^{1-(g/2)} = 4$ for all s as required. Further, $\pi_1 = \pi_2 = \pi_4 = \frac{4}{35}$; $\pi_3 = \pi_5 = \frac{32}{35}$ which are proportional to $x_i^{g/2}$ and the design is such that $\pi_{ij} > 0$ for all i and j for the estimability of the variance.

Remark 5.2.8 : T.J. Rao (1971) studied the Horvitz-Thompson estimator with designs wherein the π_i's are proportional to modified size measure. Following Hanurav (1962a), T.J. Rao proved the following theorem :

Theorem 5.2.9 (T.J. Rao (1971)) : Let D be the class of designs with expected sample size n wherein $\pi_i \propto x_i^{g/2}$ for $i = 1,2,\ldots,N$, in conjunction with which the Horvitz-Thompson estimator $\hat{Y}_{HT}$ is used for the estimation of the population total Y. In the class D, the $\theta(g)$-optimum designs for $g \varepsilon [1,2]$ which are best suited for the Horvitz-Thompson estimator are those that satisfy $\sum_{i\varepsilon s} x_i^{1-(g/2)}$ = a constant $\forall$ s with $p_s > 0$.

Remark 5.2.10 : When the auxiliary information $\mathcal{X}$ is constant on all the units in the population,

$\pi_i \propto x_i^{g/2} \Rightarrow \pi_i = n/N$ for $i = 1,2,\ldots,N$ and

$$\sum_{i\varepsilon s} x_i^{1-(g/2)} = n\,X/(\sum_{i=1}^{N} x_i^{g/2}) \Rightarrow n(s) = n \;\forall\; s \text{ with } p_s > 0,$$

nd $\hat{Y}_{HT}$ when used with a GπPS design when the x_i's are all constants therefore reduces to $N(\sum_{i \in s} y_i)/n$. Hence the (g)-optimality for $g \geq 0$ of the classical simple expansion estimator $N(\sum_{i \in s} y_i)/n$ (with $n < N$) of the population total Y coupled with a simple random sampling without replacement plan in which

$$p_s = \begin{cases} \frac{1}{\binom{N}{n}} & \text{when } n(s) = n \\ 0 & \text{otherwise} \end{cases}$$

in the class of <u>all</u> p-unbiased strategies of Y with expected sample size n follows.

CHAPTER 6

OPTIMAL EXPECTED VARIANCE AND GENERALIZED πPS DESIGNS

6.0 <u>Summary</u>

In chapter 5, for estimating the population total Y, the strategy $(G\pi PS : \hat{Y}_{HT})$ was shown to be $\theta(g)$-optimal, $g \geq 0$, in the class of <u>all</u> p-unbiased strategies of Y with expected sample size fixed ($= n$). We discuss in this chapter p-unbiased strategies of Y that attain the lower bound of the inequality (5.2.6). We completely characterize the strategies that are $\theta(g)$-optimal, $g \geq 0$, for estimating Y in the class of all p-unbiased strategies of Y with expected sample size fixed ($=n$). As a special case, the "uniqueness" of the $\theta(2)$-optimal strategy for estimating Y is obtained. We also show that Cassel <u>et al</u>.'s (1976) optimality theorem of the generalized difference estimator under $\theta(g)$ for $g \in [0, 2]$ is an immediate corollary of our characterization theorem. In section 6.2 we remark on the robustness of our $\theta(g)$-optimal strategy $(G\pi PS : \hat{Y}_{HT})$ to the possible errors of complete characterization of the underlying prior distribution $\theta(g)$ while discussing the "Generalized Regression Estimator" for estimating Y.

6.1 Optimal Expected Variance

Working with $\theta(g)$, the prior distribution of $\underset{\sim}{Y}$ further specified by the model (5.1.1) we observed in the last chapter that for any sampling design p in which the inclusion probabilities π_i's are all non-zero, and any estimator $\hat{Y}$ that is p-unbiased for Y,

$$E_{\theta(g)} \text{Var.}(p : \hat{Y}) \geq \sigma^2 \sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 1 \right) x_i^g \qquad (6.1.1)$$

(cf. Remark 5.2.2).

Definition 6.1.1 : The right hand side of (6.1.1) is defined to be the 'optimal expected variance' for p-unbiased estimators of Y under the model $\theta(g)$ where p is a design with non-zero π_i's.

Considering our model $\theta(g)$ of (5.1.1), the expected variance of the HT-estimator of the population total Y with respect to any design p with non-zero π_i's is

$$E_{\theta(g)} \text{Var.}(p : \hat{Y}_{HT}) = \sigma^2 \sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 1 \right) x_i^g + a^2 \text{Var.}\left(\sum_{i \in s} \frac{x_i}{\pi_i} \right)$$

where $\sum_{i \in s} \frac{x_i}{\pi_i}$ is the HT-estimator obtained by replacing the y_i's by x_i's for $i \in s$ $\forall$ s with $p_s > 0$.

Therefore, if the design p is such that $\sum_{i\varepsilon s} \frac{x_i}{\pi_i} = c$, a constant $\forall$ s with $p_s > 0$, we have the $\hat{Y}_{HT}$ attains the optimal expected variance under the model $\theta(g)$ of (5.1.1). Observe that if $\sum_{i\varepsilon s} \frac{x_i}{\pi_i} = c$, a constant $\forall$ s with $p_s > 0$ then c is in fact $= \sum_{i=1}^{N} x_i = X$. Hence we have that if $\hat{Y}_{HT}$ is " $\theta(g)$-unbiased" for Y then $\hat{Y}_{HT}$ attains the optimal expected variance under $\theta(g)$ when used with any design p with non-zero π_i's.

Considering p-unbiased strategies of Y with expected sample size fixed (= n), we have that the right hand side of (6.1.1) is minimised for designs such that π_i is proportional to $x_i^{g/2}$, i = 1,2,...,N. Therefore the strategy (GπPS : $\hat{Y}_{HT}$) attains the minimum optimal expected variance under $\theta(g)$ in the class of all p-unbiased strategies of Y with expected sample size fixed (= n, integral or non-integral). We illustrate this by an example below.

Example 6.1.2 :

Let g = 1.5. Consider a population consisting of four units with auxiliary information x_i, i = 1,2,3,4.

unit	x_i	$x_i^{g/2}$	$x_i^{1-(g/2)}$
1	1	1	1
2	16	8	2
3	1	1	1
4	16	8	2
Total	34	18	6

Let the expected cost of the design be fixed beforehand $= \frac{36}{17}$

Consider the following $G\pi PS$ design (T.J. Rao (1971))

sample s (units)	probabilities p_s
$\{1,2,3\}$	1/17
$\{1,3,4\}$	1/17
$\{2,4\}$	15/17

Observe that the π_{ij}'s are all > 0 for the example above so that the variance of the estimator is estimable.

Now, under $\theta(g)$

$$E_{\theta(g)} \text{Var.}(G\pi PS : \hat{Y}_{HT}) = \sigma^2 \sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 1\right) x_i^g$$

$$= 23\,\sigma^2 .$$

Consider now the following πPS design.

sample s (units)	probabilities p_s
{1,3}	$1/17^2$
{2,4}	$254/17^2$
{1,2,4}	$17/17^2$
{2,3,4}	$17/17^2$

Now

$$E_{\theta(g)} \text{ Var.}(p : \hat{Y}_{HT}) = \sigma^2 \sum_{i=1}^{N} (\frac{1}{\pi_i} - 1)x_i^g + a^2 \text{ Var.}(p : \hat{X}_{HT})$$

$$\geq \sigma^2 \sum_{i=1}^{N} (\frac{1}{\pi_i} - 1) x_i^g$$

Hence when $\hat{Y}_{HT}$ is used as an estimator of Y with the πPS design (with expected sample size $\frac{36}{17}$) given above we have

$$E_{\theta(g)} \text{ Var.}(p : \hat{Y}_{HT}) \geq \sigma^2 \frac{275}{9} > 23\,\sigma^2 .$$

One can in fact consider with respect to any design p in which the inclusion probabilities π_i's are all non-zero, the p-unbiased estimator $\hat{Y}^{**}$ of the population total Y given by

$$\hat{Y}^{**} = \hat{Y}_{HT} + c(X - \hat{X}_{HT}) \qquad (6.1.2)$$

where c is a constant and $\hat{X}_{HT}$ is the estimator $\sum_{i\epsilon s} \frac{x_i}{\pi_i} \; \forall \; s$ with $p_s > 0$ obtained by replacing the y_i's by x_i's for $i \epsilon s$ in $\hat{Y}_{HT}$. Now, under $\theta(g)$ we have

$$E_{\theta(g)} \text{Var.}(p : \hat{Y}^{**}) = E_{\theta(g)} \text{Var.}(p : \hat{Y}_{HT}) + c^2 E_{\theta(g)} \text{Var.}(p : X - \hat{X}_{HT})$$

$$+ 2c \, E_{\theta(g)} \text{Cov.}(p : (\hat{Y}_{HT}, X - \hat{X}_{HT})) \tag{6.1.3}$$

Observe that $\text{Var.}(p : X - \hat{X}_{HT}) = \text{Var.}(p : \hat{X}_{HT})$

and $E_{\theta(g)} \text{Var.}(p : \hat{X}_{HT}) = \text{Var.}(p : \hat{X}_{HT})$.

Now,

$$\text{Cov.}(p : (\hat{Y}_{HT}, X - \hat{X}_{HT})) = \sum_{s\epsilon S} (\hat{Y}_{HT} - Y)(X - \hat{X}_{HT}) \, p_s$$

$$= -\sum_{s\epsilon S} (\hat{Y}_{HT} \hat{X}_{HT} - YX) \, p_s$$

$$= -[E(p : \hat{Y}_{HT} \hat{X}_{HT}) - E(p : \hat{Y}_{HT}) \, E(p : \hat{X}_{HT})]$$

$$= -\left\{ \left[\sum_{i=1}^{N} \frac{x_i \, y_i}{\pi_i} + \sum_{\substack{i \neq j \\ 1}}^{N} \sum \frac{x_i}{\pi_i} \frac{y_j}{\pi_j} \, \pi_{ij} \right] - YX \right\}.$$

Hence

$$E_{\theta(g)} \text{ Cov.}(p : (\hat{Y}_{HT}, X - \hat{X}_{HT}))$$

$$= -a \left[\left(\sum_{i=1}^{N} \frac{x_i^2}{\pi_i} + \sum_{\substack{i \neq j \\ 1}}^{N} \sum \frac{x_i \; x_j}{\pi_i \; \pi_j} \pi_{ij} \right) - X^2 \right]$$

$$= - a \text{ Var. } (p : \hat{X}_{HT}).$$

Substituting in (6.1.3) we have

$$E_{\theta(g)} \text{ Var.}(p : \hat{Y}^{**}) = E_{\theta(g)} \text{ Var.}(p : \hat{Y}_{HT}) + c^2 \text{ Var.}(p : \hat{X}_{HT})$$

$$- 2 \; c \; a \text{ Var.}(p : \hat{X}_{HT})$$

$$= \sigma^2 \sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 1\right) x_i^g + a^2 \text{ Var.}(p : \hat{X}_{HT})$$

$$+ c^2 \text{ Var.}(p : \hat{X}_{HT}) - 2 \; c \; a \text{ Var.}(p : \hat{X}_{HT})$$

$$= \sigma^2 \sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 1\right) x_i^g + (a - c)^2 \text{ Var.}(p : \hat{X}_{HT}) \qquad (6.1.4)$$

Note that the 2nd term in the right hand side of (6.1.4) vanishes when <u>either</u> (i) $\hat{X}_{HT}$ is a constant for all s with $p_s > 0$ in which case $\hat{X}_{HT} = X \; \forall \; s$ with $p_s > 0$ <u>or</u> (ii) $c = a$, the parameter of the prior distribution $\theta(g)$ of $\underset{\sim}{Y}$. Therefore, considering our inequality (6.1.1) since in the class of all p-unbiased strategies of Y with expected sample size fixed (= n), we have that the right hand side of the inequality (6.1.1) is minimised for designs with $\pi_i \propto x_i^{g/2}$, $i = 1,2,\ldots,N$, from (6.1.4) we have the following two families of strategies that are $\theta(g)$-optimal, $g \geq 0$, for estimating Y in the class of all p-unbiased strategies of Y with expected sample size fixed(= n).

(i) With a sampling design p such that $\hat{X}_{HT}$ = a constant $\forall$ s with $p_s > 0$ and $\pi_i \propto x_i^{g/2}$ for $i = 1,2,\ldots,N$ the strategy $(p' ; \hat{Y}_{HT})$ is $\theta(g)$-optimum for estimating Y, $g \geq 0$, whatever be a and σ^2, in the class of all p-unbiased strategies of Y with expected sample size fixed (= n).

(ii) The estimator $\hat{Y}^{**} = \hat{Y}_{HT} + a(X - \hat{X}_{HT})$ with a design p in which the inclusion probability π_i is proportional to $x_i^{g/2}$ for $i = 1,2,\ldots,N$ is $\theta(g)$-optimum for estimating Y, $g \geq 0$, a known, whatever be σ^2 in the model $\theta(g)$, in the class of all p-unbiased strategies of Y with expected sample size fixed (= n).

The strategy (i) above is the (GπPS ; $\hat{Y}_{HT}$) strategy that is $\theta(g)$-optimum for estimating Y, $g \geq 0$, in the class of all p-unbiased strategies of Y with expected sample size fixed (= n) derived by us in chapter 5. The strategy derived in (ii) above is such that for integral n, the fixed expected sample size, one can in fact consider (without loss of generality) fixed sample size (= n) designs with $\pi_i \propto x_i^{g/2}$ for $i = 1,2,\ldots,N$. Hence the optimal designs with $\pi_i \propto x_i^{g/2}$ when the expected sample size n is integral, can be constructed by considering fixed sample size designs. Any πPS scheme with fixed sample size (= n) can be used to arrive at $\pi_i \propto z_i$ for $i = 1,2,\ldots,N$ where z_i is the modified size measure $x_i^{g/2}$. Observe that the expected sample size n is such that

$$\max_i \left\{ \frac{n\, x_i^{g/2}}{\sum_{i=1}^{N} x_i^{g/2}} \right\} \leq 1 \quad \text{for a design with } \pi_i \propto x_i^{g/2},$$

$i = 1,2,\ldots,N$ to be available.

Edmond Ho (see footnote below) has proved that " With a sampling design p with non-zero π_i's any p-unbiased estimator $\hat{Y}^*$ of the population total Y attains the optimal expected variance under the model ξ if, and only if, it is of the form

$$\hat{Y}^* = \hat{Y}_{HT} + [\, E_\xi (Y) - E_\xi (\hat{Y}_{HT})] \text{ "}$$

Edmond Ho (Private Communication (1977)) : Reference made to Ho, E.W.H. (1976), "Superpopulation Models in Finite Population Sampling", M.Sc. thesis, pp. 72-74, Australian National University, a copy of which the present author has not seen.

working with a superpopulation model ξ such that

$$\left.\begin{array}{ll} \text{(i)} & y_1, y_2, \ldots, y_N \text{ are independent} \\ \text{(ii)} & E_\xi(y_i) = \mu_i \\ \text{(iii)} & Var._\xi(y_i) = \sigma_i^2 \end{array}\right\} \quad \forall\ i = 1,2,\ldots,N \qquad (6.1.5)$$

As a corollary to this result Ho proves

<u>Theorem 6.1.3 (Ho (Private Communication (1977)))</u>. Under the superpopulation model ξ of (6.1.5), $\hat{Y}_{HT}$ attains the optimal expected variance if, and only if, it is ξ-unbiased for Y.

