1982

CONTRIBUTIONS TO SAMPLING THEORY AND PRACTICE USING AUXILIARY INFORMATION (INDIA).

TALAPADY. SRIVENKATARAMANA
University of Windsor

CANADIAN THESES ON MICROFICHE

I.S.B.N.

THESES CANADIENNES SUR MICROFICHE

National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

NL-339 (r. 82/08)

Canada

CONTRIBUTIONS TO SAMPLING THEORY AND PRACTICE USING AUXILIARY INFORMATION

by

© Talapady Srivenkataramana

A Dissertation

Submitted to the Faculty of Graduate Studies through the

Department of Mathematics in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy at the

University of Windsor

Windsor, Ontario, Canada

1982

Dedicated to my parents

Shri T. Madhava Bhat and Shrimati T. Indiramma

ACKNOWLEDGEMENTS

The author is greatly indebted to Dr. Derrick S. Tracy for his invaluable guidance in the research work carried out. There was always a personal touch, encouragement and affection in the relationship, besides academic guidance. The benefit derived extends much beyond the preparation of this dissertation.

Thanks are due to the Bangalore University for giving leave of absence and the University of Windsor for the financial support during the work. The author is particularly grateful to the International Development Research Centre of Canada for the Thesis Research Award during 1979-82, which supported the survey work reported in Part Two of the thesis.

Special thanks are due to Mr. Narasimha Prasad for many useful discussions regarding the material in Chapters 5 and 6 and to Mr. K. Harish Chandra for the help in the field-work and computer analysis of the survey data. The author wishes to express his appreciation of the patience exhibited by his wife and children during the entire programme. The great encouragement given by his parents, brothers and sisters instilled the necessary confidence. The help obtained from the members of the Mathematics Department at the University of Windsor and, in particular, its Chairman Dr. H.R. Atkinson are sincerely acknowledged.

ABSTRACT

Part One (Chapters 2-7) of the thesis illustrates mainly the use of auxiliary information in effecting certain transformations for improving the design/estimation strategies in sample surveys. A new product-type estimator which is complementary to the usual ratio estimator, in a certain sense, is proposed in Chapter 2. A transformation which permits the use of the product method in place of the more commonly used ratio method is discussed in the next chapter. One resulting convenience is that the bias and mean squared error of the estimator have closed form expressions, unlike that for the ratio estimator. Chapter 4 develops a change of origin for the study variate in order to improve the precision of the estimates under varying probability sampling. A change of origin under ratio method of estimation and a change of scale under the difference method are also examined. Raj (1965) has proposed a two-phase pps selection scheme in the absence of information on the auxiliary variate, but when information on some other variate is available. This also requires the knowledge of a certain parameter. Chapter 5 suggests an alternative two-phase procedure not requiring the knowledge of this parameter.

Auxiliary variates in surveys are occasionally positive and negative valued. This introduces difficulties in ratio and product methods of estimation and in pps sampling. Chapter 6 outlines two simple methods of dealing with the situation: stratification by sign of the auxiliary variate and transformation by simple translation. For the purposes of illustration simple random sampling without replacement and probability

proportional to size sampling with replacement are assumed. Finally, the scope for some future work is indicated in Chapter 7.

Part Two (Chapters 8-9) of the thesis is a brief report on a sample survey conducted during 1980-81 to assess the transition of rural people to modern ways of life in the Karnataka State in South India.

C O N T E N T S

CHAPTER - 1

GENERAL INTRODUCTION

In sample surveys emphasis is laid on the use of supplementary information for improving the precision of estimators. Such information can be incorporated either into the sampling scheme as in stratified or pps sampling, or into the estimation procedure as in the ratio and product methods or in both. Vast literature is available in standard texts and research journals in this regard. Tremendous advances have been made in methods of sampling, formation of estimators and evaluating their performance in repeated sampling which includes computer studies when the situation is difficult to be tackled directly. Attention has been paid to reducing the bias of the estimators, obtaining stable variance estimators and evolving methods which are simple for application. Among the milestones are the papers by Neyman (1934), Cochran (1942), Hansen and Hurwitz (1943), and Horvitz and Thompson (1952).

To most people in sample surveys Neyman's 1934 paper is remembered for the derivation of the formula for optimal allocation in stratified sampling. But of far greater importance to statistics in general and to sample survey theory in particular is Neyman's exposition of the logic of inference based on confidence intervals. Ratio and regression methods of estimation were introduced during the 1930s with a comprehensive account of the theory being provided by Cochran (1942). Hansen and Hurwitz (1943) introduced probability proportional to size sampling as an efficient way of

carrying out multi-stage sampling. Horvitz and Thompson (1952) provided an elegant treatment of variable probability sampling and also gave the impetus for new research at the foundations level.

Auxiliary information usually provides a basis for stratification, gives a measure of size or helps to form ratio, product or regression-type estimators. Sometimes such information is adequate for assessing the values of certain parameters or asserting that they lie in specified intervals like (1/2, 1) or (1,2). These parameters may be built into the estimators or the idea about the magnitudes of the parameters may be used to choose from a family of estimators. A number of papers have been published in this regard.

Auxiliary information may also be used for effecting certain transformations. For instance, transformations which (i) enable the use of a specific method of estimation rather than another, or (ii) make the situation satisfy certain assumptions more closely than otherwise. If these transformations are simple, they will also be well suited for large scale applications. In this context a few methods have been proposed by Srivenkataramana (1978,1980), and Srivenkataramana and Tracy (1979, 1980a,b,c). The present thesis is in the form of some contributions to sampling theory and practice using auxiliary information. Part One, Theory, compactly strings together the results in the above papers by Srivenkataramana and Tracy after filling up gaps here and there. It then tackles two new aspects. Part Two, Practice, is a brief report on an actual sample survey conducted in the villages of Karnataka State (India) during 1980-81. A brief outline of the thesis is as follows.

Chapter 2 proposes a new product-type estimator which is complementary to the common ratio estimator, for small sampling fractions.

The next chapter points out that the product method considered by Robson (1957) and Murthy (1964) has the advantage that the bias and mse of the estimator can be evaluated exactly, unlike in the case of the ratio estimator. Motivated by this, a transformation which permits the product method in place of the ratio method is discussed. Chapter 4 illustrates a few simple transformations of data after they have been collected, in order to improve the precision of estimates. Chapter 5 considers double sampling with pps selection. An alternative to the Raj (1965) scheme is suggested. The less commonly discussed situation of using positive and negative valued auxiliary variates in surveys is taken up in Chapter 6. Two methods of overcoming the associated difficulties are considered. Chapter 7 indicates some possibilities for future work.

Part Two of the thesis comprises two chapters (Chapters 8-9), giving a brief report on a sample survey conducted to assess the transition of rural people to modern ways of life in the Karnataka State in South India. The background and survey design are explained in Chapter 8. The last chapter, has a brief discussion on the findings.

PART ONE : THEORY

SOME NOTATION AND ABBREVIATIONS

1. Notation

$U_1,\ldots,U_N$	: Units of a finite population.
N,n	: Population and sample sizes.
y,x	: Variate of interest and an auxiliary variate.
Y,X	: Population totals of y and x.
$\bar{Y},\bar{X}$	: Population means of y and x.
$\bar{y},\bar{x}$	: Sample means of y and x.
$\hat{Y},\hat{X}$	: Unbiased estimators of Y and X.
V_{ij}	: Relative central moments of the joint sampling distribution of $(\hat{Y},\hat{X})$.
ρ	: Correlation between $\hat{Y}$ and $\hat{X}$.
k	: $V_{11}/V_{02} = \rho\sqrt{V_{20}/V_{02}}$.

2. Abbreviations

fpc	: Finite population correction.
mse	: Mean squared error.
ppswor	: Probability proportional to size without replacement.
ppswr	: Probability proportional to size with replacement.
srswor	: Simple random sampling without replacement.
srswr	: Simple random sampling with replacement.

CHAPTER 2.

A DUAL TO RATIO ESTIMATORS

[This chapter proposes a new product-type estimator which is complementary, in a certain sense, to the commonly used ratio estimator. Exact expressions for the bias and mse can be derived which is not the case for the ratio estimator. The correction of the new estimator for bias is examined.]

2.1. Introduction

Consider a finite population of N units $U_1,\ldots,U_N$. Let the variate of interest, y, and an auxiliary variate, x, related to y take real values y_i, x_i on U_i, $i = 1,\ldots,N$. First assume y_i, $x_i \geq 0$, since survey variates are generally non-negative with only occasional exceptions like saving and profit. Section 2.7 and Chapter 6 cover such cases. Let $\hat{Y},\hat{X}$ be unbiased estimators of the population totals Y and X corresponding to the variates y and x respectively, based on any probability sampling design. It is assumed that $\hat{X}$ is known and $X,\hat{X}$ are positive. Also let the coefficient of correlation ρ between $\hat{Y}$ and $\hat{X}$ be positive. Then the traditional ratio method estimates Y by

$$\hat{Y}_r = \hat{Y}(X/\hat{X}) \ . \tag{2.1}$$

The bias and mse of $\hat{Y}_r$ are, up to second order moments

$$B(\hat{Y}_r) = Y(1-k)\, V_{02} , \tag{2.2}$$

$$M_1(\hat{Y}_r) = Y^2[V_{20} + (1-2k)V_{02}] , \tag{2.3}$$

where V_{ij} are the relative central moments defined by

$$V_{ij} = E(\hat{Y}-Y)^i\, (\hat{X}-X)^j/Y^iX^j , \tag{2.4}$$

$$k = V_{11}/V_{02} = \rho \sqrt{V_{20}/V_{02}} \tag{2.5}$$

and E denotes averaging over all possible samples.