While Ho is interested in p-unbiased estimators of the population total Y that attain the optimal expected variance, we are looking at p-unbiased strategies of Y with expected sample size fixed (= n) that attain the minimum optimal expected variance (see example 6.1.2).

Observe that under $\theta(g)$

$$\hat{Y}^* = \hat{Y}_{HT} + [E_{\theta(g)}(Y) - E_{\theta(g)}(\hat{Y}_{HT})]$$

$$= \hat{Y}_{HT} + a(X - \hat{X}_{HT})$$

and

$$E_{\theta(g)} Var.(p : \hat{Y}^*) = \sigma^2 \sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 1\right) x_i^g . \qquad (6.1.6)$$

Thus, $\hat{Y}^* = \hat{Y}_{HT} + a(X - \hat{X}_{HT})$ is the unique estimator that attains t optimal expected variance under $\theta(g)$, when a is known, whatever be σ^2 when used with any design p with non-zero π_i's. Now, minimising the right hand side of (6.1.6) subject to $\sum_{i=1}^{N} \pi_i = n$ we have the following theorems.

Theorem 6.1.4 : When n, integral valued, is such that

$$\max_i \left\{ \frac{n\, x_i^{g/2}}{\sum_{i=1}^{N} x_i^{g/2}} \right\} \leq 1,$$

for estimating Y, when a is known, whatever be σ^2, the unique $\theta(g)$-optimal strategy for $g \geq 0$ in the class of all p-unbiased strategies of Y with expected sample size fixed (= n) is given by the strategy consisting of a design p* with fixed sample size (= n) in which $\pi_i \propto x_i^{g/2}$ for $i = 1,2,\ldots,N$ and the corresponding estimator

$$\hat{Y}^* = \hat{Y}_{HT} + a(X - \hat{X}_{HT}).$$

Theorem 6.1.5 : When n (integral or non-integral) is such that

$$\max_i \left\{ \frac{n\, x_i^{g/2}}{\sum_{i=1}^{N} x_i^{g/2}} \right\} \leq 1,$$

for estimating Y, the unique $\theta(g)$-optimal strategy for $g \geq 0$ in the class of all p-unbiased strategies of Y with expected sample size fixed (= n) is given by (GπPS : $\hat{Y}_{HT}$), whatever be a and σ^2.

Remark 6.1.6 : When $g = 2$, the design p^* in Theorem 6.1.4 is the πPS design with fixed sample size (= n). Hence $\hat{Y}_{HT}$ corresponding to this p^* is " $\theta(g)$-unbiased" for Y whatever be a . Therefore, for estimating Y, the strategy consisting of a πPS design with fixed sample size (= n) with the corresponding HT-estimator is the unique $\theta(2)$-optimal strategy, whatever be a and σ^2, in the entire class of p-unbiased strategies of Y with expected sample size fixed (= n, integral). The $\theta(2)$-optimal strategy is " unique" in this sense (cf. Godambe (1955)).

Considering a prior distribution η of $\underset{\sim}{Y}$ further specified by the model η such that

$$\left.\begin{aligned} E_\eta (y_i) &= \mu_i = \mu\ a_i + b_i, \quad \text{for } i = 1,2,\ldots,N \\ E_\eta (y_i - \mu_i)^2 &= \sigma_i^2 = \sigma^2\ a_i^2 , \qquad i = 1,2,\ldots,N \\ E_\eta \{(y_i - \mu_i)\ (y_j - \mu_j)\} &= \sigma_{ij} = \sigma^2\ \rho\ a_i\ a_j \\ & \qquad\qquad 1 \le i \ne j \le N \end{aligned}\right\} \qquad (6.1.7)$$

where $a_i > 0$, b_i, $i = 1,2,\ldots,N$ are known numbers,

$$\sum_{i=1}^{N} a_i = N, \qquad (6.1.8)$$

and μ, σ^2 and ρ are unknown and $-1/(N-1) \le \rho \le 1$, Cassel et al. (1976) prove the following theorem :

Theorem 6.1.7 (Cassel et al.(1976)) : The strategy $(p_o : \hat{\bar{Y}}_{GDO})$ such that

$$\left.\begin{array}{ll} \text{(i)} & p_o \text{ is a fixed sample size } (= n) \text{ design with} \\ & \text{the inclusion probabilities of } p_o \text{ satisfying} \\ & \pi_i = f\, a_i \text{ for } i = 1,2,\ldots,N \text{ where } f = n/N, \\ \text{(ii)} & \hat{\bar{Y}}_{GDO} = \sum_{i \in s} ((y_i - b_i)/n\, a_i) + \sum_{i=1}^{N} (b_i/N) \end{array}\right\} \quad (6.1.9)$$

is η-optimum for estimating the population mean $\bar{Y} = Y/N$ in the class of linear p-unbiased estimators of $\bar{Y}$ with a fixed sample size (= n) design p such that $\pi_i > 0$ for $i = 1,2,\ldots,N$.

Let us now look at the usual regression model ξ specified as follows :

$$\left.\begin{array}{l} E_\xi\,(y_i) = \mu_i = \beta\, x_i \\ E_\xi\,(y_i - \mu_i)^2 = \sigma_i^2 = \sigma^2\, v(x_i) \\ E_\xi\,\{(y_i - \mu_i)\,(y_j - \mu_j)\} = 0, \quad i \neq j \end{array}\right\} \quad (6.1.10)$$

where $V(x_i) = x_i^{2\upsilon} / m_\upsilon^2$; $m_\upsilon = \sum_{i=1}^{N} x_i^\upsilon / N$, υ is known, $0 \leq \upsilon \leq 1$, σ^2 unknown ; <u>when β is known</u> (6.1.10) is the special case of model (6.1.7) such that $\mu = \rho = 0$, $a_i^2 = V(x_i)$, $b_i = \beta\, x_i$. Under the model (6.1.10) as a consequence of Theorem 6.1.7 Cassel *et al.* prove

<u>Theorem 6.1.8 (Cassel et al. (1976))</u> : Provided $n\, x_i^\upsilon < N\, m_\upsilon$ for all i, the ξ-optimum strategy for estimating $\bar{Y}$ is (PPS x^υ : $\hat{\bar{Y}}_{GD\upsilon}$) in the class of <u>linear</u> p-unbiased, fixed sample size (= n), strategies of $\bar{Y}$ where

$$\hat{\bar{Y}}_{GD\upsilon} = m_\upsilon\; R_s + \beta\,(m - m_\upsilon\, q_s),$$

the generalized difference estimator wherein $R_s = \sum_{i\varepsilon s} (y_i/(n\, x_i^\upsilon))$; $q_s = \sum_{i\varepsilon s} (q_i/n)$; $q_i = x_i^{1-\upsilon}$, and writing for simplicity m in place of m_1 ; and PPS x^υ denotes a design with fixed sample size (= n), in which $\pi_i \propto x_i^\upsilon$ for $i = 1,2,\ldots,N$.

We note that Theorem 6.1.8 is a special case of our Theorem 6.1.4 which establishes the $\theta(g)$-optimality for $g \geq 0$, when a is known in the model $\theta(g)$, whatever be σ^2, of the strategy $(p^* : \hat{Y}^*)$ where p^* is a fixed sample size (= n) design with $\pi_i \propto x_i^{g/2}$ for $i = 1,2,\ldots,N$ and

$\hat{Y}^* = \hat{Y}_{HT} + a(X - \hat{X}_{HT})$ is the associated estimator of Y, in the class of all p-unbiased strategies of Y with expected sample size fixed (= n).

Remark 6.1.9 : It seems natural to conjecture here that the strategy $(p_o : \hat{\bar{Y}}_{GDO})$ of Theorem 6.1.7 is η-optimum for estimating $\bar{Y}$ in the class of all p-unbiased strategies of $\bar{Y}$ with expected sample size fixed (= n).

6.2 Generalized Regression Estimator

In section 6.1, for estimating Y, we derived the $\theta(g)$-optimality of the strategy $(p^* : \hat{Y}^*)$ when a is known, whatever be σ^2 in the model $\theta(g)$, $g \geq 0$, in the class of all p-unbiased strategies of Y with expected sample size fixed to be n such that

$\max_i \left\{ \frac{n\, x_i^{g/2}}{\sum_{i=1}^{N} x_i^{g/2}} \right\} \leq 1.$ We also noted the $\theta(g)$-optimality of the strategy $(G\pi PS : \hat{Y}_{HT})$ in the class of all p-unbiased strategies of Y with expected sample size fixed (= n such that $\max_i \left\{ \frac{n\, x_i^{g/2}}{\sum_{i=1}^{N} x_i^{g/2}} \right\} \leq 1$), whatever be a and σ^2 in the model $\theta(g)$, $g \geq 0$.

We study in this section the strategy $(p^* ; \hat{Y}^*)$ in some detail.

Consider $\hat{Y}^* = \hat{Y}_{HT} + a\,(X - \hat{X}_{HT})$ as an estimator of Y. <u>When a is known,</u> $\hat{Y}^*$ is p^*-unbiased for Y (p^* is the fixed sample size ($= n$) design with $\pi_i \propto x_i^{g/2}$ for $i = 1, 2, \ldots, N$).

If, more realistically, <u>a is unknown</u> and estimated by generalized least squares,

$$\hat{a}_g = \Big(\sum_{i \in s} (x_i\, y_i / x_i^g) \Big) \Big/ \Big(\sum_{i \in s} (x_i^2 / x_i^g) \Big)$$

then we propose along the lines of Cassel *et al* (1976) the following " $\theta(g)$-unbiased generalized regression estimator of Y "

$$\hat{Y}^*_{GRg} = \hat{Y}_{HT} + \hat{a}_g\,(X - \hat{X}_{HT}) \qquad (6.2.1)$$

with the design p^*. Observe that $\hat{Y}^*_{GRg}$ is <u>p^*-biased</u> (and $\theta(g)$-unbiased) <u>unless $g = 2$</u>.

If $g = 2$, $\hat{Y}^*_{GR2}$ reduces to $\frac{X}{n} \sum_{i \in s} \frac{y_i}{x_i}$, the p^*-unbiased Horvitz-Thompson estimator of Y.

If $g = 0$, $\hat{Y}^*_{GR0} = \frac{N}{n} \left(\sum_{i \varepsilon s} y_i \right) + \frac{\sum_{i \varepsilon s} x_i y_i}{\sum_{i \varepsilon s} x_i^2} \left(X - \frac{N}{n} \left(\sum_{i \varepsilon s} x_i \right) \right)$,

and p^* is the simple random sampling without replacement plan with fixed sample size (= n).

Considering $g = 1$, Cassel et al. (1976) show that the strategy $(p^* : \hat{Y}^*_{GR1})$ is highly efficient, in terms of expected sampling mean square error under $\theta(g)$, which they show by comparison with the best available competitors in the case $g = 1$. When $g = 1$, p^* is the fixed sample size (= n) plan with $\pi_i \propto \sqrt{x_i}$ for $i = 1, 2, \ldots, N$. Cassel et al. (1976) thus suggest a family of strategies $(p^* : \hat{Y}^*_{GRg})$ for estimating Y which when $g = 2$ and $g = 0$ correspond to well established strategies for estimating Y. They " suggest , generalizing from the limited study of gamma populations and $f = \frac{n}{N} = 0$, that the strategy obtained when $g = 1$ ought to be highly efficient for estimating Y in any practical situation where (i) sampling is from a skewed population ; (ii) the regression of $\mathcal{Y}$ on $\mathcal{X}$ is roughly linear through the origin with $Var_{\theta(g)} (y_i | x_i) \propto x_i^g$, the value of g being approximately unity" .

Regarding our study in chapter 5 and in here, observe that we are interested in obtaining the $\theta(g)$-optimal strategy for estimating Y, $g \geq 0$, whatever be a and σ^2 in the model $\theta(g)$, in the class of <u>all p-unbiased</u> strategies of Y with expected sample size fixed ($= n$ such that $\max_i \left\{ \frac{n x_i^{g/2}}{\sum_{i=1}^{N} x_i^{g/2}} \right\} \leq 1$). We would hence favour the strategy $(G\pi PS : \hat{Y}_{HT})$ for estimating Y (cf. T.M.F. Smith (1976)).

The strategy $(G\pi PS : \hat{Y}_{HT})$ is $\theta(g)$-optimal for estimating Y whatever be a and σ^2. Thus the $\theta(g)$-optimality of $(G\pi PS : \hat{Y}_{HT})$ assures the robustness of the strategy to the possible errors of complete characterization of the underlying prior distribution $\theta(g)$ of $\underset{\sim}{Y}$. Of course the design $G\pi PS$ incorporates all the prior knowledge we have about the population. Once we are careful in choosing the optimal design, the $\theta(g)$-optimal strategy in the class of all p-unbiased strategies of Y with expected sample size fixed is nice to deal with because of its robustness. On the other hand the Bayes estimator is very sensitive to the changes in the underlying prior distribution, but is independent of the sampling design (Godambe (1966)).

Royall and Herson (1973) proposed balanced sampling as a means of protection against " $\theta(g)$-bias'' when the underlying prior distribution $\theta(g)$ is such that the true regression in the corresponding model that further specifies $\theta(g)$ is a polynomial in $\mathcal{X}$.

CHAPTER 7

A CLASS OF ROBUST SAMPLING DESIGNS AND STRATIFICATION ON A SIZE VARIABLE

Summary

In section 7.1, we continue our study on p-unbiased strategies that are $\Theta(g)$-optimal and discuss an asymptotically design-unbiased strategy due to Brewer (1977). This proposed strategy turns out to be a good alternative to Cassel et. al.'s (1976) strategy $(p^* : \hat{Y}^*)$ discussed in Chapter 6 which requires the knowledge of the regression parameter a in the model $\Theta(g)$. Section 7.2 is devoted to studying stratification and allocation problems working with the model $\Theta(g)$ when the population size and the stratum sizes are large. The strategies $(G\pi PS : \hat{Y}_{HT})$, $(p^* : \hat{Y}^*)$ and Brewer's asymptotically efficient strategy are extended to stratified set up in this section. In section 7.3, a special stratification problem is discussed when the superpopulation model $\Theta(g)$, $g \in [1,2]$ provides a fair approximation to the actual situation. We then compare this stratified and unstratified strategies under $\Theta(g)$ for estimating Y. The results obtained in sections 7.2 and 7.3 are illustrated based on live data. Some of these populations were earlier considered by J.N.K. Rao (1969) and ...all (1970).

A Class of Robust Sampling Designs

In section 6.2 of Chapter 6, we studied the strategy $\hat{Y}^*_{GRg}$) for estimating Y. The strategy (p^* : $\hat{Y}^*$) which is a unbiased strategy, was derived to be $\theta(g)$-optimal in the class all p-unbiased strategies of Y with expected sample size fixed integral such that

$$\left\{\frac{n\, x_i^{g/2}}{\sum_{i=1}^{N} x_i^{g/2}}\right\} \leq 1)$$ when a is known, whatever be σ^2.

regression parameter a is not known in practice. Therefore a is estimated by $\hat{a}_g = \left(\sum_{i\in s} (x_i y_i/x_i^g)\right) / \left(\sum_{i\in s} (x_i^2/x_i^g)\right)$ generalized least squares (cf. Chapter 4), Cassel et al. 1976) obtained the strategy (p^* : $\hat{Y}^*_{GRg}$) as a compromise to the design-unbiased $\theta(g)$-optimal strategy (p^*: $\hat{Y}^*$). But observe that estimator $\hat{Y}^*_{GRg}$ is no longer p^*-unbiased and so no longer appeals to the design-oriented survey statistician. Brewer (1977) devised a sampling strategy for use in the context of large-scale surveys of populations containing units of very different sizes, such as official surveys of establishments and enterprises. Because the samples required are large, asymptotic theory is used. More specifically, working with the model $\theta(g)$ (cf. (5.1.1)),

...er (1977) considers the class of estimators (of the population ...l) given by

$$\hat{Y}_B = \sum_{i\epsilon s} y_i + \hat{a}_B \left(X - \sum_{i\epsilon s} x_i\right) \qquad (7.1.1)$$

...re $\hat{a}_B$ is an estimator of a. Since the model $\Theta(g)$ indicates ...t the dependence of y_i on x_i is linear and homogeneous, he ...siders the general homogeneous linear and $\Theta(g)$-unbiased estima-

$$\hat{a}_B = \sum_{i\epsilon s} w_i y_i \Big/ \left(\sum_{i\epsilon s} w_i x_i\right)$$

...a where the w_i are completely arbitrary. Further, the condi-...n of asymptotic design-unbiasedness is imposed, and the probabi-...ties of selection are chosen to minimize the expected mean square ...or of $\hat{Y}_B$ thus optimizing within the class. Explicitly, ...wer's asymptotic analysis is as follows :

The estimator $\hat{Y}_B$ is derived to be asymptotically unbiased ...er repeated sampling (design-unbiased) if and only if the inclus-...n probabilities with respect to the design are given by

$$\pi_i = \frac{\sum_{j=1}^{N} \pi_j w_j x_j}{\sum_{j=1}^{N} \pi_j w_j x_j + w_i \sum_{j=1}^{N} (1-\pi_j) x_j} = \frac{\alpha}{\alpha + w_i} \qquad (7.1.2)$$

where the value of α is proportional to the (arbitrary) sum of the w_i. Further, when N is large, the expected variance of $(p_B : \hat{Y}_B)$ is derived as

$$E_{\theta(g)} \text{Var.}(p_B : \hat{Y}_B) \simeq \sigma^2 \alpha^{-1} \sum_{i=1}^{N} w_i x_i^g \qquad (7.1.3)$$

where p_B is a design in which the inclusion probability π_i is as in (7.1.2) for $i = 1,2,\ldots,N$. Now, minimizing the right hand side of (7.1.3) subject to p_B being a design with expected sample size fixed (= n, integral, such that $\max_i \left\{ \frac{n x_i^{g/2}}{\sum_{i=1}^{N} x_i^{g/2}} \right\} \leq 1$)

the asymptotically efficient strategy is obtained when

$$w_i = \alpha(n^{-1} x_i^{-g/2} \sum_{i=1}^{N} x_i^{g/2} - 1) \text{ and hence when}$$

$$\pi_i = n x_i^{g/2} / (\sum_{i=1}^{N} x_i^{g/2}).$$

Thus, the above asymptotically efficient strategy $(p^* : \hat{Y}_B^*)$ is given by

$$\hat{Y}_B^* = \sum_{i\varepsilon s} y_i + \frac{\sum_{i\varepsilon s} (\frac{1}{\pi_i} - 1) y_i}{\sum_{i\varepsilon s} (\frac{1}{\pi_i} - 1) x_i} (X - \sum_{i\varepsilon s} x_i) \qquad (7.1.4)$$

with the design p^* being a fixed sample size (=n) design in the inclusion probabilities are given by $\pi_i \propto x_i^{g/2}$ for 1,2,..., N.

The expected variance of $(p^* : \hat{Y}_B^*)$ when N is large, is given by $\sigma^2 (\sum_{i=1}^{N} \pi_i^{-1} x_i^g - \sum_{i=1}^{N} x_i^g)$ which is equal to the ptotic minimum of the expected variance under $\Theta(g)$, $g \geq 0$ ny p-unbiased estimator given those π_i's as shown by Godambe Joshi (1965).

Furthermore, discussing the conventional stratified sampling me in which the classical ratio estimator is used within each tum with simple random sampling without replacement plan coupled Neyman's optimum allocation of sample size to strata and the ll and Herson (1973, 1973a) model-based robust procedure, er argues out the abolition of the need for size stratifica- while recommending his proposed strategy $(p^* : \hat{Y}_B^*)$ obtained ugh " conventional sampling rationale" .

rk 7.1.1 : Consider now the estimator $\hat{Y}^* = \hat{Y}_{HT} + a(X - \hat{X}_{HT})$ the p^* design. When a is not known in $\hat{Y}^*$ and estimated $\hat{a}$ =

$$\hat{a}_B^* = \sum_{i\epsilon s} \left(\frac{1}{\pi_i} - 1 \right) y_i \Big/ \left(\sum_{i\epsilon s} \left(\frac{1}{\pi_i} - 1 \right) x_i \right),$$

ote that $\hat{Y}^*$ with a replaced by $\hat{a}_B^*$ is given by $\hat{Y}^*_{\hat{a}_B^*}$

h is the same as $\hat{Y}_B^*$ of (7.1.4).

In the following sections of this chapter, we study the ategies (GπPS : $\hat{Y}_{HT}$), (p* : $\hat{Y}^*$), (p* : $\hat{Y}_B^*$) under a stratified up as in Chapters 2 and 3. In section 7.2 we substantiate wer's theory that the proposed strategy (p* : $\hat{Y}_B^*$) has the main ntage over current practice in that the use of (p* : $\hat{Y}_B^*$) for timating the population total Y helps remove the need for size atification.

While we have explicitly favoured the $\theta(g)$-optimal strategy PS : $\hat{Y}_{HT}$) that is (exactly) design-unbiased as compared to the biased (unless g = 2) strategy (p* : $\hat{Y}_{GRg}^*$) in Chapter 6, section 3 we note that the construction of GπPS designs for all populaons and all values of g is unduly complicated. The strategy : $\hat{Y}_B^*$) being (asymptotically) design-unbiased and asymptotically ficient as discussed above, is a good compromise to (GπPS : $\hat{Y}_{HT}$) the p*-unbiased strategy (p* : $\hat{Y}^*$) which requires the knowledge the regression parameter a .

2 Stratification on a Size Variable

In this section, we study Generalized πPS (GπPS) designs troduced earlier under a stratified set up. Values of $\mathscr{X}$, essentially positive valued characteristic highly correlated th the study variable $\mathscr{Y}$, are available for all the units the population, and the population of size N is divided into strata of sizes N_i, $i = 1,2,\ldots,k$, defined by k non-overlapping ranges of values of $\mathcal{X}$, i.e., $N = \sum_{i=1}^{k} N_i$. Let x_{ij}, denote the values of $\mathcal{X}$, $\mathcal{Y}$ respectively for the j-th unit the i-th stratum. Let a $G\pi PS$ design with expected sample be n_i be used in the i-th stratum for drawing a sample i.e., $\pi_{(i)j}$ stands for the probability of inclusion of the j-th unit of the i-th stratum in the sample from the i-th stratum, then

$$\pi_{(i)j} = n_i\, x_{ij}^{g/2} / \left(\sum_{j=1}^{N_i} x_{ij}^{g/2}\right) \quad \text{and} \quad \sum_j{}' x_{ij}^{1-(g/2)} = c, \qquad (7.2.1)$$

constant for all samples s with $p_s > 0$ from the i-th stratum where for a given i, $\sum_j{}'$ denotes summation over the units in the sample from the i-th stratum for $i = 1,2,\ldots,k$. Let sampling be carried out independently in each stratum. The expected sample sizes n_i's are such that $\sum_{i=1}^{k} n_i = n$, the total expected sample size. Observe that the constant c in (7.2.1) is $n_i X_i / \left(\sum_{j=1}^{N_i} x_{ij}^{g/2}\right)$ where $X_i = \sum_{j=1}^{N_i} x_{ij}$. <u>Throughout this chapter for a given</u> *i, $\sum_j$ will stand for $\sum_{j=1}^{N_i}$, summation over all the* <u>population units in the i-th stratum for $i = 1,2,\ldots,k$; and</u>

ll stand for $\sum_{i=1}^{k}$, summation over all strata. As an estimator

e population total $Y = \sum_i \sum_j y_{ij}$, consider the Horvitz-Thompson ator $\hat{Y}_S$ with this stratified GπPS design given by

$$\hat{Y}_S = \sum_i \left(\sum_j{}' (y_{ij}/\pi_{(i)j}) \right)$$

$$= \sum_i \left(\sum_j{}' (y_{ij}/(n_i\, x_{ij}^{g/2} / \sum_j x_{ij}^{g/2})) \right). \qquad (7.2.2)$$

consider the Horvitz-Thompson estimator $\hat{Y}_U$ of the popula- total Y based on a GπPS design with expected sample size $n_1 + n_2 + \dots + n_k$) to draw a sample from the whole (unstrati-) population given by

$$\hat{Y}_U = \sum_i \sum_j{}' (y_{ij}/\pi'_{(i)j}) \qquad (7.2.3)$$

re for a given i, $\sum_j{}'$ now denotes summation over those units of the sampled ones that belong to the i-th stratum for 1,2,...,k and $\pi'_{(i)j}$'s are the probabilities of inclusion en by

$$\pi'_{(i)j} = n\, x_{ij}^{g/2} / (\sum_i \sum_j x_{ij}^{g/2}) . \qquad (7.2.4)$$

assume that n, the values x_{ij} and the stratification pted are such that values n_i can be so chosen that none the $\pi_{(i)j}$ or $\pi'_{(i)j}$ exceeds unity).

Given $\underset{\sim}{X} = (x_{11}, x_{12}, \ldots, x_{kN_k})$, we explicitly formulate model $\Theta(g)$ thus :

$$\left.\begin{array}{l} E_{\Theta(g)}\ (y_{ij} \mid x_{ij}) = a\ x_{ij} \\ Var._{\Theta(g)}\ (y_{ij} \mid x_{ij}) = \sigma^2\ x_{ij}^{g} \\ Cov._{\Theta(g)}\ (y_{ij}, y_{rh} \mid x_{ij}, x_{rh}) = 0, \quad 1 \leq i,\ r \leq k \\ \qquad\qquad\qquad\qquad\qquad\qquad j \neq h \quad \text{if} \quad i = r \end{array}\right\} (7.2.5)$$

re a and σ^2 are the unknown parameters of the prior tribution $\Theta(g)$ of $\underset{\sim}{Y} = (y_{11}, y_{12}, \ldots, y_{kN_k})$ and $g \geq 0$. shall confine our attention to $g \in [1,2]$ in this chapter the model (7.2.5). Observe that the model (7.2.5) is the e as the model (2.1.3).

Under the model $\Theta(g)$ of (7.2.5) we now compare the two ategies, the stratified $G\pi PS$ design together with the timator $\hat{Y}_S$ and unstratified $G\pi PS$ design together with the timator $\hat{Y}_U$ in the expected variance sense.

w, Var.(stratified $G\pi PS$: $\hat{Y}_S$)

$$\sum_i \left\{ \sum_j (\pi_{(i)j}^{-1} - 1)\ y_{ij}^2 + \sum_{j \neq h}\sum (\pi_{(i)jh}\ \pi_{(i)j}^{-1}\ \pi_{(i)h}^{-1} - 1) y_{ij}\ y_{ih} \right\}$$

re $\pi_{(i)jh}$ is the joint inclusion probability of the j-th h-th units of the i-th stratum in the sample from the i-th atum for $i = 1, 2, \ldots, k$.

Further, under the model (7.2.5) we have

Var.(stratified GπPS : $\hat{Y}_S$)

$$\sum_{i=1}^{k} \left\{ \sum_{j=1}^{N_i} (\pi^{-1}_{(i)j} - 1)(\sigma^2 x^g_{ij} + a^2 x^2_{ij}) + a^2 \sum_{j\neq h}\sum (\pi_{(i)jh}\, \pi^{-1}_{(i)j}\, \pi^{-1}_{(i)h} - 1) x_{ij}\, x_{ih} \right\}$$

$$\sigma^2 \sum_i \sum_j (\pi^{-1}_{(i)j} - 1)\, x^g_{ij} + a^2 \text{ Var. (stratified GπPS : } \hat{X}_S)$$

re $\hat{X}_S$ is obtained by replacing the y_{ij}'s by x_{ij}'s in $\hat{Y}_S$.

rly Var.(stratified GπPS : $\hat{X}_S$) = 0 because

$$\hat{X}_S = \sum_{i=1}^{k} \sum_j{}' \frac{x_{ij}}{\pi_{(i)j}}$$

$$= \sum_{i=1}^{k} \left[\frac{(\sum_j x_{ij}^{g/2})}{n_i} (\sum_j{}' x_{ij}^{1-(g/2)}) \right]$$

$$= \sum_{i=1}^{k} X_i$$

$$= X$$

so

g) Var.(stratified GπPS : $\hat{Y}_S$) $= \sigma^2 \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(\frac{\sum_j x_{ij}^{g/2}}{n_i\, x_{ij}^{g/2}} - 1 \right) x^g_{ij}$

$$= \sigma^2 \left\{ \sum_{i=1}^{k} \left[\frac{(\sum_j x_{ij}^{g/2})^2}{n_i} \right] - \left(\sum_{i=1}^{k} \sum_{j=1}^{N_i} x_{ij}^{g} \right) \right\} . \qquad (7.2.6)$$

r $i = 1,2,\ldots,k$, $\mu_i = \sum_j x_{ij}^{g/2}$.

Minimizing the right hand side of (7.2.6) subject to the tion $\sum_{i=1}^{k} n_i = n$, we have that for $i = 1,2,\ldots,k$

$$-\sigma^2 \frac{\mu_i^2}{n_i^2} = -\lambda$$ where λ is the Lagrangian

plier and therefore we have that

$$n_i = n \frac{\mu_i}{\sum_i \mu_i}, \quad \text{for } i = 1,2,\ldots,k \qquad (7.2.7)$$

e " optimum allocation of the expected sample size to the la that minimizes the expected variance of the strategy atified GπPS : $\hat{Y}_S$) under the model $\Theta(g)$ of (7.2.5) " .

With this optimum allocation n_i of (7.2.7) we have that

$$\text{Var.(stratified G}\pi\text{PS} : \hat{Y}_S^{\text{opt.}}) = \sigma^2 \left\{ \frac{(\sum_i \mu_i)^2}{n} - \sum_i \sum_j x_{ij}^{g} \right\} . \qquad (7.2.8)$$

rem 7.2.1 : In the sense of expected variance under the l $\theta(g)$ of (7.2.5), for any allocation of the expected le size to the strata, apart from the optimum allocation

.2.7), stratified GπPS sampling strategy is inferior to

atified GπPS sampling strategy for estimating Y and with

ptimum allocation of the expected sample size to the strata,

strategies are equivalent for estimating Y.