Let $n(< N)$ be the sample size. Then

$$\hat{X}^* = (NX - n\hat{X})/(N-n) \tag{2.6}$$

is also unbiased for X, and $\mathrm{cor}(\hat{Y},\hat{X}^*) = -\mathrm{cor}(\hat{Y},\hat{X}) = -\rho$. An interpretation of $\hat{X}^*$ is given in Section 2.7. Since $\hat{X}^*$ is negatively correlated with $\hat{Y}$, we form a product-type estimator based on $\hat{X}^*$. Note that $\hat{X}^*$ can be easily computed once $\hat{X}$ is known. Therefore as an estimator of Y consider

$$\hat{Y}_a = \hat{Y}(\hat{X}^*/X) . \tag{2.7}$$

2.2 Bias and MSE of $\hat{Y}_a$

Write $\hat{Y} = Y(1+e_1)$, $\hat{X} = X(1+e_2)$ with $E(e_1) = E(e_2) = 0$.

Then

$$\begin{aligned}\hat{X}^* &= (NX - n\hat{X})/(N-n)\\ &= [NX - nX(1+e_2)]/(N-n)\\ &= X(1-ge_2) , \text{ where } g = n/(N-n) .\end{aligned}$$

And

$$\hat{Y}_a = \hat{Y}\hat{X}^*/X = Y(1+e_1)(1-ge_2) = Y(1+e_1-ge_2-ge_1e_2) \ . \quad \text{Hence}$$

$$B(\hat{Y}_a) = E(\hat{Y}_a) - Y$$

$$= - gYE(e_1e_2) \ , \text{ since } E(e_1) = E(e_2) = 0 \ ,$$

$$= - gYV_{11} = - gkYV_{02} \ . \qquad (2.8)$$

Next, the mse of $\hat{Y}_a$ is

$$M(\hat{Y}_a) = E(\hat{Y}_a-Y)^2$$

$$= Y^2E(e_1-ge_2-ge_1e_2)^2$$

$$= Y^2E(e_1^2+g^2e_2^2-2ge_1e_2-2ge_1^2e_2+2g^2e_1e_2^2+g^2e_1^2e_2^2)$$

$$= Y^2(V_{20}+g^2V_{02}-2gV_{11}-2gV_{21}+2g^2V_{12}+g^2V_{22})$$

$$= Y^2[V_{20}+g(g-2k)V_{02} + 2g(gV_{12}-V_{21}) + g^2V_{22}] \ . \qquad (2.9)$$

From (2.8) we see that when $V_{11} > 0$ the bias is always negative. Usually $g \leq 0.10$ and it is seen in (2.13) that $\hat{Y}_a$ is preferred to $\hat{Y}_r$ when $k < (1+g)/2$. In such cases (2.8) implies that $|B(\hat{Y}_a)|/Y \leq 0.055V_{02}$. Therefore the relative bias in $\hat{Y}_a$ is likely to be negligible, assuming V_{02} to be small.

Now in particular consider srswr or varying probability sampling with replacement or any other scheme involving independent subsamples. Let $\hat{Y}_t, \hat{X}_t$ be unbiased estimators of Y and X respectively, based on the t^{th} selection or subsample; $t = 1,\ldots,n$. Here n is the sample size or the number of subsamples, as the case may be. Let V'_{ij} denote

the relative central moments of the joint sampling distribution of $\hat{Y}_t, \hat{X}_t$. These are independent of t . Also the estimators $\hat{Y}_t, \hat{Y}_s, \hat{X}_t$, $\hat{X}_s$ are pairwise independent for $s \neq t$, except for the pairs $(\hat{Y}_t, \hat{X}_t)$ and $(\hat{Y}_s, \hat{X}_s)$. Suppose we take $\hat{Y} = \Sigma \hat{Y}_t/n$, $\hat{X} = \Sigma \hat{X}_t/n$ as the unbiased estimators of Y and X to be used in the estimator $\hat{Y}_a$, and V_{ij} denote the relative central moments of the joint distribution of $(\hat{Y}, \hat{X})$. Then

$$V_{20} = E\left(\frac{\Sigma \hat{Y}_t}{n} - Y\right)^2 / Y^2 = \frac{1}{n^2 Y^2} \sum_t E(\hat{Y}_t - Y)^2 = \frac{V'_{20}}{n} .$$

Similarly $V_{02} = V'_{02}/n$ and $V_{11} = V'_{11}/n$. Next

$$V_{21} = E\left(\frac{\Sigma \hat{Y}_t}{n} - Y\right)^2 \left(\frac{\Sigma \hat{X}_t}{n} - X\right)/Y^2 X$$

$$= \frac{1}{n^3 Y^2 X} E[(\hat{Y}_1 - Y) + \ldots + (\hat{Y}_n - Y)]^2 [(\hat{X}_1 - X) + \ldots + (\hat{X}_n - X)]$$

$$= \frac{1}{n^3 Y^2 X} \sum_t E(\hat{Y}_t - Y)^2 (\hat{X}_t - X) \text{ , since the other terms vanish,}$$

$$= V'_{21}/n^2$$

and similarly $V_{12} = V'_{12}/n^2$. Finally

$$V_{22} = E\left(\frac{\Sigma \hat{Y}_t}{n} - Y\right)^2 \left(\frac{\Sigma \hat{X}_t}{n} - X\right)^2 / Y^2 X^2$$

$$= \frac{1}{n^4 Y^2 X^2} E[(\hat{Y}_1 - Y) + \ldots + (\hat{Y}_n - Y)]^2 [(\hat{X}_1 - X) + \ldots + (\hat{X}_n - X)]^2$$

$$= \frac{1}{n^4 Y^2 X^2} E[\Sigma_t (\hat{Y}_t - Y)^2 (\hat{X}_t - X)^2 + \Sigma_t \Sigma_{s \neq t} (\hat{Y}_t - Y)^2 (\hat{X}_s - X)^2$$

$$+ 2\sum_{t}\sum_{s\neq t}(\hat{Y}_t-Y)(\hat{Y}_s-Y)(\hat{X}_t-X)(\hat{X}_s-X)]$$

$$= \frac{1}{n^4}[nV'_{22} + n(n-1)V'_{20}V'_{02} + 2n(n-1)V'^2_{11}]$$

$$= [V'_{22} + (n-1)(V'_{20}V'_{02} + 2V'^2_{11})]/n^3 .$$

Therefore the bias and mse of $\hat{Y}_a$ are

$$B(\hat{Y}_a) = -gYV'_{11}/n , \tag{2.10}$$

$$M(\hat{Y}_a) = Y^2[\{V'_{20} + g(g-2k)V'_{02}\}/n + 2g(gV'_{12} - V'_{21})/n^2 + g^2\{V'_{22} + (n-1)(V'_{20}V'_{02} + 2V'^2_{11})\}/n^3] . \tag{2.11}$$

For srswor, (2.10) and (2.11) are applicable only if N is large enough for the fpc to be ignored. These expressions show that when n is at least moderately large the contribution of bias to mse will be small, and that the terms involving $1/n^2$ or $1/n^3$ in the mse can be neglected, in which case (2.11) reduces to

$$M_1(\hat{Y}_a) = Y^2[V'_{20} + g(g-2k)\ V'_{02}]/n .$$

2.3 Efficiency Comparisons

To keep comparisons tangible, the mse is taken only up to second order moments in this section. Thus

$$M_1(\hat{Y}_a) = Y^2[V_{20} + g(g-2k)V_{02}] . \tag{2.12}$$

The estimator $\hat{Y}_a$ is more precise than $\hat{Y}$ when $M_1(\hat{Y}_a) < V(\hat{Y}) = Y^2V_{20}$.

This is the case when $g(g-2k)V_{02} < 0$, which implies $2k > g$. On the other hand $\hat{Y}_a$ is to be preferred to $\hat{Y}_r$ when $M_1(\hat{Y}_a) < M_1(\hat{Y}_r)$. From (2.3) and (2.12) we see that this needs $g(g-2k) < (1-2k)$. That is $2k(1-g) < 1 - g^2$, or $2k < 1 + g$, assuming $1 - g > 0$. Thus $\hat{Y}_a$ is more efficient that $\hat{Y}$ or $\hat{Y}_r$ when

$$g/2 < k < (1+g)/2 , \tag{2.13}$$

regarded as a condition on k . While deriving this it is assumed that $(1-g) > 0$, that is $n < N/2$ which may be taken as a typical survey situation. Also it is seen that $\hat{Y}_r$ is more efficient than $\hat{Y}$ when

$$k > 1/2 . \tag{2.14}$$

Since g is usually small, (2.13) and (2.14) imply that for the most part $\hat{Y}_a$ is superior, in terms of mse , to $\hat{Y}$ just when $\hat{Y}_r$ is inferior to $\hat{Y}$. In this sense $\hat{Y}_a$ and $\hat{Y}_r$ are complementary. To get a simpler idea let $V_{20} = V_{02}$ and $n = N/5$. Then $\bar{k}$ reduces to ρ and (2.13) is satisfied by any ρ in (0. 125, 0.625) while (2.14) needs $\rho > 0.5$. In general (2.13) specifies an interval of length 1/2 for k . This interval slides to the right from (0, 1/2) to (1/2, 1) as n is increased from 0 to $N/2$. In the less likely case of $n > N/2$, (2.13) is to be replaced by

$$k > (1+g)/2 , \tag{2.13a}$$

and $\hat{Y}_a$ becomes an alternative to $\hat{Y}_r$. In any case the reductions in

mse are

$$V(\hat{Y}) - M_1(\hat{Y}_a) = gY^2 (2k-g)V_{02} ,$$
$$M_1(\hat{Y}_r) - M_1(\hat{Y}_a) = (1-g) Y^2(1+g-2k)V_{02} . \qquad (2.15)$$

The expression for $M_1(\hat{Y}_a)$ in (2.12) reveals an unusual type of relation between the mse of $\hat{Y}_a$ and sample size. For example, under srswor if $\hat{Y}$ and $\hat{X}$ are the simple expansion estimators then $M_1(\hat{Y}_a)$ decreases as n increases only as long as $n < Nk/(\rho+k)$. Thereafter $M_1(\hat{Y}_a)$ increases with n . In general for any sampling design where V_{20}, V_{11} and V_{02} are proportional to $(N-n)/n$, the dependence of the mse on n will be similar.