Considering $\hat{Y}_U$ of (7.2.3) we have

$$\text{Var.}(G\pi PS : \hat{Y}_U) = \sigma^2 \sum_i \sum_j (\pi'^{-1}_{(i)j} - 1) x^g_{ij} + a^2 \text{Var.}(G\pi PS : \hat{X}_U)$$

$\hat{X}_U$ is obtained by replacing the y_{ij}'s by x_{ij}'s in

But since a GπPS design with expected sample size n is

to draw the sample from the whole (unstratified) population,

$(G\pi PS : \hat{X}_U) = 0$. Therefore,

$$\text{Var.}(G\pi PS : \hat{Y}_U) = \sigma^2 \sum_i \sum_j \left(\frac{\sum_i \sum_j x^{g/2}_{ij}}{n \; x^{g/2}_{ij}} - 1\right) x^g_{ij}$$

$$= \sigma^2 \left\{ \frac{(\sum_i \mu_i)^2}{n} - \sum_i \sum_j x^g_{ij} \right\} . \qquad (7.2.9)$$

(7.2.6) and (7.2.9) we have

$$\{\text{Var.}(\text{stratified } G\pi PS : \hat{Y}_S) - \text{Var.}(G\pi PS : \hat{Y}_U)\}$$

$$= \sigma^2 \left\{ \sum_{i=1}^{k} \frac{\mu_i^2}{n_i} - \frac{(\sum_i \mu_i)^2}{n} \right\}$$

$$\geq 0 , \qquad (7.2.10)$$

ity occurring if, and only if $n_i \propto \mu_i$ for $i = 1,2,\ldots,k$ with the optimum allocation (7.2.7) of the expected size to the strata, stratified $G\pi PS$ sampling strategy unstratified $G\pi PS$ sampling strategy are equivalent under for estimating Y.

7.2.2 : Since from (7.2.10) we have that

$$\{\text{Var.}(\text{stratified } G\pi PS : \hat{Y}_S) - \text{Var.}(G\pi PS : \hat{Y}_U)\} \geq 0,$$

ever be a and σ^2 in the model $\Theta(g)$, we have proved that unstratified $G\pi PS$ strategy is better than the stratified strategy under $\Theta(g)$ for estimating Y, with any allocation the expected sample size to the strata.

7.2.3 : Consider the model $\Theta(g)$ of (7.2.5). Let within stratum a design p^* with fixed sample size n_i be used drawing the sample for which $\pi_{(i)j}$'s are proportional to

i.e., $\pi_{(i)j} = n_i \, x_{ij}^{g/2} / (\sum_{j=1}^{N_i} x_{ij}^{g/2})$ for $j = 1,2,\ldots,N_i$;

1,2,...,k. Let sampling be carried out independently in each tum. The sample sizes n_i's are each that $\sum_{i=1}^{k} n_i = n$.

a is known, whatever be σ^2 in the model $\Theta(g)$ of (7.2.5), estimator of the population total Y consider

$$\hat{Y}^*_S = \sum_{i=1}^{k} (\sum_j{}' \frac{y_{ij}}{\pi_{(i)j}} + a\ (X_i - \sum_j{}' \frac{x_{ij}}{\pi_{(i)j}})) \qquad (7.2.11)$$

he stratified p^* design described above. In (7.2.11)

$\sum_j x_{ij}$ for $i = 1,2,\ldots,k$ and for a given i, $\sum_j{}'$ denotes

ion over the sampled n_i units from the i-th stratum for

,2,...,k. From our results of chapter 6 we have that

$$\text{Var.}(\text{stratified } p^* : \hat{Y}^*_S) = \sigma^2 \sum_{i=1}^{k} \sum_{j=1}^{N_i} (\frac{1}{\pi_{(i)j}} - 1)\ x^g_{ij}\ . \qquad (7.2.12)$$

consider the estimator $\hat{Y}^*_U$ of the population total Y

the design p^* with fixed sample size $n\ (= \sum_{i=1}^{k} n_i)$ in

the inclusion probabilities $\pi'_{(i)j}$'s are proportional to

is used for drawing a sample from the whole (unstratified)

ation, i.e.,

$$= n\ x^{g/2}_{ij} / (\sum_i \sum_j x^{g/2}_{ij}), \quad \text{given by}$$

$$= \sum_{i=1}^{k} \sum_j{}' \frac{y_{ij}}{\pi'_{(i)j}} + a(X - \sum_{i=1}^{k} \sum_j{}' \frac{x_{ij}}{\pi'_{(i)j}}) \qquad (7.2.13)$$

a is known, whatever be σ^2 in the model $\theta(g)$ of (7.2.5).

7.2.13) above $\sum_j{}'$ for a given i denotes summation over those

s out of the sampled n that belong to the i-th stratum for

1,2,...,k. Now, under $\theta(g)$ we have

$$\text{(g)}\ Var.(p^* : \hat{Y}^*_U) = \sigma^2 \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(\frac{1}{\pi'_{(i)j}} - 1\right) x^g_{ij} . \qquad (7.2.14)$$

aring (7.2.12) and (7.2.14) we have

orem 7.2.4 : The unstratified strategy $(p^* : \hat{Y}^*_U)$ is better the strategy (stratified $p^* : \hat{Y}^*_S$) for estimating Y in the ected variance sense under the model $\theta(g)$ of (7.2.5) when is known, whatever be σ^2, with any allocation of the sample e to the strata.

of : Immediate from the proof of Theorem 7.2.1.

ark 7.2.5 : Now consider the more realistic case when a is t known in the model $\theta(g)$ of (7.2.5). Suppose now within ch stratum a design p^* with fixed sample size n_i is used r drawing the sample in which $\pi_{(i)j} = n_i\, x^{g/2}_{ij} / (\sum_j x^{g/2}_{ij})$ for $= 1,2,\ldots,N_i$; $i = 1,2,\ldots,k$ as in Remark 7.2.3. Let sampling carried out independently in each stratum. The sample sizes n such that $\sum_{i=1}^{k} n_i = n$. As an estimator of $Y = \sum_i \sum_j y_{ij}$ onsider now

$$\sum_{i=1}^{k} \left(\sum_j{}' y_{ij} + \frac{\sum_j{}' \left(\frac{1}{\pi_{(i)j}} - 1\right) y_{ij}}{\sum_j{}' \left(\frac{1}{\pi_{(i)j}} - 1\right) x_{ij}} \left(X_i - \sum_j{}' x_{ij}\right)\right) \qquad (7.2.15)$$

the stratified p^* design described above. For a given i, (7.2.15) denotes summation over the sampled n_i units from ith stratum for $i = 1,2,\ldots,k$; $X_i = \sum_j x_{ij}$ for $i=1,2,\ldots,k$. [...]e that $\hat{Y}^*_{B,st.}$ of (7.2.15) above is obtained from (7.2.11) by [...]ting a by $\hat{a}^*_B$ within each stratum (cf. Remark 7.1.1). [...]ection 7.1 we have that when the stratum sizes N_i are large [...]ewer's asymptotic analysis to hold within each stratum

$$\text{Var.}(\text{stratified } p^* : \hat{Y}^*_{B,st.}) \simeq \sigma^2 \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(\frac{1}{\pi_{(i)j}} - 1\right) x_{ij}^{g}. \qquad (7.2.16)$$

[...], as an estimator of Y consider

$$= \sum_{i=1}^{k} \sum_j{}' y_{ij} + \frac{\sum_{i=1}^{k} \sum_j{}' \left(\frac{1}{\pi'_{(i)j}} - 1\right) y_{ij}}{\sum_{i=1}^{k} \sum_j{}' \left(\frac{1}{\pi'_{(i)j}} - 1\right) x_{ij}} \left(X - \sum_{i=1}^{k} \sum_j{}' x_{ij}\right) \qquad (7.2.17)$$

the design p^* with fixed sample size n in which the [...]sion probabilities $\pi'_{(i)j}$ are proportional to $x_{ij}^{g/2}$ is used [...]rawing a sample from the whole (unstratified) population,

$\pi'_{(i)j} = n\, x_{ij}^{g/2} / (\sum_i \sum_j x_{ij}^{g/2})$. In (7.2.17) above $\sum_j'$ for a i indicates summation over those units out of the sampled [th]at belong to the i-th stratum for $i = 1,2,\ldots,k$. Note that [..] above is obtained from (7.2.13) by estimating a using $\hat{a}_B^*$ [..] the p^* design is used for drawing a sample from the whole [..]lation (cf. Remark 7.1.1). Now, under $\Theta(g)$ of (7.2.5) when [..] large, we have

$$\text{Var.}(p^* : \hat{Y}^*_{B,Un.}) \simeq \sigma^2 \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(\frac{1}{\pi'_{(i)j}} - 1\right) x_{ij}^{g}, \qquad (7.2.18)$$

[..] Brewer (1977) (see section 7.1). Thus, comparing (7.2.16) (7.2.18) we have

[The]orem 7.2.6 : The unstratified (asymptotically design-unbiased) [stra]tegy $(p^* : \hat{Y}^*_{B,Un.})$ is better than the stratified (asymptotically [desi]gn-unbiased) strategy (stratified $p^* : \hat{Y}^*_{B,st.}$) for estimating [..] in the expected variance sense under $\Theta(g)$ of (7.2.5) with any [all]ocation of sample size to the strata.

[Pro]of : Immediate from the proof of Theorem 7.2.1.

We illustrate our Theorem 7.2.6 by considering live data [of] six populations, data for which are provided in appendix AI, [..]I, AIV, AV and AVI. Four out of these six populations were

considered by J. N. K. Rao (1969) and Royall (1970). We
ribe the populations and the auxiliary information $\mathcal{X}$ consider
each of the populations. We illustrate for each population,
fficiency of the unstratified strategy (p^* : $\hat{Y}^*_{B,Un.}$) as compared
he stratified strategy (stratified p^* : $\hat{Y}^*_{B,st.}$) for values of
<u>1.0(0.1)2.0</u> i.e., Table 7.2.7 to Table 7.2.12 provide
) Var.(stratified p^* : $\hat{Y}^*_{B,st.}$)/$E_{\theta(g)}$Var.(p^* : $\hat{Y}^*_{B,Un.}$) for
1.0(0.1) 2.0.

For the populations considered either 3 or 4 strata were
ed for illustration. The total sample size was fixed to be
er 9 or 12. <u>Arbitrary allocations</u> of the total sample size
considered with the restriction that at least two units be
n from each stratum for the estimability of the variance of
estimator. Optimum allocation of the sample size to the
ta that minimizes the expected variance of the strategy
atified p^* : $\hat{Y}^*_{B,st.}$) under $\theta(g)$ was not computed. Therefore,
entries are ≥ 1 in Table 7.2.7 through Table 7.2.12.

<u>lation 1</u> : (data provided in Appendix AI).

We consider real data on crops and grass acreage ($\mathcal{X}$)
en by Sampford (1962, p.61) which relates to 35 farms in
ey. The population was divided into three strata (Sampford,
2) containing farms 1-12, farms 13-24 and farms 25-35. Here

rata sizes are $N_1 = 12$, $N_2 = 12$ and $N_3 = 11$. An overall size $n = 9$ is taken for illustration and the allocations ered were (2,3,4), (2,4,3), (3,2,4), (3,4,2), (3,3,3), (4,2,3),) and (2,2,5). Table 7.2.7 provides

Var. (stratified $p^* : \hat{Y}^*_{B,st.}$)/$E_{\Phi(g)}$ Var. ($p^* : \hat{Y}^*_{B,Un.}$) for each of allocations for $g = 1.0(0.1)2.0$.

Table 7.2.7

e efficiency of the strategy ($p^* : \hat{Y}^*_{B,Un.}$) as compared to e strategy (stratified $p^* : \hat{Y}^*_{B,st.}$) for a fixed overall sample ze $n = 9$ for $g = 1.0(0.1)2.0$.

Population given in Sampford (1962 , p.61).

Allocation							
(2,3,4)	(2,4,3)	(3,2,4)	(3,4,2)	(3,3,3)	(4,2,3)	(4,3,2)	(2,2,5)
1.0019	1.1198	1.1110	1.4647	1.1225	1.2834	1.5192	1.0802
1.0057	1.1426	1.1203	1.5308	1.1524	1.3144	1.5881	1.0701
1.0127	1.1695	1.1319	1.6024	1.1859	1.3482	1.6620	1.0625
1.0228	1.2007	1.1459	1.6797	1.2229	1.3854	1.7413	1.0573
1.0360	1.2362	1.1623	1.7628	1.2635	1.4257	1.8260	1.0545
1.0524	1.2759	1.1813	1.8518	1.3078	1.4692	1.9163	1.0540
1.0719	1.3199	1.2027	1.9469	1.3558	1.5162	2.0123	1.0559
1.0945	1.3683	1.2267	2.0482	1.4077	1.5666	2.1143	1.0602
1.1202	1.4211	1.2533	2.1559	1.4634	1.6207	2.2224	1.0668
1.1491	1.4783	1.2826	2.2701	1.5231	1.6785	2.3368	1.0757
1.1812	1.5399	1.3146	2.3910	1.5869	1.7401	2.4577	1.0870

Of the 88 ratios in the above table, 86 entries are greater one. When $g = 1.0$ and 1.1, for the allocation (2,3,4) of size to strata the efficiency of the unstratified strategy ($\hat{Y}^{*}_{B,Un.}$) as compared to the stratified strategy is nearly 1. is so, since the optimum allocation $\underset{\sim}{n}$ of sample size to given by $n_i = n \sum_j x_{ij}^{g/2} / \sum_i \sum_j x_{ij}^{g/2}$, $i = 1,2,3$ for $g = 1.0$ 1.1 is nearly (2.3,4).

lation 2 : (data provided in Appendix AIII).

Live example of the population consisting of four States India was considered. The district-wise 1961 census population res (rounded off to thousands) was considered as $\mathcal{X}$, the liary information. The four States were treated as four strata. overall sample size $n = 12$ was considered. The allocations 2,4), (3,3,4,2), (2,2,2,6), (2,3,2,5), (2,3,4,3) and (3,3,3,3) considered for illustration.

Table 7.2.8

The efficiency of the strategy $(p^* : \hat{Y}^*_{B,Un.})$ as compared to the strategy (stratified $p^* : \hat{Y}^*_{B,st.}$) for a fixed overall sample size $n = 12$ for $g = 1.0(0.1)2.0$.

	Allocation				
4,2,4)	(3,3,4,2)	(2,2,2,6)	(2,3,2,5)	(2,3,4,3)	(3,3,3,3)
.0391	1.2271	1.3949	1.1322	1.0957	1.0988
.0349	1.2359	1.3945	1.1288	1.1016	1.1020
.0311	1.2451	1.3943	1.1256	1.1075	1.1054
.0275	1.2548	1.3941	1.1227	1.1136	1.1091
.0243	1.2648	1.3940	1.1199	1.1198	1.1130
.0214	1.2754	1.3939	1.1173	1.1261	1.1173
.0186	1.2864	1.3939	1.1149	1.1326	1.1217
.0162	1.2981	1.3939	1.1125	1.1392	1.1265
.0140	1.3103	1.3939	1.1103	1.1461	1.1316
.0120	1.3231	1.3939	1.1082	1.1532	1.1370
.0103	1.3367	1.3939	1.1062	1.1605	1.1427

The efficiency of the unstratified strategy $(p^* : \hat{Y}^*_{B,Un.})$ mpared to the strategy (stratified $p^* : \hat{Y}^*_{B,st.}$) is nearly ber cent for the allocation (2,2,2,6) of sample size to is for all values of g from 1.0 through 2.0 in Table 7.2.8. the 66 entries are greater than 1 in the above table.

lation 3 : (data provided in the Appendix AIV).