2.4 Unbiased Estimators

It was pointed out in Sec. 2.2 that the bias in $\hat{Y}_a$ is likely to be small. However unusual situations may exist where the coefficient of variation of $\hat{X}$ is large and consequently the bias becomes significant. In such cases the use of exactly unbiased estimators may be of great advantage (Rao, 1969). Then we may consider the following alternatives:

(i) We may make $\hat{Y}$ and $\hat{X}$ uncorrelated, so that $\hat{Y}_a$ becomes unbiased as is apparent from (2.8). But this situation is not good since there will be an unacceptable increase in variance relative to $V(\hat{Y})$.

(ii) We may draw the sample in the form of n independent interpenetrating subsamples. Then let $\hat{Y}_i$, $\hat{X}_i$ be unbiased estimators of Y and X based on

the i^{th} subsample, $\mathrm{cor}(\hat{Y}_i, \hat{X}_i) > 0$ and $\hat{X}_i^* = (NX - n\hat{X}_i)/(N-n)$ for $i = 1, \ldots, n$. Consider the estimators

$$\hat{Y}_1 = (\sum^n \hat{Y}_i/n)(\sum^n \hat{X}_i^*/n)/X, \qquad (2.16)$$

$$\hat{Y}_2 = \sum^n \hat{Y}_i \hat{X}_i^*/(nX) .$$

Following Murthy (1964), we can show that $B(\hat{Y}_2) = nB(\hat{Y}_1)$ and hence that

$$\hat{Y}_3 = (n\hat{Y}_1 - \hat{Y}_2)/(n-1)$$

is unbiased for Y. The conditions for $\hat{Y}_3$ to be more efficient than $\hat{Y}_1$ are similar to those given by Murthy and Nanjamma (1959) in the case of obtaining an almost unbiased ratio estimator.

In particular consider srswr. Let $\bar{x}$, $\bar{y}$ be the sample means and $\hat{X} = N\bar{x}$, $\hat{Y} = N\bar{y}$. Then (2.10) reduces to $B(\hat{Y}_a) = -N^2 S_{11}/[(N-n)X]$, where S_{11} is the population covariance

$$\sum^N (x_i - \bar{X})(y_i - \bar{Y})/(N-1) .$$

This bias can be estimated in an unbiased way by $-N^2 s_{11}/[(N-n)X]$, where s_{11} is the sample covariance $\sum^n (x_i - \bar{x})(y_i - \bar{y})/(n-1)$. Using this to correct $\hat{Y}_a$ for its bias, we get the estimator $\hat{Y}_4 = \hat{Y}_a + N^2 s_{11}/[(N-n)X]$. A little manipulation shows that $\hat{Y}_3$ in fact reduces to $\hat{Y}_4$ in this case. And for srswor $B(\hat{Y}_a) = -NS_{11}/X$, which is seen to be independent of n. The corresponding unbiased estimator is $\hat{Y}_5 = \hat{Y}_a + Ns_{11}/X$.

Srivastava et al.(1981)have shown that this method of adjusting the product-type estimator for its bias by subtracting an unbiased estimate of bias (Robson, 1957) is to be generally preferred when x and y have a joint distribution with central moments $\mu_{12} = \mu_{21} = 0$. In other cases we may apply the technique developed by Quenouille (1956), as in the next possibility.

(iii) Take $n = 2m$ and split the sample at random into two subsamples of m units each. Let $\hat{Y}_i$, $\hat{X}_i$ $(i = 1, 2)$ be unbiased estimators of Y and X based on the subsamples and $\hat{Y}$, $\hat{X}$ those based on the entire sample. Take $\hat{X}_i^* = (NX-n\hat{X}_i)/(N-n)$ and $\hat{X}^* = (NX-n\hat{X})/(N-n)$. And consider the product-type estimators $\hat{Y}_{ai} = \hat{Y}_i\hat{X}_i^*/X$, $(i = 1,2)$ and $\hat{Y}_a = \hat{Y}\hat{X}^*/X$. Then it can be shown that an unbiased estimator of Y is

$$\hat{Y}_6 = (2N-n)\,\hat{Y}_a/N - (N-n)(\hat{Y}_{a1}+\hat{Y}_{a2})/2N . \qquad (2.17)$$

The variance of $\hat{Y}_6$ can also be evaluated. Details are similar to those in Sukhatme and Sukhatme (1970, pp.161-5). Interestingly enough, the variance of $\hat{Y}_6$ and the mse of $\hat{Y}_a$ are equal up to second order moments. Since $\hat{Y}_6$ is unbiased, while $\hat{Y}_a$ is not, the former may be preferred to the latter.

2.5 Empirical Study

For illustration srswor is assumed in this section and $\hat{Y} = N\bar{y}$, $\hat{X} = N\bar{x}$. First consider a hypothetical population, I, of 5 units having

(4,6), (5,5), (7,7), (8,2) and (12,10) as values of (y,x). Here $Y = 36$, $X = 30$ and $k = 0.44$. Assuming $n = 2$, the 10 possible samples were listed and the biases and mse's of the estimators were computed from first principles, to avoid approximations. The results are in Table 2.1. A poor performance by $\hat{Y}_r$ may be anticipated here, since the points (x_i, y_i) do not lie near a line through the origin.

Table - 2.1: Performance of estimators for population I

Estimator	MSE	Bias
$\hat{Y}$	72.75	0
$\hat{Y}_r$	118.77	1.82
$\hat{Y}_a$	60.08	-0.75

Next, the estimators are applied to eight populations, seven of which have been published and used for comparative purposes in the literature. The other one is the population of livestock and cows yielding milk as noted in 50 of the villages in the survey discussed in Part Two of this thesis. Table 2.2 gives the source, nature of y and x and the value of k for these populations. Let ϕ denote the ratio of $V(\hat{Y})$ to the mse of an alternative estimator, expressed as a percentage. This can be taken as an indicator of the performance of the alternative estimator. The values of ϕ and bias of the estimators are summarized in Table 2.3. Computations were done only up to second order moments as is the usual practice. The sampling fractions of $f = 0.02$, 0.05, 0.10 and 0.15 are used to demonstrate

the increase, with f, in the efficiency of $\hat{Y}_a$ relative to that of $\hat{Y}$.

Table - 2.2: Description of the populations II to IX

Population	Source	y	x	N	k
II	Sampford (1967)	Hypothetical	Hypothetical	10	0.23
III	Murthy (1967, p.228)	Output in a factory	Number of workers	80	0.35
IV	Murthy (1967, p.228)	Output in a factory	Fixed capital	80	0.44
V	Yates (1960, p.163)	Measured vol. of timber	Eye-estimated vol. of timber	25	0.57
VI	Cochran (1977, p.325)	Number of persons in a block	Number of rooms in a block	10	0.74
VII	Table 3.4	Number of livestock in the village	Number of cows yielding milk in the village	50	0.82
VIII	Yates (1960, p.159)	Total number of persons in a kraal	Number of persons absent from a kraal	43	0.89
IX	Kish (1965, p.625)	Number of dwellings occupied by renters	Number of dwellings	270	1.20

Table - 2.3: The values of ϕ for $\hat{Y}$, $\hat{Y}_r$ and $\hat{Y}_a$ and values of $100|Bias|/Y$ for $\hat{Y}_r$ and $\hat{Y}_a$

Population	k	$\hat{Y}$	$\hat{Y}_r$	$\hat{Y}_a$				$100\|B\|/Y$	
				f=0.02	f=0.05	f=0.10	f=0.15	$\hat{Y}_r$	$\hat{Y}_a$
II	0.23	100	34.54	103.20	107.99	115.43	120.79	12.66	0.42
III	0.35	100	31.78	111.06	132.37	188.64	297.35	6.57	0.39
IV	0.44	100	64.91	109.00	125.79	167.39	241.25	3.69	0.32
V	0.57	100	113.54	102.06	105.33	111.24	117.68	3.06	0.45
VI	0.74	100	158.83	102.37	106.21	113.42	121.80	0.43	0.13
VII	0.82	100	337.50	101.10	102.90	106.10	110.00	0.03	0.03
VIII	0.89	100	147.68	101.51	103.92	108.33	113.30	0.48	0.43
IX	1.20	100	1178.57	103.28	108.80	119.98	134.58	0.65	0.44

For populations I - IV , cases of moderate k , the estimator $\hat{Y}_a$ is seen to be appropriate, for population V , $k = 0.57$, $\hat{Y}_r$ and $\hat{Y}_a$ are almost equally good, while for others, large k, $\hat{Y}_r$ is better. From Table 2.3, it may be noted that the relative bias of $\hat{Y}_a$ is quite small. Thus the empirical study indicates (i) that $\hat{Y}_a$ is dual to $\hat{Y}_r$ in that it performs well just when $\hat{Y}_r$ does not, and (ii) that the general conclusions of Sections 2.2 and 2.3, based on approximations apt for large n , do apply even on a scale as small as that of populations I, II or VI.

2.6 Optimality of the Estimators

Suppose that the population is divided into r classes and that in the i^{th} class $x = x_i$ for all the units: Let n be allocated to the different classes proportional to their sizes and sampling be simple random in each class. Then it is easy to see that the sample mean $\bar{x}$ will always equal the population mean $\bar{X}$. Now if $\hat{X} = N\bar{x}$, then the estimators $\hat{Y}$, $\hat{Y}_r$ and $\hat{Y}_a$ coincide. Therefore in a situation approximating to the above the three estimators are expected to perform equally well.

In general, for $i = 1,\ldots, N$, define $u_i = a - bx_i$, where a,b are positive constants and $u_i > 0$. With srswor and $\hat{X} = N\bar{x}$, the estimator $\hat{Y}_a$ proposed in this chapter is the same as the product estimator $\hat{Y}\bar{u}/\bar{U}$ for the choice $a = N\bar{X}/(N-n)$, $b = n/(N-n)$. Here $\bar{U}$, $\bar{u}$ are population and sample means for u . The conditions under which a ratio estimator is optimal are known (Brewer, 1963; Royall, 1970; Royall and Herson, 1973).