Data from a random sample of 43 Kraals was considered ws (1960), p.159). Total number of persons (including ntees) was treated as the auxiliary information $\mathcal{X}$. The y variable $\mathcal{Y}$ was the number of absentees. The population divided into 4 strata for illustration containing Kraals , Kraals 11-22, Kraals 23-35, Kraals 36-43. Here $N_1 = 10$, 12, $N_3 = 13$, $N_4 = 8$. The overall sample size was fixed n = 12. The allocations (2,4,2,4), (3,3,4,2), (2,2,2,6), ,2,5), (2,3,4,3) and (3,3,3,3) were considered for illustra-

Table 7.2.9

The efficiency of the strategy $(p^* : \hat{Y}^*_{B,Un.})$ as compared to the strategy (stratified $p^* : \hat{Y}^*_{B,st.}$) for an overall sample size $n = 12$ for $g = 1.0(0.1)2.0$.

Population given in Yates (1960), p.159.

Allocation					
(2,4,2,4)	(3,3,4,2)	(2,2,2,6)	(2,3,2,5)	(2,3,4,3)	(3,3,3,3)
1.2822	1.0206	1.5736	1.3655	1.0529	1.0488
1.2819	1.0214	1.5750	1.3656	1.0519	1.0492
1.2816	1.0222	1.5766	1.3658	1.0510	1.0496
1.2814	1.0231	1.5783	1.3661	1.0501	1.0500
1.2813	1.0240	1.5802	1.3665	1.0493	1.0503
1.2813	1.0249	1.5823	1.3670	1.0485	1.0507
1.2814	1.0259	1.5845	1.3676	1.0478	1.0510
1.2815	1.0269	1.5869	1.3683	1.0472	1.0513
1.2818	1.0279	1.5894	1.3691	1.0465	1.0516
1.2821	1.0289	1.5921	1.3700	1.0460	1.0519
1.2825	1.0300	1.5950	1.3710	1.0454	1.0522

For the arbitrary allocations of sample size to strata ...sidered in the above table, the efficiency of the unstratified ...ategy ($p^* : \hat{Y}^*_{B,Un.}$) as compared to the stratified strategy, ...ges between 102 per cent and 159 per cent. All 66 entries greater than one. For the equal allocation of sample size strata the efficiency is nearly 105 per cent for all values g from 1.0 through 2.0 in Table 7.2.9.

...ulation 4 : (data provided in the Appendix AV).

Eye estimated volume of timber, x (cu. ft./ 1/10 acre), considered on 25 sample plots (Yates (1960), p.163). The ...dy variable y was the measured volume of timber. The ...ulation was divided for illustration into 3 strata containing ...ots 1-8, plots 9-18, plots 19-25. Thus $N_1 = 8$, $N_2 = 10$ ... $N_3 = 7$. An overall sample size $n = 9$ was taken and the ...ocations (2,3,4), (2,4,3), (3,2,4), (3,4,2), (3,3,3), (4,2,3), ...3,2) and (2,2,5) were considered for illustration.

Table 7.2.10

The efficiency of the strategy $(p^* : \hat{Y}^*_{B,Un.})$ as compared to the strategy (stratified $p^* : \hat{Y}^*_{B,st.}$) for an overall sample size $n = 9$ for $g = 1.0(0.1)2.0$.

Population given in Yates (1960 , p.163).

Allocation							
(2,3,4)	(2,4,3)	(3,2,4)	(3,4,2)	(3,3,3)	(4,2,3)	(4,3,2)	(2,2,5)
1.0792	1.0202	1.2981	1.1211	1.0447	1.3486	1.2306	1.3868
1.0727	1.0181	1.2975	1.1335	1.0478	1.3552	1.2459	1.3785
1.0667	1.0166	1.2970	1.1465	1.0512	1.3620	1.2617	1.3707
1.0612	1.0156	1.2967	1.1601	1.0550	1.3690	1.2778	1.3632
1.0561	1.0152	1.2966	1.1742	1.0590	1.3761	1.2945	1.3561
1.0514	1.0154	1.2968	1.1889	1.0634	1.3835	1.3116	1.3493
1.0471	1.0161	1.2971	1.2042	1.0681	1.3912	1.3292	1.3429
1.0431	1.0172	1.2977	1.2200	1.0731	1.3991	1.3473	1.3369
1.0395	1.0189	1.2985	1.2365	1.0784	1.4073	1.3660	1.3312
1.0363	1.0210	1.2996	1.2536	1.0840	1.4159	1.3852	1.3258
1.0333	1.0235	1.3008	1.2713	1.0900	1.4248	1.4051	1.3207

In the above table, corresponding to the allocation (2,4,3) mple size to strata, observe that the efficiency of the atified strategy (p^* : $\hat{Y}^*_{B,Un.}$) as compared to the strategy atified p^* : $\hat{Y}^*_{B,st.}$) is nearly one for values of g ranging een 1.1 and 1.8. The optimum allocation of sample size to ta by chance effect of rounding off is nearly (2,4,3) for 1.1 through 1.8.

lation 5 and Population 6 : (data provided in the Appendix AVI).

Area under wheat in 1936 ($\mathcal{Y}$ in acres) and total cultivated in 1931 ($\mathcal{X}$ in acres) for 34 villages in Lucknow Subdivision ttar Pradesh State of India (Sukhatme (1954), p. 183) was idered. The 3 strata formed consisted of villages 1-12, lages 13-23, villages 24-34. Thus $N_1 = 12$, $N_2 = 11$ and = 11. The total sample size was fixed to be $n = 9$ and the ocations (2,3,4), (2,4,3), (3,2,4), (3,4,2), (3,3,3), (4,2,3), 3,2) and (2,2,5) were considered for illustration. Table 7.2.11 Table 7.2.12 below give the efficiency comparisons for popula- n 5 ($\mathcal{Y}$ being area under wheat in 1936) and population 6 (being total cultivated area in 1931) respectively.

The efficiency of the unstratified strategy (p^* : $\hat{Y}^*_{B,Un.}$) compared to the strategy (stratified p^* : $\hat{Y}^*_{B,st.}$) is > 1 all the allocations considered in the following tables for values of $g = 1.0(0.1)2.0$.

Table 7.2.11

he efficiency of the strategy $(p^* : \hat{Y}^*_{B,Un.})$ as compared to the trategy (stratified $p^* : \hat{Y}^*_{B,st.}$)for an overall sample size = 9 for $g = 1.0(0.1)2.0$.

Population given in Sukhatme (1954 , p.183). $\mathcal{X}$ being area nder wheat in 1936.

Allocation							
4,3,4)	(2,4,3)	(3,2,4)	(3,4,2)	(3,3,3)	(4,2,3)	(4,3,2)	(2,2,5)
.2817	1.2952	1.1089	1.1493	1.0328	1.0359	1.0629	1.3989
1.2975	1.3110	1.1129	1.1534	1.0371	1.0341	1.0611	1.4144
1.3138	1.3272	1.1173	1.1577	1.0415	1.0326	1.0595	1.4307
1.3306	1.3439	1.1223	1.1621	1.0461	1.0314	1.0580	1.4479
1.3481	1.3611	1.1277	1.1669	1.0509	1.0306	1.0567	1.4660
1.3662	1.3789	1.1336	1.1719	1.0559	1.0300	1.0555	1.4851
1.3850	1.3974	1.1399	1.1771	1.0611	1.0297	1.0546	1.5051
1.4045	1.4166	1.1466	1.1828	1.0665	1.0297	1.0538	1.5263
1.4249	1.4366	1.1537	1.1888	1.0721	1.0299	1.0532	1.5485
1.4461	1.4574	1.1613	1.1952	1.0780	1.0302	1.0528	1.5719
1.4684	1.4793	1.1694	1.2021	1.0840	1.0308	1.0526	1.5966

Table 7.2.12

The efficiency of the strategy $(p^* : \hat{Y}^*_{B,Un.})$ as compared to the strategy (stratified $p^* : \hat{Y}^*_{B,st.}$) for an overall sample size $n = 9$ for $g = 1.0(0.1)2.0$.

Population given in Sukhatme(1954,p.183). $\mathcal{X}$ being total cultivated area in 1931.

Allocation							
(2,3,4)	(2,4,3)	(3,2,4)	(3,4,2)	(3,3,3)	(4,2,3)	(4,3,2)	(2,2,5)
1.3097	1.3085	1.1330	1.1296	1.0399	1.0436	1.0413	1.4392
1.3297	1.3278	1.1392	1.1336	1.0460	1.0421	1.0384	1.4587
1.3505	1.3479	1.1460	1.1380	1.0525	1.0411	1.0358	1.4793
1.3722	1.3686	1.1534	1.1428	1.0595	1.0405	1.0334	1.5008
1.3947	1.3902	1.1615	1.1480	1.0669	1.0404	1.0314	1.5235
1.4182	1.4126	1.1702	1.1536	1.0749	1.0407	1.0297	1.5473
1.4426	1.4359	1.1797	1.1597	1.0833	1.0415	1.0282	1.5722
1.4680	1.4601	1.1898	1.1662	1.0921	1.0428	1.0271	1.5984
1.4945	1.4854	1.2006	1.1731	1.1016	1.0445	1.0262	1.6258
1.5221	1.5116	1.2122	1.1806	1.1115	1.0467	1.0256	1.6546
1.5509	1.5390	1.2245	1.1886	1.1221	1.0493	1.0254	1.6848

Further Extensions

We shall denote throughout this section the $G\pi PS$ design a fixed $g = g_o \varepsilon [1,2]$ by " $G\pi PSg_o$ design".

Now, considering a stratified set up as in section 7.2 let $G\pi PSg_o$ design with expected sample size n_i be used for wing a sample in the i-th stratum for $i = 1,2,...,k$. The ected sample sizes n_i's are such that $\sum_{i=1}^{k} n_i = n$. Let sampling carried out independently in each stratum. In our notation of tion 7.2,

$$\pi_{(i)j} = n_i \, x_{ij}^{g_o/2} / \left(\sum_{j=1}^{N_i} x_{ij}^{g_o/2} \right) \quad \text{and} \quad \sum_j{}' \, x_{ij}^{1-(g_o/2)} = c, \qquad (7.3.1)$$

constant $(= n_i X_i / (\sum_{j=1}^{N_i} x_{ij}^{g_o/2}))$ for all samples s with $p_s > 0$ from the i-th stratum where $\sum_j{}'$ for a given i denotes summation er the sampled units from the i-th stratum for $i = 1,2,...,k$. rking with the super-population model $\Theta(g)$ of (7.2.5) for $\varepsilon[1,2]$, we derive below the optimum allocation of the expected mple size to strata that minimizes the expected variance, under $g)$, of the strategy consisting of the stratified $G\pi PSg_o$ design gether with the estimator $\hat{Y}_S = \sum_{i=1}^{k} (\sum_j{}' (y_{ij}/\pi_{(i)j}))$ of the pulation total $Y = \sum_i \sum_j y_{ij}$. Under the model $\Theta(g)$ of (7.2.5) have for $g \varepsilon [1,2]$, g_o fixed $\varepsilon [1,2]$.

$$\text{Var}(\text{stratified } G\pi PSg_o : \hat{Y}_S) = \sigma^2 \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(\frac{\sum_j x_{ij}^{g_o/2}}{n_i \, x_{ij}^{g_o/2}} - 1 \right) x_{ij}^{g}$$

$$= \sigma^2 \left\{ \sum_{i=1}^{k} \frac{(\sum_j x_{ij}^{g-(g_o/2)})(\sum_j x_{ij}^{g_o/2})}{n_i} - \sum_{i=1}^{k} \sum_{j=1}^{N_i} x_{ij}^{g} \right\}. \tag{7.3.2}$$

Let $\delta_i = \sum_j x_{ij}^{g_o/2}$ and $\eta_i = \sum_j x_{ij}^{g-(g_o/2)}$ for $i = 1,2,\ldots,k$.

...zing the right hand side of (7.3.2) subject to the condition

... $= n$, we have that for $i = 1,2,\ldots,k$

$$\frac{-\delta_i \eta_i}{n_i^2} = -\lambda \quad \text{where}$$

the Lagrangian multiplier and therefore for $i = 1,2,\ldots,k$

$$n_i = n \sqrt{\delta_i \eta_i} \Big/ \left(\sum_{i=1}^{k} \sqrt{\delta_i \eta_i} \right) \tag{7.3.3}$$

... " optimum allocation of the expected sample size to the ... that minimizes the expected variance of the strategy ...tified $G\pi PSg_o : \hat{Y}_S$) under $\Theta(g)$ ".