An account of these is given by Cochran (1977, pp.158-160). The corresponding conditions for optimality of a product estimator based on arithmetic means are difficult to obtain. However, the following may be mentioned.

Suppose that the N population values (y_i, u_i) are a random sample from a superpopulation in which

$$y_i = B/u_i + e_i' , \tag{2.18}$$

where the e_i' are independent of the u_i. In arrays in which u_i is fixed, e_i' has mean zero and variance proportional to $1/u_i$. Then the estimator $\hat{Y}_h = \hat{Y}\tilde{u}/\tilde{U}$ where $\tilde{u}$ and $\tilde{U}$ are respectively sample and population harmonic means for u, is optimal in the same sense as in Brewer-Royall results. But the use of harmonic means is not favoured by survey practitioners. However, in the cases where it is felt that $\bar{u}/\bar{U} = \tilde{u}/\tilde{U}$ at least approximately, the estimator $\hat{Y}_a$ will be nearly optimal under model (2.18). The practical relevance of this result is that it suggests conditions under which $\hat{Y}_a$ may be the best in an entire class of estimators.

2.7 Remarks

For srswor and with $\hat{X} = N\bar{x}$, $\hat{X}^*$ as defined in (2.6) reduces to the simple expansion estimator of X based on the $(N-n)$ population units not included in the sample.

The expressions (2.8) and (2.9) for bias and mse of $\hat{Y}_a$ are exact.

These can be estimated by replacing the concerned parameters by the corresponding sample statistics.

If the auxiliary variate assumes both negative and positive values, the use of the ratio or the product method of estimation based directly on x is better avoided since $\hat{X}$ or X may happen to be close to zero. Chapter 6 considers this problem. In general, the results of the present chapter hold whenever $V_{11} > 0$, that is $\rho XY > 0$. If $\rho XY < 0$, a ratio estimator can be formed with $\hat{X}^*$ in the denominator. Conditions for its optimality follow easily from Brewer-Royall results.

CHAPTER 3.

EXTENDING PRODUCT METHOD OF ESTIMATION TO THE POSITIVE CORRELATION CASE

[Continuing with the idea of the previous chapter a general transformation is suggested below to permit a product method of estimation rather than a ratio method in the common situation of positive correlation between $\hat{Y}$ and $\hat{X}$. This leads to the advantage that the bias and mse have exact expressions. An extension to use multiauxiliary information is outlined.]

3.1 Introduction

The previous chapter proposed a product-type estimator as a dual to the ratio estimator. This was achieved through a simple transformation of $\hat{X}$ defined in (2.6). The present chapter seeks a general transformation which allows the use of the product method even in cases appropriate for ratio estimaor, since the former admits closed form expressions for the bias and mse while the latter does not. In fact the closeness of the expressions (2.2), (2.3) respectively to the actual bias and mse of the ratio estimator $\hat{Y}_r$ depends much on the composition of the population, the sampling design and the sample size. Hence these expressions must be taken with reservation (Murthy, 1967, p.365). Motivated by this, consider simple transformations

that render the situation suitable for a product method instead of the ratio method. First assume the variates to be nonnegative and ρ to be positive. Let $\hat{Z} = L - \hat{X}$, where L is a scalar to be chosen. Clearly $\hat{Z}$ is unbiased for $Z = L - X$ and $\text{cor}(\hat{Y},\hat{Z}) = -\rho$. Now consider the following estimator of Y :

$$\hat{Y}_p = \hat{Y}(\hat{Z}/Z) . \tag{3.1}$$

The bias and mse of $\hat{Y}_p$ are, with $\Theta = X/(L-X)$,

$$B(\hat{Y}_p) = -\Theta Y V_{11} \tag{3.2}$$

$$M(\hat{Y}_p) = Y^2[V_{20} + \Theta(\Theta-2k)V_{02} + 2\Theta(\Theta V_{12}-V_{21}) + \Theta^2 V_{22}] . \tag{3.3}$$

These are respectively the same as the expressions for bias and mse of the estimator $\hat{Y}_a$ of the previous chapter as given in (2.2), (2.3) except that g is now replaced by Θ. The variance estimators for products of estimators have been considered by Goodman (1960).

3.2 The Choice of L

$M(\hat{Y}_p)$ is minimized when $\Theta_{opt} = (kV_{02}+V_{21})/(V_{02} + 2V_{12}+V_{22})$ and the corresponding L is

$$\begin{aligned} L_{opt} &= X(1 + \Theta_{opt})/\Theta_{opt} \\ &= X(1 + 1/k) + X(2V_{12}+V_{22}-V_{21}/k)/(kV_{02}+V_{21}) . \end{aligned} \tag{3.4}$$

Let $\hat{Y}_p^*$ denote $\hat{Y}_p$ for optimum L. The bias and mse of $\hat{Y}_p^*$ are obtained by replacing Θ in (3.2) with Θ_{opt}. Ideally, one should know L_{opt}, but in practice this is seldom possible. At the best an approximation to L_{opt} may be obtained. For example as $X(1 + 1/k)$; the second term on the right side of (3.4) is likely to be unimportant since V_{ij} with $i + j > 2$ are generally small (Murthy, 1967, pp.380-81). The quantity X is known and it is often possible to assess the value of k. This problem of assessing certain parameters has been studied among others by Murthy (1967, pp.96-99) and Reddy (1978). Let L_o denote the corresponding approximation to L_{opt}. We examine below to what extent L_o may deviate from L_{opt} and yet give an estimator better than $\hat{Y}_r$ and $\hat{Y}$.

To get specific ideas, the V_{ij} with $i + j > 2$ are ignored in the expressions. Thus $\Theta_{opt} = k$ and

$$M(\hat{Y}_p^*) = (1-\rho^2)\, Y^2 V_{20} = (1-\rho^2)\, V(\hat{Y}) , \tag{3.5}$$

$$\begin{aligned} M(\hat{Y}_p) &= Y^2[V_{20} + \Theta_o(\Theta_o-2k)V_{02}] \\ &= M(\hat{Y}_p^*)[1 + e^2\rho^2/(1-\rho^2)] , \end{aligned} \tag{3.6}$$

if $\Theta = \Theta_o = k(1+e)$ when $L = L_o$. It is seen that $M(\hat{Y}_p^*)$ is the same as the variance of the difference estimator $\hat{Y} - B(\hat{X}-\bar{X})$ in the ideal case, namely when B is the coefficient of regression of $\hat{Y}$ on $\hat{X}$. $M(\hat{Y}_p^*)$ is also the same as the large sample approximation to the mse of the regression estimator. From (3.6) it follows that the proportional increase in the mse

of $\hat{Y}_p$ over that of $\hat{Y}_p^*$ is less than d if

$$|e| < \sqrt{d(1-\rho^2)/\rho^2} \quad . \tag{3.7}$$

Thus to ensure only a small relative increase in mse, $|e|$ must be close to 0 if ρ is high, but can depart considerably from 0 if ρ is just moderate. Also from (2.3) and (3.6) we get

$$M(\hat{Y}_r) - M(\hat{Y}_p) = Y^2[(k-1)^2 - (\Theta_o - k)^2]V_{02} > 0$$

when

$$\Theta \text{ lies between } (2k-1) \text{ and } 1 \; . \tag{3.8}$$

Similarly a necessary and sufficient condition for $M(\hat{Y}_p) < V(\hat{Y})$ is

$$0 < \Theta_o < 2k \; . \tag{3.9}$$

To investigate where (3.8), (3.9) are satisfied simultaneously we distinguish between the cases $0 \leq k \leq 1$ and $k > 1$.

Case I. $0 \leq k \leq 1$.

Here choose $L_o > 2X$ so that Θ_o is in (0,1) . If k happens to be in (0,0.5), the condition (3.8) is automatically met since $(2k-1) < 0$, but (3.9) needs

$$L_o > (1 + 1/2k)X \; . \tag{3.10}$$

On the other hand if k is in (0.5,1), then (3.9) is always met since $2k > 1$, but (3.8) requires

$$L_o < [1 + 1/(2k-1)]X \; . \tag{3.11}$$

Thus any $L_o > 2X$ satisfying (3.10) or (3.11) as the case may be will make $\hat{Y}_p$ an improved estimator.

Case II. $k > 1$.

Here choose $L_o < 2X$. In addition we need only that $L_o > [1 + 1/(2k-1)]X$ for $\hat{Y}_p$ to be more precise than $\hat{Y}_r$ or $\hat{Y}$. To give a clearer idea some typical situations are presented in Table 3.1.

Table 3.1: Optimum L and lower and upper bounds on the choice of L for typical values of k .

k	Lower bound on L	Optimum L	Upper bound on L
0.1	6.00X	11.00X	∞
0.2	3.50X	6.00X	∞
0.3	2.67X	4.33X	∞
0.4	2.25X	3.50X	∞
0.5	2.00X	3.00X	∞
0.6	2.00X	2.67X	6.00X
0.7	2.00X	2.43X	3.50X
0.8	2.00X	2.25X	2.66X
0.9	2.00X	2.11X	2.25X
1.0	2.00X	2.00X	2.00X
1.1	1.83X	1.91X	2.00X
1.3	1.63X	1.77X	2.00X
1.5	1.50X	1.67X	2.00X
2.0	1.33X	1.50X	2.00X
2.5	1.25X	1.40X	2.00X
3.0	1.20X	1.33X	2.00X

Interestingly the choice $L_o = 2.25X$ covers a fairly wide range for k from 0.4 to 0.9, being actually optimum for $k = 0.8$. Similarly $L_o = 3.5X$ suits the range from 0.2 to 0.7 for k with optimum at $k = 0.4$. In fact the choice of L_o is very flexible when k is moderate, say in (0,0.7). This flexibility disappears in the neighbourhood of $k = 1$. However a value like $L_o = 1.9X$ is safe virtually for all $k > 1$. Better selections can be made when k is known more precisely.