Suppose now, a $G\pi PSg_o$ design with expected sample size $\sum_{i=1}^{k} n_i$) is used to draw a sample from the whole (unstratified) ...tion we have

$$\pi_{ij} = n\, x_{ij}^{g_o/2} / \Big(\sum_{i=1}^{k} \sum_{j=1}^{N_i} x_{ij}^{g_o/2} \Big) \quad \text{and} \quad \sum_{i=1}^{k} \sum_j{}' \, x_{ij}^{1-(g_o/2)} = \text{a constant} \tag{7.3.4}$$

all samples s with $p_s > 0$ from the whole (unstratified) population where $c = n\,X/(\sum_i \sum_j x_{ij}^{g_o/2})$. For a given i, $\sum_j{}'$ in (7.3.4) denotes summation over those units out of the sampled ones that belong to the i-th stratum for $i = 1,2,\dots,k$. As an estimator of the population total with this design to draw a sample from the whole population consider

$$\hat{Y}_U = \sum_{i=1}^{k} \sum_j{}' \, (y_{ij}/\pi'_{(i)j})$$

$$= \sum_{i=1}^{k} \sum_j{}' \, \Big(y_{ij}/\Big(n\, x_{ij}^{g_o/2} / \sum_i \sum_j x_{ij}^{g_o/2}\Big)\Big).$$

Considering $\theta(g)$ of (7.2.5) we have

$$\theta(g)\,\text{Var.}(G\pi PSg_o : \hat{Y}_U) = \sigma^2 \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(\frac{\sum_i \sum_j x_{ij}^{g_o/2}}{n\, x_{ij}^{g_o/2}} - 1 \right) x_{ij}^{g}$$

$$= \sigma^2 \left\{ \frac{(\sum_i \delta_i)\,(\sum_i \eta_i)}{n} - \sum_{i=1}^{k} \sum_{j=1}^{N_i} x_{ij}^{g} \right\}. \tag{7.3.5}$$

Remark 7.3.1 : We assume that n, the x_{ij}'s and the stratification are such that the values n_i's can be so chosen that none of

$\pi_{(i)j}$'s or $\pi'_{(i)j}$'s in (7.3.1) and (7.3.4) exceeds unity.

orem 7.3.2 : The strategy (stratified $G\pi PSg_o$: $\hat{Y}_S$) with the imum allocation (7.3.3) of the expected sample size to strata superior to the unstratified strategy ($G\pi PSg_o$: $\hat{Y}_U$) for imating Y under $\Theta(g)$ for all values of $g \in [1,2]$ different m g_o. At $g = g_o$, both the strategies are equivalent for imating Y.

of : With the optimum allocation n_i of (7.3.3) to the strata, have that

$$_{(g)}\text{Var.}(\text{stratified } G\pi PSg_o : \hat{Y}_S^{\text{ opt.}}) = \sigma^2 \left\{ \frac{(\sum_{i=1}^{k} \sqrt{\delta_i \eta_i})^2}{n} - \sum_{i=1}^{k} \sum_{j=1}^{N_i} x_{ij}^g \right\}$$

om (7.3.2),

w,

$$_{(g)}\text{Var.}(G\pi PSg_o : \hat{Y}_U) = \sigma^2 \left\{ \frac{(\sum_i \delta_i)(\sum_i \eta_i)}{n} - \sum_{i=1}^{k} \sum_{j=1}^{N_i} x_{ij}^g \right\}.$$

erefore,

$$_{(g)}[\text{Var.}(\text{stratified } G\pi PSg_o : \hat{Y}_S^{\text{ opt.}}) - \text{Var.}(G\pi PSg_o : \hat{Y}_U)]$$

$$= \frac{\sigma^2}{n} \left\{ (\sum_{i=1}^{k} \sqrt{\delta_i \eta_i})^2 - (\sum_i \delta_i)(\sum_i \eta_i) \right\}$$

$$\leq 0, \quad \forall\, g \in [1,2], \text{ and } \underline{\text{fixed}}\ g_o \in [1,2] \qquad (7.3.6)$$

ever be σ^2, equality occuring if, and only if, δ_i is ortional to η_i for $i = 1,2,\ldots,k$. Observe that when g_o, $\delta_i = \eta_i$ for $i = 1,2,\ldots,k$ and so at $g = g_o$, with optimum allocation (7.3.3) both the strategies are equivalent estimating Y.

rk 7.3.3 : The stratified $G\pi PSg_o$ sampling strategy with optimum allocation (7.3.3) of the expected sample size to ata is better than the unstratified $G\pi PSg_o$ sampling strategy estimating Y under $\Theta(g)$ for all $g \in [1,2]$, whatever be and σ^2 in the model $\Theta(g)$ of (7.2.5).

ark 7.3.4 : Let within each stratum a design $p^*_{g_o/2}$ with ed sample size n_i be used for drawing the sample in which inclusion probabilities $\pi_{(i)j}$'s are proportional to $x_{ij}^{g_o/2}$., $\pi_{(i)j} = n_i\, x_{ij}^{g_o/2} / (\sum_j x_{ij}^{g_o/2})$ for $i=1,2,\ldots,k$; $j=1,2,\ldots,N_i$; fixed $\in [1,2]$. Let sampling be carried out independently in ch stratum. The sample sizes n_i's are such that $\sum_{i=1}^{k} n_i = n$.

en a is known, whatever be σ^2 in the model $\Theta(g)$ of (7.2.5), an estimator of the population total $Y = \sum_i \sum_j y_{ij}$ consider

$$\hat{Y}_S^* = \sum_{i=1}^{k} \left(\sum_j{}' \frac{y_{ij}}{\pi_{(i)j}} + a\left(X_i - \sum_j{}' \frac{x_{ij}}{\pi_{(i)j}}\right)\right)$$

e $X_i = \sum_{j=1}^{N_i} x_{ij}$ for $i = 1,2,\ldots,k$ and for a given i,

enotes summation over the sampled n_i units from the i-th

tum for $i = 1,2,\ldots,k$. Under the model $\theta(g)$ of (7.2.5)

ave

$$\text{Var.}(\text{stratified } p^*_{g_o/2} : \hat{Y}^*_S) = \sigma^2 \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(\frac{\sum_j x_{ij}^{g_o/2}}{n_i \, x_{ij}^{g_o/2}} - 1 \right) x_{ij}^{g} \,. \tag{7.3.7}$$

Next consider the estimator $\hat{Y}^*_U$ of the population total

hen a is known in the model $\theta(g)$ of (7.2.5) with the fixed

ple size (= n) design $p^*_{g_o/2}$ used to draw a sample from

whole (unstratified) population in which the inclusion

ability $\pi'_{(i)j} = n \, x_{ij}^{g_o/2} / (\sum_i \sum_j x_{ij}^{g_o/2})$ for $i = 1,2,\ldots,k$;

$1,2,\ldots,N_i$; given by

$$\hat{Y}^*_U = \sum_{i=1}^{k} \sum_j{}' \frac{y_{ij}}{\pi'_{(i)j}} + a\left(X - \sum_{i=1}^{k} \sum_j{}' \frac{x_{ij}}{\pi'_{(i)j}}\right)$$

re for a given i, $\sum_j'$ now denotes summation over those units

of the sampled n that belong to the i-th stratum for

$1,2,\ldots,k$. Under $\theta(g)$ of (7.2.5) we have now

$$V(p^*_{g_0/2} : \hat{Y}^*_U) = \sigma^2 \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(\frac{\sum_i \sum_j x_{ij}^{g_0/2}}{n\, x_{ij}^{g_0/2}} - 1 \right) x_{ij}^{g} . \qquad (7.3.8)$$

ing (7.3.7)and (7.3.8) we have

7.3.5 : Under $\theta(g)$ of (7.2.5) <u>when a is known,whatever</u> , with the allocation of the sample size to strata given by

$$n \frac{\sqrt{(\sum_j x_{ij}^{g_0/2})(\sum_j x_{ij}^{g-(g_0/2)})}}{\sum_{i=1}^{k} \sqrt{(\sum_j x_{ij}^{g_0/2})(\sum_j x_{ij}^{g-(g_0/2)})}}$$

ratified strategy (stratified $p^*_{g_0/2} : \hat{Y}^*_S$) <u>for a fixed</u> <u>1,2]</u> is superior to the unstratified strategy ($p^*_{g_0/2} : \hat{Y}^*_U$) stimating Y for all $g \in [1,2]$ different from g_0. At , both the strategies are equivalent for estimating Y.

: Immediate from the proof of Theorem 7.3.2.

7.3.6 : Let us now consider situations when a is not , as happens often in practice.

Considering a stratified set up as in Remark 7.3.4, let each stratum a design $p^*_{g_0/2}$ with fixed sample size n_i ed for drawing the sample in which the inclusion probabilities are proportional to $x_{ij}^{g_o/2}$, g_o fixed $\varepsilon [1,2]$, i.e.,

$$\cdot n_i\, x_{ij}^{g_o/2} / (\sum_j x_{ij}^{g_o/2}) \quad \text{for} \quad j = 1,2,\dots,N_i; \quad i = 1,2,\dots,k.$$

pling be carried out independently in each stratum. The sizes n_i's are such that $\sum_i n_i = n$. Whatever be a and the model $\theta(g)$ of (7.2.5), as an estimator of the popula- otal $Y = \sum_i \sum_j y_{ij}$ consider

$$= \sum_{i=1}^{k} \{ \sum_j' y_{ij} + \frac{\sum_j' (\pi_{(i)j}^{-1} - 1) y_{ij}}{\sum_j' (\pi_{(i)j}^{-1} - 1) x_{ij}} (X_i - \sum_j' x_{ij}) \} \tag{7.3.9}$$

the stratified $p^*_{g_o/2}$ design discussed above. In (7.3.9), given i, $\sum_j'$ denotes summation over the sampled n_i units the i-th stratum for $i = 1,2,\dots,k$. Observe that $\hat{Y}^*_{B,st.}$ is obtained from $\hat{Y}^*_S$ in Remark 7.3.4 by estimating a using ithin each stratum (cf. Remark 7.1.1). Now, when the stratum N_i are large and Brewer's asymptotic analysis is carried ithin each stratum, the expected variance of the strategy tified $p^*_{g_o/2} : \hat{Y}^*_{B,st.}$) under $\theta(g)$ is given by (analogous .1.3))

$$\text{Var.}(\text{stratified } p^*_{g_o/2} : \hat{Y}^*_{B,st.}) \simeq \sigma^2 \sum_{i=1}^{k} \sum_{j=1}^{N_i} (\pi_{(i)j}^{-1} - 1) x_{ij}^{g} \tag{7.3.10}$$

Suppose now, a $p^*_{g_o/2}$ design (g_o fixed ε [1,2]) with d sample size n ($= \sum_i n_i$) is used for drawing a sample from whole (unstratified) population in which the inclusion probalities are $\pi'_{(i)j} = n\, x_{ij}^{g_o/2} / (\sum_i \sum_j x_{ij}^{g_o/2})$, $j = 1,2,\ldots,N_i$; 1,2,...,k. As an estimator of the population total Y sider

$$\hat{Y}_{Un.} = \sum_{i=1}^{k} \sum_j{}' y_{ij} + \frac{\sum_{i=1}^{k} \sum_j{}' (\pi'^{-1}_{(i)j} - 1) y_{ij}}{\sum_{i=1}^{k} \sum_j{}' (\pi'^{-1}_{(i)j} - 1) x_{ij}} (X - \sum_{i=1}^{k} \sum_j{}' x_{ij}) \qquad (7.3.11)$$

the $p^*_{g_o/2}$ design described above for drawing a sample of e n from the whole population. For a given i, $\sum_j{}'$ in (7.3.11) otes summation over those units out of the sampled n that ong to the i-th stratum for $i = 1,2,\ldots,k$. Note that $\hat{Y}_{Un.}$ of (7.3.11) above is obtained from $\hat{Y}^*_U$ in Remark 7.3.4 by imating a using $\hat{a}^*_B$ with the $p^*_{g_o/2}$ design (cf. Remark 7.1.1). n N is large, under the model $\theta(g)$ of (7.2.5), the mptotically $p^*_{g_o/2}$-unbiased estimator $\hat{Y}^*_{B,Un.}$ is such that

$$_{g)}\mathrm{Var.}(p^*_{g_o/2} : \hat{Y}^*_{B,Un.}) \simeq \sigma^2 \sum_{i=1}^{k} \sum_{j=1}^{N_i} (n^{-1} x_{ij}^{-g_o/2} (\sum_i \sum_j x_{ij}^{g_o/2}) - 1) x_{ij}^{g}, \qquad (7.3.12)$$

ich is obtained analogous to (7.1.3).

__orem 7.3.7__ : The strategy (stratified $p^*_{g_0/2}$: $\hat{Y}^*_{B,st.}$) with allocation $\underset{\sim}{n} = (n_1, n_2, \dots, n_k)$ of sample size to strata given

$$n_i = n \frac{\sqrt{(\sum_j x_{ij}^{g_0/2})(\sum_j x_{ij}^{g-(g_0/2)})}}{\sum_{i=1}^{k} \sqrt{(\sum_j x_{ij}^{g_0/2})(\sum_j x_{ij}^{g-(g_0/2)})}}$$

superior to the unstratified strategy ($p^*_{g_0/2}$: $\hat{Y}^*_{B,Un}$) for timating Y under $\Theta(g)$ for all values of $g \in [1,2]$ different om g_0. At $g = g_0$, both the strategies are equivalent for timating Y.

__oof__ : Follows from the proof of Theorem 7.3.2.

__mark 7.3.8__ : Results of sections 7.2 and 7.3 relate as follows our results of Chapter 2. Under a stratified set up, assuming e model $\Theta(g)$ of (7.2.5) we observe in Chapter 2 that, whatever the allocation $\underset{\sim}{n}$ of sample size to strata,

$$\varepsilon_{(g)}[\text{Var.}(\text{stratified } \pi PS : \hat{Y}_S) - \text{Var.}(\pi PS : \hat{Y}_U)] \geq 0, \qquad (7.3.13)$$

en $g = 2$. We also observed that equality occurs in (7.3.13) en and only when X-proportional allocation (allocation proport- nal to the stratum total X_i) is used. Considering a πPS design fixed sample size n_i in the i-th stratum for $i = 1,2,\ldots,k$ the corresponding Horvitz-Thompson estimator $\hat{Y}_S$ of the ulation total Y (which is obtained if we substitute $g_o = 2$ the stratified strategy (stratified $p^*_{g_o/2}$: $\hat{Y}^*_{B,st.}$)), T.J.Rao 68) derives the optimum allocation of the sample size to strata t minimizes the expected variance of the strategy (stratified : $\hat{Y}_S$) under $\theta(g)$ for $g \in [1,2]$ to be

$$n_i = n \sqrt{(\sum_j x_{ij})(\sum_j x_{ij}^{g-1})} \Big/ \sum_{i=1}^{k} \sqrt{(\sum_j x_{ij})(\sum_j x_{ij}^{g-1})}$$

r $i = 1,2,\ldots,k$, where $n = \sum_{i=1}^{k} n_i$ (cf. (2.1.7)).

serve that when $g = 2$, n_i above reduces to $n_i = n(X_i/X)$ $i = 1,2,\ldots,k$.

<u>lustrations</u> : We illustrate Theorem 7.3.7 considering 4 popula-ions data for which are provided in Appendix AIV, AV and AVI. For ch of the populations <u>g_o was fixed to be equal to 1.</u> The total mple size n was fixed to be either 9 or 12 for the populations nsidered. Either 3 or 4 strata were formed and the optimum allo-ation

$$n_i = n \sqrt{(\sum_j x_{ij}^{1/2})(\sum_j x_{ij}^{g-(1/2)})} \Big/ \sum_{i=1}^{k} \sqrt{(\sum_j x_{ij}^{1/2})(\sum_j x_{ij}^{g-(1/2)})} \qquad (7.3.14)$$

i = 1,2,...,k was computed for $g = 1.1(0.1)2.0$. With this
m allocation $\underset{\sim}{n}$, $E_{\theta(g)}$Var.(stratified $p^*_{1/2}$: $\hat{Y}^{*,opt.}_{B,st.}$)
valuated for $g = 1.1(0.1)2.0$. Table 7.3.9 to Table 7.3.12
de the optimum allocation $\underset{\sim}{n} = (n_1,n_2,\ldots,n_k)$ of the sample
to the strata (n_i for $i = 1,2,\ldots,k$ evaluated from 7.3.14)
$E_{\theta(g)}$Var.(stratified $p^*_{1/2}$: $\hat{Y}^{*,opt.}_{B,st.}$)/$E_{\theta(g)}$Var.($p^*_{1/2}$: $\hat{Y}^{*}_{B,Un.}$)
$g = 1.1(0.1)2.0$.

ation 1 : (data provided in the Appendix AIV).

The population of 43 Kraals (Yates (1960), p.159) was
ied into 4 strata for illustration containing Kraals 1-10,
s 11-22, Kraals 23-35, Kraals 36-43. Total number of persons
uding absentees) was the auxiliary information $\mathcal{X}$. The
um sizes here are $N_1 = 10$, $N_2 = 12$, $N_3 = 13$ and $N_4 = 8$.
overall sample size was fixed as $n = 12$ for illustration.