3.3 Case of Negative Correlation

When $\rho < 0$, take $\hat{Z} = L + \hat{X}$ so that $\text{cor}(\hat{Y},\hat{Z}) = \text{cor}(\hat{Y},\hat{X})$. Here it is appropriate to compare $\hat{Y}_p$ with the traditional product estimator $\hat{Y}(\hat{X}/X)$. An approximation to L_{opt} is $-(1 + 1/k)X$. The restrictions on the choice L_o for L_{opt} can be investigated. It turns out that $L_o = 0.25X$ covers the range -0.9 to -0.4 for k, being the best $k = -0.8$, while $L_o = 1.5X$ is suitable for k in $(-0.7, -0.2)$ being actually optimum at $k = -0.4$. And a choice like $L_o = -0.10X$ is practically safe for all $k < -1$. A better selection can be made if k is more precisely known. Figure 3.1 presents the results in a graphical form.

The rules of thumb for choosing L when a firm guess of the value of k cannot be made, but only an interval containing k can be specified, are given in Table 3.2.

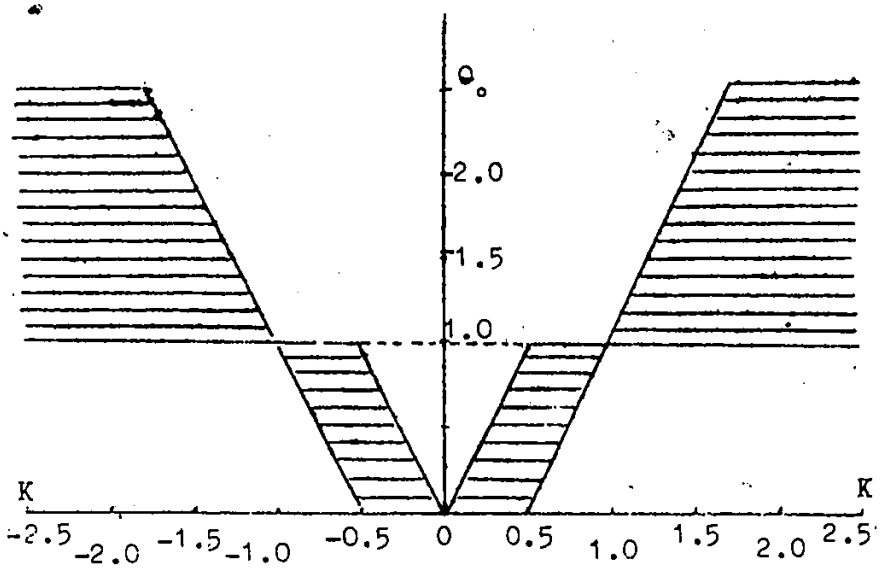


Fig. 3.1 Region where $\hat{Y}_p$ has smaller variance

Table - 3.2: Rules of thumb for choosing L

k > 0		k < 0	
Case	L	Case	L
$0 < k \leq \frac{1}{2}$	3.50X	$-\frac{1}{2} \leq k < 0$	1.50X
$\frac{1}{2} < k < 1$	2.25X	$-1 < k < -\frac{1}{2}$	0.25X
$k > 1$	1.90X	$k < -1$	-0.10X

These rules are not applicable when k is in the neighbourhood of either 0 or ± 1. In fact if k is (i) close to 0 the simple estimator $\hat{Y}$ is to be used, (ii) close to 1 either $\hat{Y}_r$ or $\hat{Y}_p$ with L = 2X may be used, and (iii) close to -1, $\hat{Y}_p$ with L = 0 (which is the same as the usual product estimator) may be used.

If it is desirable to use an exactly unbiased estimator we may consider the alternatives suggested in Sec. 2.4, with the change that in the place of $\hat{X}^*$ and X we now use $\hat{Z}$ and Z.

3.4 Use of Multiauxiliary Information

Frequently information on several x-variates may be used: for instance utilizing census data to adjust current estimates. Let x_t, $t = 1,\ldots, q+s$ be the auxiliary variates and $\hat{Y},\hat{X}_t$ be unbiased estimators of the totals Y,X_t, based on any probability sampling design. All values are real, nonnegative and X_t are known. Also let $\text{cor}(\hat{Y},\hat{X}_t) > 0$

for $t = 1,\ldots,q$ and $\mathrm{cor}(\hat{Y},\hat{X}_t) < 0$ for $t = q+1,\ldots,q+s$. Then Rao and Mudholkar (1967) have suggested the following linear combination of estimators of ratio and product-types as an estimator of Y:

$$\hat{Y}_{RM} = \hat{Y}\left[\sum_{1}^{q} a_t(X_t/\hat{X}_t) + \sum_{q+1}^{q+s} a_t(\hat{X}_t/X_t)\right]$$

where the a_t are the weights; $\sum_{1}^{q+s} a_t = 1$. As an alternative we consider the following extension of the methods of Sections 3.1 and 3.3.

The estimator $\hat{Z}_t = L_t + h_t\hat{X}_t$ is unbiased for $Z_t = L_t + h_tX_t$ for each t. Take $h_t = -1$ for $t = 1,\ldots,q$ and $h_t = 1$ for $t = q+1,\ldots,q+s$. Then an estimator of Y is

$$\hat{Y}_p = \hat{Y}\sum_{}^{q+s} W_t(\hat{Z}_t/Z_t) \tag{3.12}$$

where $W = (W_1,\ldots,W_{q+s})$, $\sum_{}^{q+s} W_t = 1$, is a vector of weights. And

$$B(\hat{Y}_p) = Y\left(\sum_{q+1}^{q+s} W_t\Theta_t V_{11}^{(t)} - \sum_{1}^{q} W_t\Theta_t V_{11}^{(t)}\right) \tag{3.13}$$

and the mse, up to second order moments, is

$$M(\hat{Y}_p) = Y^2\sum_{t,u=1}^{q+s} W_tW_ud_{tu} = Y^2.\ WDW' \tag{3.14}$$

where the elements of the matrix D are given by

$$d_{tu} = V_{20} + h_t\Theta_tV_{11}^{(t)} + h_u\Theta_uV_{11}^{(u)} + h_th_u\Theta_t\Theta_uV_{11}^{(tu)}$$

with $\Theta_t = X_t/(L_t + h_tX_t)$; $t, u = 1,\ldots,q+s$. Here $V_{ij}^{(t)}$ stands for the

relative central moments V_{ij} defined in (2.4) with $(\hat{X}-X)$ replaced by $\hat{X}_t - X$, X by X_t and

$$V_{11}^{(tu)} = E(\hat{X}_t - X_t)(\hat{X}_u - X)/X_t X_u .$$

As shown in Rao and Mudholkar (1967) the matrix D is positive definite if the $(q+s+1) \times (q+s+1)$ matrix of the coefficients of variation of $\hat{Y}$ and $\hat{Z}_t$ is positive definite.

Theoretically the L_t can be determined to minimize $M(\hat{Y}_p)$. But a practicable alternative is to choose L_t such that d_{tt} is controlled. Thus as a rule of thumb, L_t may be $3.5X_t$ if $k_t = V_{11}^{(t)} / V_{02}^{(t)}$ is positive but moderate, while L_t may be $1.5X_t$ when k_t is negative but moderate. Other choices may be made as discussed in Sections 3.2 and 3.3.

Next, applying the generalized Cauchy inequality (see Olkin, 1958) the W_t optimum in the sense of minimizing $M(\hat{Y}_p)$ for given D are provided by

$$W_{opt} = eD^{-1}/eD^{-1}e'$$

where $e = (1,\ldots,1)$. Substituting W_{opt} in (3.14),

$$M_{min}(\hat{Y}_p) = Y^2/eD^{-1}e' .$$

However, in surveys W_{opt} can rarely be computed and used since the matrix D is usually unknown. Theoretically W_t will all be equal $(= 1/(q+s))$ if and only if the column sums of D are equal. A hypothetical example of this occurs when the population coefficients of variation of the $\hat{Z}_t$ are

all equal, $\hat{Y}$ is equally correlated with all $\hat{Z}_t$ and all pairs of two different estimators $\hat{Z}_t$ have the same correlation. Usually the W_t are selected from experience and theoretical considerations. In small scale surveys of specialized scope it may be feasible to estimate W_{opt} from the sample itself. The sampling error of these estimates must be examined.

3.5 Empirical Performance of $\hat{Y}_p$

For purposes of illustration, srswor is assumed in this section. The following five populations are considered.

Population 1. Hypothetical population of 7 pairs of (y,x) values: (1,6), (2,5), (4,7), (5,2), (6,4), (8,10) and (9,8).

Population 2. In the survey of villages reported in Part Two of the thesis, the Heads of 20 of the households who had bank accounts were asked to state the number of withdrawals (y) during July-December 1980, number of deposits during July-December 1980 (x_1) and during January-June 1980 (x_2). The respondents were asked to refer their bank passbooks to be able to get these numbers as accurately as possible. The responses are recorded in Table 3.3 as population 2.

Table - 3.3: Number of bank deposits (y) during a 6-month period and withdrawals (x_1,x_2) during two 6-month periods by 20 households in the survey of Karnataka.

y	:	12	22	38	15	18	31	15	20	10	25	11	17	12	22	14	26	08	16	13	19
x_1	:	14	25	37	18	20	30	15	21	12	28	14	19	12	23	16	28	09	15	15	20
x_2	:	30	25	09	30	28	12	30	24	36	28	30	30	31	25	31	25	35	25	30	28

Population 3. Data on the number of workers (x_1), fixed capital x_2 and output (y) for 80 factories in a certain region (Murthy, 1967, p.228).

Population 4. Data on cultivated area (y) and area under wheat (x_1,x_2) during two different years for 34 villages in a certain region (Murthy, 1967, p.399, Table 10.6).

Population 5. Data on the number of livestock (y), number of cows yielding milk (x_1) and the number of farms (x_2) in 50 of the villages in a study of the Karnataka State in India, reported in Part Two of the thesis. Details are in Table 3.4.