Table 7.3.9

Population given in Yates (1960 , p. 159).

optimum allocation of sample size	$\frac{E_{\theta(g)} \text{Var.}(\text{stratified } p^*_{1/2} : \hat{Y}^{*,\text{opt.}}_{B,\text{st.}})}{E_{\theta(g)} \text{Var.}(p^*_{1/2} : \hat{Y}^*_{B,\text{Un.}})}$
(3,3,4,2)	0.9999
(3,3,4,2)	0.9999
(3,3,4,2)	0.9999
(3,3,4,2)	0.9998
(3,3,4,2)	0.9997
(3,3,4,2)	0.9996
(3,3,4,2)	0.9996
(3,3,4,2)	0.9994
(3,3,4,2)	0.9993
(3,3,4,2)	0.9992

ation 2 : (data provided in the Appendix AV). Eye estimated e of timber, $\mathcal{X}$ (cu. ft./ $\frac{1}{10}$ acre), was considered on 25 plots s (1960), p.163). The population was divided for illustration 3 strata containing plots 1-8, plots 9-18, plots 19-25. Thus stratum sizes are $N_1=8$, $N_2=10$ and $N_3 = 7$. The total sample was fixed as $n = 9$ for illustration.

Table 7.3.10

Population given in Yates (1950, p.163).

g	optimum allocation of sample size	$\frac{E_{\theta(g)} Var.(\text{stratified } p^*_{1/2} : \hat{Y}^{*,opt.}_{B,st.})}{E_{\theta(g)} Var.(p^*_{1/2} : \hat{Y}^*_{B,Un.})}$
1.1	(2,4,3)	0.9997
1.2	(2,4,3)	0.9990
1.3	(2,4,3)	0.9979
1.4	(2,4,3)	0.9965
1.5	(2,4,3)	0.9948
1.6	(2,4,3)	0.9927
1.7	(2,4,3)	0.9904
1.8	(2,4,3)	0.9879
1.9	(2,4,3)	0.9852
2.0	(2,4,3)	0.9823

Population 3 and Population 4 : (data provided in the Appendix AVI).

Area under wheat in 1936 ($\mathcal{Y}$ in acres) and total cultivated area in 1931 ($\mathcal{X}$ in acres) for 34 villages in Lucknow subdivision in Uttar Pradesh State of India (Sukhatme (1954), p.183) was considered. Three strata were formed consisting of villages 1-12, villages 13-23, villages 24-34. The stratum sizes were therefore $N_1=12$, $N_2=11$ and $N_3=11$. The total sample size for illustration was fixed to be $n=9$. Table 7.3.11 and Table 7.3.12 below provide the efficiency parisons for population 3 ($\mathcal{X}$ being area under wheat in 1936) population 4 ($\mathcal{X}$ being total cultivated area in 1931) spectively.

Table 7.3.11

Population given in Sukhatme (1954 , p.183). $\mathcal{X}$ being area under wheat in 1936.

	optimum allocation of sample size	$\frac{E_{\theta(g)} \text{Var.}(\text{stratified } p^*_{1/2} : \hat{Y}^{*,opt.}_{B,st.})}{E_{\theta(g)} \text{Var.}(p^*_{1/2} : \hat{Y}^{*}_{B,Un.})}$
1	(3,3,3)	0.9999
2	(3,3,3)	0.9996
3	(4,2,3)	0.9992
4	(4,2,3)	0.9987
.5	(4,2,3)	0.9981
6	(4,3,2)	0.9976
.7	(4,3,2)	0.9970
8	(4,3,2)	0.9964
.9	(4,3,2)	0.9959
0	(4,3,2)	0.9954

Table 7.3.12

Population given in Sukhatme (1954, p.183). $\mathcal{X}$ being total cultivated area in 1931.

optimum allocation of sample size	$\dfrac{E_{\theta(g)} \text{Var.}(\text{stratified } p^*_{1/2} : \hat{Y}^{*,opt.}_{B,st.})}{E_{\theta(g)} \text{Var.}(p^*_{1/2} : \hat{Y}^*_{B,Un.})}$
(3,3,3)	0.9998
(3,3,3)	0.9992
(3,3,3)	0.9984
(4,2,3)	0.9972
(4,3,2)	0.9957
(4,3,2)	0.9940
(4,3,2)	0.9919
(4,3,2)	0.9896
(4,3,2)	0.9870
(4,3,2)	0.9841

The 4 populations chosen for our illustration in section above were among the 16 populations considered by J.N.K.Rao (1969) and Royall (1970). J.N.K.Rao, restricting attention to simple random sampling without replacement plans, considered the ratio estimator, which is p-biased, along with seven other ratio type estimators, some of which are p-unbiased, the classical regression estimator and a p-unbiased regression type estimator, empirically compares the performance of each of the strategies.

chose to estimate the ratio Y/X. He computed the sampling an square error of each of these estimators, for various sample zes, ranging from $n = 2$ to $n = 12$ for each of the 16 popula-tions considered. Royall (1970) predicts that for all the 16 opulations, a plausible model, apriori, is $\Theta(g)$ with $g = 1$. ince the classical ratio estimator ($\hat{Y}_1/X$) is the best ' $\Theta(1)$ - nbiased' linear estimator of (Y/X) we have that ($\hat{Y}_1/X$) cannot e improved upon, with respect to expected mean square error, hatever be the sample size or sampling plan. Royall's prediction is generally concordant with J.N.K. Rao's results, which suggest that, with regard to sampling mean square error, three estimators, one of which is the ratio estimator, appear to be generally superior to the rest. J.N.K.Rao (1969, p.222) states that the choice is not clearcut between these three estimators. Royall (1970) considers the purposive sampling plan which selects the largest n units for fixed n, evaluates for each of the 16 populations and each fixed sample size $n = 2$ to 12 the ratio

$$\frac{\text{minimum sampling mean square error among simple random sample procedures}}{\text{squared error of ratio estimator with purposive sampling}}$$

serves that the strategy (purposive sampling : $\hat{Y}_1/X$) is
iently *robust* to maintain better performance than the
tional (SRSWOR : $\hat{Y}_1/X$) strategy under the variety of condi-
represented by the 16 populations.

We considered 4 out of these 16 populations and fixed
= 1 since, apriori, a plausible model is $\theta(1)$ for
populations in view of the above comments.

A P P E N D I X

APPENDIX

A I

Recorded acreage of crops and grass for 1947, and acreage under oats in 1957, for 35 farms in Orkney, extracted from the records of the Scottish Fertiliser Practice Survey, 1957.

(Source : Sampford (1962, p.61))

Farm number	Recorded crops & grass (x)	Oats 1957 (y)	Farm number	Recorded crops & grass (x)	Oats 1957 (y)
(1)	(2)	(3)	(1)	(2)	(3)
1	50	17	18	110	24
2	50	17	19	140	43
3	52	10	20	140	48
4	58	16	21	156	44
5	60	6	22	156	45
6	60	15	23	190	60
7	62	20	24	198	63
8	65	18	25	209	70
9	65	14	26	240	28
10	68	20	27	274	62
11	71	24	28	300	59
12	74	18	29	303	66
13	78	23	30	311	58
14	90	0	31	324	128
15	91	27	32	330	38
16	92	34	33	356	69
17	96	25	34	410	72
			35	430	103
			Total:	5759	1384

A II

Data from the Census of India 1971 document consisting of 1961 Census population (x) and 1971 Census population (y) for 142 cities/urban agglomerations of India with 1971 population size 100,000 and above.

(Figures in '00s)

Sl. no.	Name of the city/urban agglomeration	1961 population (no. of persons)	1971 population (no. of persons)
(1)	(2)	(3)	(4)
1	Hyderabad	12490	17989
2	Visakhapatnam	2112	3623
3	Vizayawada	2344	3437
4	Guntur	1871	2699
5	Warangal	1561	2071
6	Rajahmundray	1300	1888
7	Kakinada	1229	1642
8	Kurnool	1008	1367
9	Nellore	1068	1336
10	Eluru	1083	1270
11	Nizamabad	791	1149
12	Machilipatnam	1014	1126
13	Tenali	785	1029
14	Gauhati	1007	1230
15	Patna	3646	4903
16	Jamshedpur	3280	4652
17	Dhanbad	2006	4331
18	Ranchi	1402	2560
19	Gaya	1511	1798
20	Bhagalpur	1438	1727

Contd.

A II (Continued)

(1)	(2)	(3)	(4)
21	Darbhanga	1030	1321
22	Muzaffarpur	1091	1270
23	Bokaro Steel City	751*	1080
24	Monghyr	898	1025
25	Bihar	786	1001
26	Ahmedabad	11499	15884
27	Surat	2880	4718
28	Baroda	2984	4674
29	Rajkot	1941	3002
30	Bhavnagar	1765	2261
31	Jamnagar	1486	2149
32	Nadiad	790	1083
33	Rohtak	882	1248
34	Ambala Cantt.	1055	1025
35	Srinagar	2853	4036
36	Jammu	1027	1552
37	Cochin	2807	4384
38	Trivandrum	2398	4098
39	Calicut	1925	3340
40	Alleppey	1388	1601
41	Quilon	910	1241
42	Indore	3949	5726
43	Jabalpur	3670	5338
44	Gwalior	3006	4068
45	Bhopal	2229	3921
46	Durg-Bhilainagar	1332	2453
47	Ujjain	1442	2091

Contd.

A II (Continued)

(1)	(2)	(3)	(4)
48	Raipur	1398	2059
49	Sagar	1047	1548
50	Bilaspur	867	1308
51	Ratlam	875	1186
52	Burhanpur	821	1053
53	Greater Bombay	41520	59685
54	Nagpur	6436	8661
55	Poona	5976	8532
56	Sholapur	3376	3981
57	Kolhapur	1874	2591
58	Amravati	1379	1936
59	Malegaon	1214	1918
60	Nasik	1311	1762
61	Thana	1011	1702
62	Akola	1158	1685
63	Ulhasnagar	1078	1681
64	Aurangabad	876	1505
65	Dhulia	989	1371
66	Nanded	811	1264
67	Ahmednagar	970	1173
68	Sangli	738	1151
69	Jalgaon	804	1067
70	Bangalore	11526	16482
71	Hubli-Dharwar	2485	3796
72	Mysore	2539	3556
73	Mangalore	1745	2141
74	Belgaum	1468	2138
75	Gulbarga	971	1456

Contd

A II (Continued)

(1)	(2)	(3)	(4)
76	Bellary	857	1251
77	Devanagere	781	1210
78	Bijapur	789	1030
79	Shimoga	638	1027
80	Bhadravathi	658	1013
81	Cuttack	1463	1940
82	Rourkela	903	1725
83	Berhampur	769	1176
84	Bhubaneswar	382	1055
85	Amritsar	3763	4327
86	Ludhiana	2440	4011
87	Jullundur	2226	2961
88	Patiala	1252	1519
89	Jaipur	4034	6131
90	Jodhpur	2248	3189
91	Ajmer	2312	2625
92	Kota	1203	2130
93	Bikaner	1506	1886
94	Udaipur	1111	1629
95	Alwar	727	1008
96	Madras City	17292	24703
97	Madurai	4248	5483
98	Coimbatore	2863	3535
99	Salem	2492	3083
100	Tiruchirapalli	2499	3062
101	Tuticorin	1242	1548
102	Nagercoil	1062	1412
103	Thanjavur	1111	1405

Contd.

A II (Continued)

(1)	(2)	(3)	(4)
104	Vellore	1137	1382
105	Dindigul	929	1274
106	Singanallur	247	1134
107	Tiruppur	798	1132
108	Kumbakonam	926	1130
109	Kanchipuram	927	1105
110	Tirunelveli	880	1085
111	Erode	738	1037
112	Cuddalore	792	1013
113	Kanpur	9710	12730
114	Lucknow	6557	8262
115	Agra	5087	6378
116	Varanasi	4898	5829
117	Allahabad	4307	5140
118	Meerut	2840	3678
119	Bareilly	2728	3261
120	Moradabad	1918	2724
121	Aligarh	1850	2540
122	Gorakhpur	1802	2307
123	Saharanpur	1852	2257
124	Dehra Dun	1563	1994
125	Jhansi	1697	1981
126	Rampur	1354	1618
127	Shahjahanpur	1177	1441
128	Mathura	1253	1405
129	Firozabad	986	1339
130	Ghaziabad	704	1280

Contd.

A II (Continued)

(1)	(2)	(3)	(4)
131	Muzaffarnagar	876	1149
132	Farrukhabad-cum-Fatehgarh	946	1114
133	Faizabad	883	1098
134	Mirzapur-cum-Vindhyachal	1001	1059
135	Calcutta	57369	70054
136	Durgapur	417	2072
137	Kharagpur	1473	1619
138	Asansol	1034	1574
139	Burdwan	1082	1450
140	Chandigarh	993	2330
141	Delhi	23593	36298
142	Imphal	677	1006

* Bokaro Steel City came into existence only after 1961. To ensure uniformity, a hypothetical 1961 population is estimated based on growth rate in other cities of Bihar State in which Bokaro is situated.

A III

Live example consisting of the district-wise population figures of four States in India. The four States were treated as four strata.

Stratum 1

Sl. no.	Name of the district	1951 census population (rounded off to thousands) $\mathcal{X}$	1961 census population $\mathcal{Y}$
(1)	(2)	(3)	(4)
1	Cannanore	1375	1,780,294
2	Kozhikode	2065	2,617,189
3	Palghat	1565	1,776,566
4	Trichur	1363	1,639,862
5	Ernakulam	1530	1,859,913
6	Kottayam	1328	1,732,880
7	Alleppey	1521	1,811,252
8	Quilon	1474	1,941,228
9	Trivandrum	1328	1,744,531
	Total	13549	16,903,715

A III (Continued)

Stratum 2

Sl. no.	Name of the district	1951 census population (rounded off to thousands) $\mathcal{X}$	1961 census population $\mathcal{Y}$
(1)	(2)	(3)	(4)
1	Srikakulam	2123	2,340,878
2	Visakhapatnam	2073	2,290,759
3	East Godavari	2302	2,608,375
4	West Godavari	1698	1,978,257
5	Krishna	1736	2,076,956
6	Guntur	2560	3,009,900
7	Nellore	1795	2,033,679
8	Chittor	1666	1,914,639
9	Cuddapah	1163	1,342,015
10	Anantapur	1484	1,767,464
11	Kurnool	1617	1,908,740
12	Mahbubnagar	1447	1,590,686
13	Hyderabad	1822	2,062,995
14	Medak	1110	1,227,361
15	Nizamabad	835	1,022,013
16	Adilabad	832	1,009,292
17	Karimnagar	1428	1,621,515
18	Warangal	1330	1,545,435
19	Khammam	808	1,057,542
20	Nalgonda	1287	1,574,946
	Total	31,116	35,983,447

A III (Continued)

Stratum 3

Sl. no.	Name of the district	1951 census population (rounded off to thousands) X	1961 census population Y
(1)	(2)	(3)	(4)
1	Jamnagar	617	828,419
2	Rajkot	930	1,208,519
3	Surendranagar	506	663,206
4	Bhavnagar	886	1,119,435
5	Amreli	539	667,823
6	Junagadh	988	1,245,643
7	Kutch	568	696,440
8	Banaskantha	774	996,144
9	Sabarkantha	684	918,587
10	Mehsana	1394	1,689,963
11	Ahmedabad	1676	2,210,199
12	Kaira	1612	1,977,540
13	Panchmahals	1131	1,468,946
14	Baroda	1212	1,527,326
15	Broach	718	891,969
16	Surat	1982	2,451,624
17	Dangs	47	71,567
	Total	16,264	20,633,350

A III (Continued)

Stratum 4

Sl. no.	Name of the district	1951 census population (rounded off to thousands) x	1961 census population y
(1)	(2)	(3)	(4)
1	Darjeeling	460	624,640
2	Jalpaiguri	915	1,359,292
3	Cooch Behar	671	1,019,806
4	West Dinajpur	979	1,323,797
5	Malda	938	1,221,923
6	Murshidabad	1716	2,290,010
7	Nadia	1145	1,713,324
8	24-Parganas	4459	6,280,915
9	Calcutta	2698	2,927,289
10	Howrah	1611	2,038,477
11	Hooghly	1604	2,231,418
12	Burdwan	2192	3,082,846
13	Birbhum	1067	1,446,158
14	Bankura	1319	1,664,513
15	Midnapur	3359	4,341,855
16	Purulia	1169	1,360,016
	Total	26,302	34,926,279

A IV

Data from a random sample of 43 Kraals : Total number of persons (including absentees), $\mathcal{X}$, and number of absentees, $\mathcal{Y}$.