Table 3.4. Number of livestock (y), milk yielding cows (x_1) and farms (x_2) in 50 villages of Karnataka.

y	x_1	x_2	y	x_1	x_2
400	70	55	715	190	170
630	160	155	845	360	335
1200	320	285	1016	235	219
1170	445	381	184	74	63
1060	250	278	282	62	79
823	120	112	195	71	60
1737	560	632	439	137	100
1061	254	278	854	195	142
360	102	112	821	260	265
945	359	345	740	142	85
470	110	100	750	142	190
1625	481	489	625	130	140
827	125	113	600	125	138
90	5	8	75	5	4
1300	428	340	152	60	38
337	78	82	65	40	35
260	75	105	235	142	128
186	45	28	280	162	190
1760	564	515	335	180	200
605	238	245	432	200	235
700	92	85	645	148	162
524	247	220	800	300	80
571	134	133	926	400	230
962	131	145	421	220	150
407	129	102	850	321	140

For populations 2, 3, 4 and 5, three subcases are studied: using x_1 alone, x_2 alone, and x_1, x_2 as auxiliary variates. These are respectively denoted by 2a, 2b, 2c etc. In the case of population 1, all possible samples of $n = 3$ units were listed and the biases and mse's were computed from first principles to avoid approximation, while in 2a the exact expressions for bias and mse were used. In the remaining cases computations were made only up to second order moments. The rules of thumb in Table 3.2 were applied for choosing L for $\hat{Y}_p$, while L was taken to be $(1 + 1/k)X$ or $-(1 + 1/k)X$ as the case may be for $\hat{Y}_p^*$. When information on x_1 and x_2 was utilized, $\hat{Y}_p$ was compared with the generalized multivariate estimator $\hat{Y}_{RM}$ discussed in Rao and Mudholkar (1967), with weights $W_1 = W_2 = 1/2$. The results are summarized in Table 3.5. For srswor the relative bias of the suggested estimator $\hat{Y}_p$ reduces to

$$B(\hat{Y}_p)/Y = -(N-n)\Theta S_{11}/Nn\bar{Y}\bar{X},$$

where S_{11} is the population covariance between y and x. Hence

$$[Nn/(N-n)] \cdot |B(\hat{Y}_p)|/Y = \Theta S_{11}/\bar{Y}\bar{X},$$

where the right side expression is seen to be independent of the sample size n. To have this convenience the values of $100[Nn/(N-n)] \cdot |\text{Bias}|/Y$ are reported in Table 3.5 within parentheses.

Table 3.5: The values of ϕ and $100\{Nn/(N-n)\}\,|\text{Bias}|/Y$ for different estimators.

Population	k	Estimator			
		$\hat{Y}$	Ratio(product) estimator	$\hat{Y}_p$	$\hat{Y}_p^*$
1	0.57	100	94 (10)	115 (9)	121 (6)
2a	1.11	100	2433 (1)	4253 (16)	4253 (16)
2b	-0.50	100	100 (3)	109 (1)	109 (1)
2c	1.11,-0.50	100	218 (2)	315 (1)	798 (1)
3a	0.35	100	32 (58)	729 (13)	837 (11)
3b	0.44	100	65 (33)	1088 (10)	1197 (11)
3c	0.35,0.44	100	45 (46)	246 (11)	1052 (11)
4a	0.75	100	376 (13)	565 (31)	577 (29)
4b	0.71	100	318 (16)	514 (32)	549 (29)
4c	0.75,0.71	100	365 (15)	570 (32)	590 (29)
5a	0.82	100	338 (38)	381 (31)	382 (28)
5b	0.20	100	56 (86)	101 (18)	105 (17)
5c	0.82,0.20	100	342 (42)	365 (28)	396 (30)

ϕ = 100V($\hat{Y}$)/(mse of the estimator) .

The value of k ranges from -0.50 to 1.11 in Table 3.5. Here substantial gain in efficiency is seen when $\hat{Y}_p$ is used instead of the traditional estimators, in most of the cases. Also $\hat{Y}_p$ compares quite well with the ideal case of $\hat{Y}_p^*$. Thus the illustrations indicate that (i) the use of $\hat{Y}_p$ is desirable in practice, and (ii) the rules of thumb for choosing L work well. If an unbiased estimator is preferred then the techniques outlined in Sec. 2.4 may be employed.

CHAPTER 4.

VARIATE TRANSFORMATIONS AFTER SAMPLING

[A change of origin for y is illustrated for improving the precision of estimators under varying probability sampling. A change of origin under ratio method of estimation and a change of scale under the difference method are also considered.]

4.1 Introduction

It is known that ppswr sampling is very effective when the value of the study variate y is proportional to the value of the measure of size x. In ppswor sampling, proportionality between value of y and the probability of including the unit in the sample is desirable. Sections 4.2 to 4.8 discuss a change of origin for y for achieving such proportionality. Sections 4.9 and 4.10 examine changes of origin and scale for $\hat{Y},\hat{X}$ in the context of ratio and difference methods of estimation. Assume that x is a positive valued variate.

4.2 PPSWR Sampling

Consider with replacement pps sampling of n units, using x as the measure of size. The usual unbiased estimator of Y is

$$\hat{Y}_{pps} = (X/n) \sum^{n} y_i/x_i \ . \qquad (4.1)$$

Let $z_i = y_i - a$; $i = 1,\ldots,N$, a being a scalar to be chosen. As an

estimator of Y consider

$$\hat{Y}_1 = (X/n)\ [\overset{n}{\Sigma}\ (y_i-a)/x_i] + Na\ . \tag{4.2}$$

It is easy to show that $\hat{Y}_1$ is unbiased for Y. On the other hand, if srswr is used an unbiased estimator of Y would have been $\hat{Y} = (N/n)\ \overset{n}{\Sigma}\ y_i$.

The variances of the three competing estimators are given by

$$V(\hat{Y}_{pps}) = (X/n)\ \overset{N}{\Sigma}\ y_i^2/x_i - Y^2/n\ , \tag{4.3}$$

$$V(\hat{Y}_1) = (X/n)\ \overset{N}{\Sigma}\ (y_i-a)^2/x_i - (Y-Na)^2/n\ , \tag{4.4}$$

$$V(\hat{Y}) = (N/n)\ \overset{N}{\Sigma}\ y_i^2 - Y^2/n\ . \tag{4.5}$$

4.3 The Choice of a

From (4.3) and (4.4) we obtain

$$\begin{aligned} V(\hat{Y}_{pps}) - V(\hat{Y}_1) &= \frac{X}{n}\ [2a\overset{N}{\Sigma}(y_i/x_i) - a^2\overset{N}{\Sigma}(1/x_i)] \\ &\quad + [\frac{N^2a^2}{n} - \frac{2NaY}{n}] \\ &= \frac{N\bar{X}}{n}\ [2aN\bar{R} - \frac{a^2N}{\tilde{X}}] + \frac{N^2a}{n}\ [a - 2\bar{Y}] \\ &= \frac{N^2a}{n\tilde{X}}\ [2\bar{X}\tilde{X}\bar{R} - a\bar{X}] + \frac{N^2a}{n\tilde{X}}[a\tilde{X} - 2\tilde{X}\bar{Y}] \\ &= \frac{N^2a}{n\tilde{X}}\ [2\bar{X}\tilde{X}(\bar{R}-R) - a(\bar{X}-\tilde{X})] \end{aligned} \tag{4.6}$$

where $\tilde{X} = N/\overset{N}{\Sigma}(1/x_i)$ is the harmonic mean of x, and $\bar{R} = (1/N)\overset{N}{\Sigma}(y_i/x_i)$.

Therefore $V(\hat{Y}_{pps}) - V(\hat{Y}_1) > 0$, and hence $\hat{Y}_1$ is more precise than $\hat{Y}_{pps}$ when both the factors in (4.6) have the same sign. Since $X - \tilde{X} > 0$ when at least two x_i are unequal, this needs that

$$a > 0 \quad \text{and} \quad a < 2\bar{X}\tilde{X}(\bar{R}-R)/(\bar{X}-\tilde{X})$$

or

$$a < 0 \quad \text{and} \quad a > 2\bar{X}\tilde{X}(\bar{R}-R)/(\bar{X}-\tilde{X}) .$$

That is

$$a \text{ lies between } 0 \text{ and } 2a^* , \tag{4.7}$$

where

$$a^* = \bar{X}\tilde{X}(\bar{R}-R)/(\bar{X}-\tilde{X}) . \tag{4.8}$$

It is interesting to note that the sign of a^* is the same as that of $(\bar{R}-R)$, since $(\bar{X}-\tilde{X}) > 0$. In particular the value of a minimizing $V(\hat{Y}_1)$ is $a_{opt} = a^*$, which is the midpoint of the interval for a , specified in (4.7). Denote the estimator $\hat{Y}_1$ for the choice $a_{opt} = a^*$ by $\hat{Y}_1^*$. The expression for $V(\hat{Y}_1^*)$ is the same as (4.4) with a replaced by a^* .

To get a simpler idea consider the case $y_i = A + Bx_i$, so that there is perfect correlation between y and x . Then

$$\begin{aligned} \bar{R} - R &= (1/N) \sum^{N} (y_i/x_i) - \sum^{N} y_i / \sum^{N} x_i \\ &= (1/N) \sum^{N} (A+Bx_i)/x_i - \sum^{N} (A+Bx_i) / \sum^{N} x_i \\ &= A(\bar{X} - \tilde{X})/\bar{X}\tilde{X} , \end{aligned} \tag{4.9}$$

and hence $a^* = A$. Thus in general when the regression of y on x is linear, a can be interpreted as an approximation to the intercept A. Also

when the regression line passes through a point far from the origin (which corresponds to nonproportionality between y_i and x_i) the value of a^* will tend to be large and there will be sufficient flexibility in the choice of a.

Reddy and Rao (1977) have suggested a change of origin for x in order to improve the proportionality between values of y and x. The suggested estimator is

$$\hat{Y}^*_{pps} = (X'/n) \sum^{n} (y_i/x'_i) , \qquad (4.10)$$

where $x'_i = x_i + d^*\bar{X}$ are the size measures, $X' = \sum^{N} x'_i$ and $d^* = (1-k)/k$. If we are interested in a single y, this manipulation of the size measures works well. However, a transformation of x has the following practical difficulties, unlike that of y :

(i) The transformation is to be made before the sample is drawn.

(ii) The transformed x has to be positive for every unit.

(iii) When multiple characteristics are being estimated from the same sample, a single transformation of x may not suit all the cases.

4.4. Empirical Efficiency of $\hat{Y}^*_1$

Five populations, which are the same as those used by Reddy and Rao (1977), are employed. This allows comparisons.

Population 1: Consists of data on number of workers (x) and output (y) for 80 factories in a certain region (Murthy, 1967, p.228).

Population 2: Consists of data on fixed capital (x) and output (y) for the factories in population 1 .

The other three are the hypothetical populations A, B and C considered by Yates and Grundy (1953), with details as in Table 4.1.

The value of a and n. (estimator variance) for the populations are in Table 4.2.

Table 4.1: Populations A, B and C

Unit	x_i	A	B	C
1	0.1	0.5	0.8	0.2
2	0.2	1.2	1.4	0.6
3	0.3	2.1	1.8	0.9
4	0.4	3.2	2.0	0.8

Table 4.2: Sampling variance for the five populations

Population	a*	$nV(\hat{Y})$	$nV(\hat{Y}_{pps})$	$nV(\hat{Y}^*_{pps})$	$nV(\hat{Y}^*_1)$
1	2810.2763	213×10^8	676×10^8	41×10^8	90×10^8
2	2144.1418	213×10^8	282×10^8	30×10^8	57×10^8
A	- 0.4138	16.360	1.000	0.198	0.172
B	0.4138	3.360	1.000	0.176	0.172
C	0	1.150	0.250	0.282	0.250

We note that for populations A,B the estimator $\hat{Y}_1^*$ performs the best, it is the same as $\hat{Y}_{pps}$ for C, and for populations 1 and 2, it is next only to $\hat{Y}_{pps}^*$. Percentage efficiencies of $\hat{Y}$, $\hat{Y}_{pps}^*$ and $\hat{Y}_1^*$ relative to that of $\hat{Y}_{pps}$ are in Table 4.3, while values of $nV(\hat{Y}_1)$ for typical choices of a for populations 1 and 2 are in Table 4.4.

Table 4.3: Percentage efficiencies of the different estimators

Population	$\hat{Y}_{pps}$	$\hat{Y}$	$\hat{Y}_{pps}^*$	$\hat{Y}_1^*$
1	100	318	1644	753
2	100	133	930	493
A	100	6	505	581
B	100	30	568	581
C	100	22	89	100

Table 4.4: Values of $V(\hat{Y}_1) \times 10^{-8}$ for typical choices of a

a	$V(\hat{Y}_1) \times 10^{-8}$	
	Pop. 1	Pop. 2
0	676	282
500	452	189
1000	333	121
1500	217	77
2000	139	58
2500	97	63
3000	92	93
3500	125	147
4000	195	226

From Table 4.4 we note that there is enough flexibility in the choice of, a . Finally, Table 4.5 shows the performances of $\hat{Y}^*_{pps}$, $\hat{Y}^*_1$, Horvitz-Thompson estimator $\hat{Y}_{HT}$, symmetrized Des Raj estimator $\hat{Y}_{SD}$ and Rao-Hartley-Cochran estimator $\hat{Y}_{RHC}$ in relation to that of $\hat{Y}_{pps}$ for the populations A, B, C and $n = 2$.

Table 4.5: Performances of the different estimators

Estimator	Population A		Population B		Population C	
	Variance	Efficiency	Variance	Efficiency	Variance	Efficiency
$\hat{Y}^*_{pps}$	0.099	505	0.088	568	0.141	89
$\hat{Y}^*_1$	0.081	617	0.081	617	0.125	100
$\hat{Y}_{HT}$	0.823	61	0.057	877	0.059	212
$\hat{Y}_{SD}$	0.333	150	0.333	150	0.083	151
$\hat{Y}_{RHC}$	0.333	150	0.333	150	0.083	151
$\hat{Y}_{pps}$	0.500	100	0.500	100	0.125	100

It is seen that $\hat{Y}^*_1$ is the best (among the estimators considered) for population A and does better than $\hat{Y}^*_{pps}$ in each of the three cases. These empirical studies, though of limited scope, throw some light on the relative performances of the estimators considered.

4.5 PPSWOR Sampling

Consider the Midzuno-Sen scheme of varying probability sampling using x as the measure of size. The first sample unit is selected with probability proportional to the value of x and the subsequent units are selected with equal probability and without replacement. Under this scheme the probability of selecting a specified sample is proportional to the total size of the units included in the sample. The probability of including U_i in a sample of size n is

$$\pi_i = \frac{(N-n)x_i}{(N-1)X} + \frac{(n-1)}{(N-1)} , \tag{4.11}$$

while the probability of including both U_i and U_j $(i \neq j)$ is

$$\pi_{ij} = \frac{(n-1)}{(N-1)(N-2)} \{(N-n)(x_i+x_j)X^{-1} + (n-2)\} . \tag{4.12}$$

Then the Horvitz-Thompson estimator of Y is

$$\hat{Y}_{HT} = \sum^{n} (y_i/\pi_i) ,$$

which is unbiased and

$$V(\hat{Y}_{HT}) = \sum_{i>j=1}^{N} (\pi_i\pi_j - \pi_{ij})(y_i/\pi_i - y_j/\pi_j)^2 . \tag{4.13}$$

An advantage with the Midzuno-Sen scheme is that the Yates-Grundy estimate of $V(\hat{Y}_{HT})$ is never negative.

4.6 A Modification to $\hat{Y}_{HT}$

Expression (4.13) shows that $V(\hat{Y}_{HT})$ will be small if y_i/π_i is

nearly the same for all the population units. However, even when $y_i = Bx_i$ for all i, y_i/π_i will not be a constant because of the second term on the right-hand side of (4.11). On the other hand, $[y_i + BX(n-1)/(N-n)]\pi_i^{-1}$ will be a constant. Motivated by this, consider the transformation

$$z_i = y_i + (n-1)b/(N-n) , \quad (i=1,\ldots,N) \tag{4.14}$$

where b is a scalar to be chosen. Then an unbiased estimator of Y is

$$\hat{Y}_2 = \sum^{n} (z_i/\pi_i) - N(n-1)b/(N-n) . \tag{4.15}$$

It is seen that

$$\begin{aligned} V(\hat{Y}_2) &= \sum_{i>j=1}^{N} (\pi_i\pi_j - \pi_{ij})(z_i/\pi_i - z_j/\pi_j)^2 \\ &= \sum_{i>j=1}^{N} (\pi_i\pi_j - \pi_{ij})\{y_i/\pi_i - y_j/\pi_j) + \frac{(n-1)b}{N-n}(1/\pi_i - 1/\pi_j)\}^2 \\ &= V(\hat{Y}_{HT}) + b^2D_1 + 2bD_2 , \end{aligned} \tag{4.16}$$

where

$$D_1 = \{(n-1)/(N-n)\}^2 \sum_{i>j=1}^{N} (\pi_i\pi_j - \pi_{ij})(1/\pi_i - 1/\pi_j)^2 ;$$

$$D_2 = \{(n-1)/(N-n)\} \sum_{i>j=1}^{N} (\pi_i\pi_j - \pi_{ij})(y_i/\pi_i - y_j/\pi_j)(1/\pi_i - 1/\pi_j) .$$

And $V(\hat{Y}_2)$ is minimized for the choice $b_{opt} = -D_2/D_1$ with corresponding $V_{min}(\hat{Y}_2) = V(\hat{Y}_{HT}) - D_2^2/D_1$.

Result (4.16) demonstrates that $\hat{Y}_2$ will have a variance smaller than that of $\hat{Y}_{HT}$ when the last two terms on the right-hand side add up to a negative number. That is $b^2D_1 + 2bD_2 < 0$. Since $D_1 > 0$, this will happen if and only if the roots of the equation $b^2D_1 + 2bD_2 = 0$ are real and distinct and b lies between them. Here these roots are real and distinct, and they are $b = 0$ and $b = -2D_2/D_1 = 2b_{opt}$. Hence for $\hat{Y}_2$ to be more efficient than $\hat{Y}_{HT}$ we obtain the condition that

b lies between 0 and $2b_{opt}$.

Thus when D_2 is large relative to D_1, b_{opt} will be large and hence there will be considerable flexibility in choosing b such that $\hat{Y}_2$ is more efficient than $\hat{Y}_{HT}$. Note that D_2 is a quantity that reflects the nonconstancy of y_i/π_i.

4.7 Choice of b

Since in practice y_i is not known for all the units in the population, b_{opt} can not be computed. However, a reasonable choice of b_{opt} can be made. Two methods for doing this are outlined below.

Method I. Consider the case $y_i = Bx_i$ for all i. Then

$$D_2 = \{(n-1)/(N-n)\} \sum_{i>j=1}^{N} (\pi_i\pi_j - \pi_{ij})(y_i/\pi_i - y_j/\pi_j)(1/\pi_i - 1/\pi_j)$$

$$= B\{(n-1)/(N-n)\} \sum_{i>j=1}^{N} (\pi_i\pi_j - \pi_{ij})(x_i/\pi_i - x_j/\pi_j)(1/\pi_i - 1/\pi_j)$$

$$= - BX\{(n-1)/(N-n)\}^2 \sum_{i>j=1}^{N} (\pi_i\pi_j - \pi_{ij})(1/\pi_i - 1/\pi_j)^2 ,$$

since, in view of the expression (4.11) for the first order inclusion probabilities, we have

$$(x_i/\pi_i - x_j/\pi_j) = - \{(n-1)/(N-n)\}(1/\pi_i - 1/\pi_j)X .$$

Hence b_{opt} reduces to BX in this case and the range for a suitable b is 0 to $2BX$. An estimate of B may be obtained by plotting y_i/π_i against x_i/π_i for the sample units and gauging the slope of the best fitting line.