(Source : Yates (1960) p.159))

Sl. no.	$\mathcal{X}$	$\mathcal{Y}$	Sl. no.	$\mathcal{X}$	$\mathcal{Y}$
1	95	18	22	64	12
2	79	14	23	75	12
3	30	6	24	69	16
4	45	3	25	63	9
5	28	5	26	83	14
6	142	15	27	124	25
7	125	18	28	31	3
8	81	9	29	96	45
9	43	12	30	42	25
10	53	4	31	85	35
11	148	31	32	91	28
12	89	7	33	73	13
13	57	9	34	159	36
14	132	26	35	54	26
15	47	7	36	69	27
16	43	17	37	61	2
17	116	24	38	164	69
18	65	16	39	132	41
19	103	18	40	82	10
20	52	16	41	33	8
21	67	27	42	86	22
			43	51	19
			Total	3427	799

A V

Measured volume of timber, $\mathcal{Y}$, on 25 sample plots and Eye estimated volume of timber, $\mathcal{X}$, of corresponding stands (cu. ft. per 1/10 acre).

(Source : Yates (1960), p. 163)

Sl. no.	$\mathcal{X}$	$\mathcal{Y}$	Sl. no.	$\mathcal{X}$	$\mathcal{Y}$
1	102	170	14	128	112
2	14	47	15	79	153
3	57	64	16	177	216
4	70	91	17	65	125
5	95	126	18	196	100
6	92	146	19	167	287
7	110	87	20	268	261
8	208	195	21	152	169
9	208	255	22	152	182
10	110	135	23	148	74
11	110	146	24	207	24
12	110	154	25	167	255
13	110	110	Total	3302	3684

A VI

Values of total cultivated area and of area under wheat in two consequtive years for a sample of 34 villages in Lucknow Subdivision (India).

(Source : Sukhatme (1954), p. 183)

Sl. no. of village	Total cultivated area in 1931 (in acres)	Area under wheat 1936 (in acres)	Area under wheat 1937 (in acres)
(1)	(2)	(3)	(4)
1	401	75	52
2	634	163	149
3	1194	326	289
4	1770	442	381
5	1060	254	278
6	827	125	111
7	1737	559	634
8	1060	254	278
9	360	101	112
10	946	359	355
11	470	109	99
12	1625	481	498
13	827	125	111
14	96	5	6
15	1304	427	399

Contd.

A VI (Continued)

(1)	(2)	(3)	(4)
16	377	78	79
17	259	78	105
18	186	45	27
19	1767	564	515
20	604	238	249
21	701	92	85
22	524	247	221
23	571	134	133
24	962	131	144
25	407	129	103
26	715	192	179
27	845	663	330
28	1016	236	219
29	184	73	62
30	282	62	79
31	194	71	60
32	439	137	100
33	854	196	141
34	824	255	261

REFERENCES

Basu, D. (1958). On sampling with and without replacement. Sankhyā, 20, 287-294.

________ (1970). An essay on the logical foundations of survey sampling, Part one. Symposium on the Foundations of Statistical Inference, Waterloo, Canada.

Brewer, K.R.W. and Undy, G.C. (1962). Samples of two units drawn with unequal probabilities without replacement. Austral. J. Statist., 4, 89-100.

Brewer, K.R.W. (1963). Ratio estimation and finite populations. Austral. J. Statist., 5, 93-105.

________ (1963a). A model of systematic sampling with unequal probabilities. Austral. J. Statist., 5, 5-13.

________ (1975). A simple procedure for sampling πPSwor. Austral. J. Statist., 17, 166-172.

________ (1977). A class of robust sampling designs for large scale surveys (to appear). Preprint by courtesy of the author.

Brewer, K.R.W., Foreman, E.K., Mellor, R.W. and Trewin, D.J. (1977). Use of experimental design and population modelling in survey sampling. Presented at the 41st session of the International Statistical Institute held at New Delhi.

Cassel, C.M., Särndal, C.E. and Wretman, J.H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. Biometrika, 63, 615-620.

Chaudhuri, A. (1975). A simple method of sampling without replacement with inclusion probabilities exactly proportional to size. _Metrika_, _22_, 147-152.

Cochran, W.G. (1942). Sampling theory when the sampling units are of unequal sizes. _J. Amer. Statist. Assoc._, _37_, 199-212.

________ (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. _Ann. Math. Statist._, _17_, 164-177.

________ (1947). Recent developments in sampling theory in the United States. _Rev. Int. Stat. Inst._, _3_, 40-66.

________ (1963). _Sampling Techniques._ New York : Wiley.

Dalenius, T. (1962). Recent advances in sample survey theory and methods. _Ann. Math. Statist._, _33_, 325-349.

Das, M.N. and Mohanty, S. (1973). On PPS sampling without replacement ensuring selection probabilities exactly proportional to sizes. _Austral. J. Statist._, _15_, 87-94.

Des Raj (1956). Some estimators in sampling with varying probabilities without replacement. _J. Amer. Statist. Assoc._, _51_, 269-284.

______ (1956a). A note on the determination of optimum probabilities in sampling without replacement. Sankhyā, _17_, 197-200.

es Raj and Khamis, S.H. (1958). Some remarks on sampling with replacement. Ann. Math. Statist., 39, 550-557.

es Raj (1963). A note on stratification in unequal probability sampling. Sankhyā Ser. B, 25, 215-216.

____ (1968). Sampling Theory, Mcgraw-Hill, New York.

odds, D.J. and Fryer, J.G. (1971). Some families of selection probabilities for sampling with probability proportional to size. J. Roy. Statist. Soc. Ser. B, 33, 263-274.

Durbin, J. (1953). Some results in sampling theory when the units are selected with unequal probabilities. J. Roy. Statist. Soc. Ser. B, 15, 262-269.

Fellegi, I.P. (1963). Sampling with varying probabilities without replacement : Rotating and nonrotating samples. J. Amer. Statist. Assoc., 58, 183-201.

Foreman, E.K. and Brewer, K.R.W. (1971). The efficient use of supplementary information in standard sampling procedures. J. Roy. Statist. Soc. Ser. B, 33, 391-400.

Godambe, V.P. (1955). A unified theory of sampling from finite populations. J. Roy. Statist. Soc. Ser. B, 17, 269-278.

____ and Joshi, V.M. (1965). Admissibility and Bayes estimation in sampling finite populations - I. Ann. Math. Statist., 36, 1707-1722.

Godambe, V.P. (1966). A new approach to sampling from finite populations. J. Roy. Statist. Soc. Ser. B, 28, 310-328.

Goodman, R. and Kish, L. (1950). Controlled selection — a technique in probability sampling. J. Amer. Statist. Assoc., 45, 350-372.

Grundy, P.M. (1954). A method of sampling with probability exactly proportional to size. J. Roy. Statist. Soc. Ser. B, 16, 236-238.

Hájek, J. (1959). Optimum strategy and other problems in probability sampling. Časopis Pěst. Mat., 84, 387-428.

_______ (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. Ann. Math. Statist., 35, 1491-1523.

Hansen, M.H. and Hurwitz, W.N. (1943). On the theory of sampling from finite populations. Ann. Math. Statist., 14, 333-362.

_______ (1949). On the determination of optimum probabilities in sampling. Ann. Math. Statist., 20, 426-432.

Hanurav, T.V. (1962). An existence theorem in sampling. Sankhyā Ser. A, 24, 227-230.

_______ (1962a). On Horvitz-Thompson estimator. Sankhyā Ser. A, 24, 429-436.

Hanurav, T.V. (1962b). Some sampling schemes in probability sampling. Sankhyā Ser. A, 24, 421-428.

________ (1965). Optimum Sampling Strategies and Some Related Problems. Ph.D. thesis, Indian Statistical Institute, Calcutta.

________ (1966). Some aspects of unified sampling theory. Sankhyā Ser. A, 28, 175-204.

________ (1966a). An optimum πPS sampling, (an abstract). Ann. Math. Statist., 37, 1859.

________ (1967). Optimum utilization of auxiliary information —πPS sampling of two units from a stratum. J. Roy. Statist. Soc. Ser. B, 29, 374-391.

Hartley, H.O. and Rao, J.N.K. (1962). Sampling with unequal probabilities and without replacement. Ann. Math. Statist., 33, 350-374.

Ho, E.W.H. (1977). Private communication.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. J. Amer. Statist. Assoc., 47, 663-685.

Jessen, R.J. (1942). Statistical investigation of a sample survey for obtaining farm facts. Iowa Agr. Exp. Sta. Res. Bull., No. 304.

Joshi, V.M. (1965a). Admissibility and Bayes estimation in sampling finite populations - II. Ann. Math. Statist., 36, 1723-1729.

__________ (1965b). Admissibility and Bayes estimation in sampling finite populations - III. Ann. Math. Statist., 36, 1730-1742.

Mahalanobis, P.C. (1944). On large scale sample surveys. Philos. Trans. Roy. Soc. Ser.B. 231, 329-451.

_______ (1946). Recent experiment in statistical sampling in the Indian Statistical Institute. J. Roy. Statist. Soc., 109, 325-378.

Murthy, M.N. (1957). Ordered and unordered estimators in sampling without replacement. Sankhyā, 18, 379-390.

_______ (1963). Some recent advances in sampling theory. J. Amer. Statist. Assoc., 58, 737-755.

Neyman, J. (1934). On two different aspects of the representative method, the method of stratified sampling and the method of purposive selection. J. Roy. Statist. Soc., 97, 558-606.

Prabhu Ajgaonkar, S.G. (1965). On a class of linear estimators in sampling with varying probabilities without replacement. J. Amer. Statist. Assoc., 60, 637-642.

Ramachandran, G. and Rao, T.J. (1974). Allocation to strata and relative efficiencies of stratified and unstratified πPS sampling schemes. J. Roy. Statist. Soc. Ser. B, 36, 292-298.

______ (1977). Probability proportional to size sampling strategy and stratification. Bull. Int. Stat. Inst., 47, Book 4, 410-412. Presented at the 41st session of the International Statistical Institute held at New Delhi.

Rao, J.N.K., Hartley, H.O. and Cochran, W.G. (1962). On a simple procedure of unequal probability sampling without replacement. J.Roy.Statist. Soc. Ser. B, 27, 482-491.

Rao, J.N.K. (1963). On three procedures of unequal probability sampling without replacement. J. Amer. Statist. Assoc., 58, 202-215.

________ (1965). On two simple schemes of unequal probability sampling without replacement. J. Indian Statist. Assoc., 3, 173-180.

________ (1966). On the relative efficiency of some estimators in PPS sampling for multiple characteristics. Sankhyā, Ser.A, 28, 61-70.

________ (1969). Ratio and regression estimators. In New Developments in Survey Sampling, eds. N.L.Johnson and H. Smith. New York : Wiley, 213-234.

Rao, T.J. (1968). On the allocation of sample size in stratified sampling. Ann. Inst. Statist. Math., 18, 117-121.

_______ (1971). πPS sampling designs and Horvitz-Thompson estimator. J.Amer.Statist. Assoc., 66, 872-875.

_______ (1972). Horvitz-Thompson and Des Raj estimators revisited. Austral. J. Statist., 14, 227-230.

Roy, J. and Chakravarty, I.M. (1960). Estimating the mean of a finite population. Ann. Math. Statist., 31, 392-398.

Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. Biometrika, 57, 377-387.

Royall, R.M. and Herson, J. (1973). Robust estimation in finite populations I. J. Amer. Statist. Assoc., 68, 880-889.

_________________ (1973a). Robust estimation in finite populations II : Stratification on a size variable. J. Amer. Statist. Assoc., 68, 890-893.

Sampford, M.R. (1962). An Introduction to Sampling Theory with Applications to Agriculture. Edinburgh : Oliver and Boyd.

_______ (1967). On sampling without replacement with unequal probabilities. Biometrika, 54, 499-513.

Sankaranarayanan, K. (1969). An IPPS sampling scheme using Lahiri's method of selection. J. Indian Soc. Agric. Statist., 21, 58-66.

Seng, Y.P. (1951). Historical survey of the development of sampling theories and practice. J. Roy. Statist. Soc. Ser.A, 114, 214-231.

Smith, H.F. (1938). An empirical law governing soil heterogeneity. J. Agr. Sci., 28, 1-23.

Smith, T.M.F. (1976). Comments on paper by Cassel, Särndal and Wretman. Biometrika, 63, 620.

________ (1976a). The foundations of survey sampling : a review (with discussion). J. Roy. Statist. Soc. Ser.A, 139, 183-204.

Stephen, F.F. (1948). History of the use of modern sampling procedures. J. Amer. Statist. Assoc., 43, 12-39.

Stuart, A. (1963). Some remarks on sampling with unequal probabilities. Bull. Int. Stat. Inst., 40, 773-779.

________ (1964). Multistage sampling with preliminary random stratification of first stage units. Rev. Int. Stat. Inst., 32, 193-201.

Sukhatme, P.V. (1954). Sampling Theory of Survey with Applications. Iowa State College Press.

___________ (1959). Major developments in the theory and application of sampling during the last twenty-five years. Estadistica, 17, 652-679.

Sukhatme, P.V. (1966). Major developments in sampling theory and practice. Research papers in Statistics, Festschrift for J. Neyman, Wiley, New York, 367-409.

Vijayan, K. (1966). On Horvitz-Thompson and Des Raj estimators. Sankhyā Ser.A, 28, 87-92.

__________ (1968). On an exact πPS sampling scheme - generalization of a method of Hanurav. J.Roy. Statist. Soc. Ser. B, 30, 556-566.

Vos, J.W.E. (1974). Steekproeftrekking met ongelijke Kansen. Statistica Neerlandica, 28, Part I, 11-49, Part II, 69-108.

Yates, F. and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. J. Roy. Statist. Soc. Ser. B, 15, 253-261.

Yates, F. (1946). A review of recent statistical developments in sampling and sampling surveys. J. Roy. Statist. Soc., 109, 12-43.

_______ (1960). Sampling Methods for Censuses and Surveys. 3rd edition. London : Griffin.