Method II. Computing b with the summations in the expressions for D_1 and D_2 taken over only the sample units and the terms weighted by $1/\pi_{ij}$. That is

$$b = - \left[\frac{N-n}{n-1}\right] \frac{\Sigma(\pi_i\pi_j - \pi_{ij})(y_i/\pi_i - y_j/\pi_j)(1/\pi_i - 1/\pi_j)\pi_{ij}^{-1}}{\Sigma(\pi_i\pi_j - \pi_{ij})(1/\pi_i - 1/\pi_j)^2 \pi_{ij}^{-1}} ,$$

where the summations are over $i > j = 1,\ldots,n$. This method computes b from current data. The sampling variability introduced by this factor is not investigated here.

4.8 Empirical Study

To compare the efficiency of the estimator $\hat{Y}_2$ with that of the usual estimator under the Midzuno-Sen scheme, five populations are considered.

Population I consists of the number of cattle, y, and the number of farms, x, in 13 village blocks as noted in the Dakshina Kannada district of Karnataka during the survey discussed in Part Two of the thesis. Details are in Table 4.6. Population II is the population A considered by Sampford (1978, p.37, Table 2), where the concepts of predictive estimation and internal congruency are presented. The other three are populations A, B and C considered by Yates and Grundy (1953), as in Table 4.1.

Table 4.6: Number of cattle, y, and the number of farms, x, in 13 village blocks of the Dakshina Kannada district of Karnataka (Popn.I).

x :	19	28	28	30	31	46	51	53	55	56	61	64	83
y :	168	326	396	360	331	697	586	739	914	930	619	784	906

For the above populations, values of b_{opt}, $V(\hat{Y}_{HT})$, $V(\hat{Y}_2)$ and percentage efficiency of $\hat{Y}_2$ relative to $\hat{Y}_{HT}$, that is $100V(\hat{Y}_{HT})/V(\hat{Y}_2)$, are given in Table 4.7.

Table 4.7: Efficiency of $\hat{Y}_2$

Population	Sample Size	b_{opt}	$V(\hat{Y}_{HT})$	$V(\hat{Y}_2)$	Efficiency (%)
I	2	8434.8803	2227.56×10^3	932.27×10^3	239
II	4	303.5651	10.96×10^3	6.20×10^3	177
A	2	9.6781	2.8841	0.0510	5665
B	2	3.3324	0.3839	0.0530	724
C	2	2.2959	0.2416	0.0821	294

These empirical studies indicate that $\hat{Y}_2$ can be considerably more efficient than $\hat{Y}_{HT}$. For population II the variance of the biased internally congruent estimator suggested by Sampford (1978) was computed and found to be 6948 which is higher than $V(\hat{Y}_2)$. Finally Table 4.8 shows the effect of deviations from the optimum value of b on the efficiency of $\hat{Y}_2$.

Table 4.8: Sensitivity of efficiency of $\hat{Y}_2$ to departures from the optimum choice of b.

$100\|1 - b/b_{opt}\|$	Value of efficiency for populations				
	I	II	A	B	C
0	239	177	5655	724	294
20	226	171	1755	586	273
40	200	160	644	389	231
60	159	139	269	224	173
80	126	118	152	143	130
100	100	100	100	100	100

4.9 Ratio Method of Estimation

Let $\hat{Y}$, $\hat{X}$ be unbiased estimators of Y and X, based on any probability sampling design. Consider the following modification to the ratio estimator $\hat{Y}_r = \hat{Y}(X/\hat{X})$.

$$\hat{Y}_3 = (\hat{Y}-Na)(X/\hat{X}) + Na \quad . \tag{4.17}$$

The bias and mse of $\hat{Y}_3$ are given approximately by

$$B(\hat{Y}_3) = Y(\Theta - k)V_{02} \tag{4.18}$$

$$M(\hat{Y}_3) = Y^2(V_{20} - 2\Theta V_{11} + \Theta^2 V_{02}) \tag{4.19}$$

where the V_{ij} are the relative central moments defined in (2.4), $\Theta = 1 - Na/Y$, $\Theta_{opt} = V_{11}/V_{02} = k$, and the corresponding $a_{opt} = (1-k)\bar{Y}$. For this choice of a, $B(\hat{Y}_3) = 0$. In general there will be reduction in absolute bias relative to that of $\hat{Y}_r$ when $|\Theta - k| < |1-k|$, that is, k is closer to Θ than to 1.

In order to have an increase in precision relative to the estimators $\hat{Y}$ or $\hat{Y}_r$ it is necessary and sufficient that Θ lies between 0 and $2k$, and between $2k-1$ and 1. The situation is similar to that in Sec. 3.2. The implication of these conditions on the choice of a are given for typical values of k in Table 4.9.

Table 4.9: Optimum a and lower and upper bounds on the choice of a for typical values of k.

k	Lower bound on a	Optimum a	Upper bound on a
0.1	$0.8\bar{Y}$	$0.9\bar{Y}$	$\bar{Y}$
0.2	$0.6\bar{Y}$	$0.8\bar{Y}$	$\bar{Y}$
0.3	$0.4\bar{Y}$	$0.7\bar{Y}$	$\bar{Y}$
0.4	$0.2\bar{Y}$	$0.6\bar{Y}$	$\bar{Y}$
0.5	0	$0.5\bar{Y}$	$\bar{Y}$
0.6	0	$0.4\bar{Y}$	$0.8\bar{Y}$
0.7	0	$0.3\bar{Y}$	$0.6\bar{Y}$
0.8	0	$0.2\bar{Y}$	$0.4\bar{Y}$
0.9	0	$0.1\bar{Y}$	$0.2\bar{Y}$
1.0	0	0	0
1.2	$-0.4\bar{Y}$	$-0.2\bar{Y}$	0
1.4	$-0.8\bar{Y}$	$-0.4\bar{Y}$	0
1.6	$-1.2\bar{Y}$	$-0.6\bar{Y}$	0
1.8	$-1.6\bar{Y}$	$-0.8\bar{Y}$	0
2.0	$-2.0\bar{Y}$	$-1.0\bar{Y}$	0

This table indicates that there is reasonable flexibility in the choice of a, except when k is close to 0 or 1. In these cases the simple estimator $\hat{Y}$ and the ratio estimator $\hat{Y}_r$ may be used respectively.

4.10 Difference Method of Estimation

Define $\hat{W} = \hat{X}/\bar{X}$, $\hat{Z} = \hat{Y}/Y_k$ where Y_k is a scalar to be chosen. Consider the difference estimator

$$\hat{Y}_4 = Y_k[(\hat{Z}-\hat{W}) + N] . \tag{4.20}$$

This is unbiased for Y since

$$E(\hat{Y}_4) = EY_k[(\hat{Z}-\hat{W}) + N] = Y_k[(Y/Y_k - X/\bar{X}) + N] = Y .$$

Further

$$V(\hat{Y}_4) = Y^2\{V_{20} - 2\Theta V_{11} + \Theta^2 V_{02}\} , \tag{4.21}$$

with $\Theta = Y_k/\bar{Y}$. The restrictions on Θ in order to have increased precision relative to that of the ratio estimator are the same as those in the previous section. The implications of these on the choice of Y_k are summarized in Table 4.10.

Table 4.10: Optimum Y_k and lower and upper bounds on the choice of Y_k for typical values of k .

k	Lower bound on Y_k	Optimum Y_k	Upper bound on Y_k
0.1	0	$0.1\bar{Y}$	$0.2\bar{Y}$
0.2	0	$0.2\bar{Y}$	$0.4\bar{Y}$
0.3	0	$0.3\bar{Y}$	$0.6\bar{Y}$
0.4	0	$0.4\bar{Y}$	$0.8\bar{Y}$
0.5	0	$0.5\bar{Y}$	$1.0\bar{Y}$
0.6	$0.2\bar{Y}$	$0.6\bar{Y}$	$\bar{Y}$
0.7	$0.4\bar{Y}$	$0.7\bar{Y}$	$\bar{Y}$
0.8	$0.6\bar{Y}$	$0.8\bar{Y}$	$\bar{Y}$
0.9	$0.8\bar{Y}$	$0.9\bar{Y}$	$\bar{Y}$
1.0	$1.0\bar{Y}$	$1.0\bar{Y}$	$\bar{Y}$
1.2	$\bar{Y}$	$1.2\bar{Y}$	$1.4\bar{Y}$
1.4	$\bar{Y}$	$1.4\bar{Y}$	$1.8\bar{Y}$
1.6	$\bar{Y}$	$1.6\bar{Y}$	$2.2\bar{Y}$
1.8	$\bar{Y}$	$1.8\bar{Y}$	$2.6\bar{Y}$
2.0	$\bar{Y}$	$2.0\bar{Y}$	$3.0\bar{Y}$

Again there is sufficient flexibility in the choice of Y_k except when k is close to 0 or 1 . The particular case of srswor design for the estimators $\hat{Y}_3$, $\hat{Y}_4$ has been discussed by Srivenkataramana (1978), and Srivenkataramana and Srinath (1982).

CHAPTER 5.

DOUBLE SAMPLING WITH PPS SELECTION

[Raj (1965) has proposed a pps selection scheme for estimating the population total in the absence of information on the auxiliary variate, but when information on some other variate is available. This requires the knowledge of a certain parameter h. But in practice the value of h may not easily be known. Accordingly this chapter proposes an alternative scheme which does not need the knowledge of h.]

5.1 Introduction

Let y and x be the study and auxiliary variates; (y_i, x_i), $i = 1,\ldots,N$ being real values, which are however unknown. Let z be another variate for which the value is known for each unit in the population. It is assumed that $x_i, z_i > 0$ for $i = 1,\ldots,N$. For instance, in estimating the total yield of wheat in a village having N farms, the yield and area under the crop in each farm are likely to be unknown. But the total area in each farm may be known from village records or may be obtained at a low cost. Then y, x and z are respectively yield, area under wheat and area under cultivation.

In this context Raj (1965) has proposed the following scheme.

Scheme I

Suppose an initial sample of n' units is selected with probabilities $p_i = z_i/Z$; $i = 1,\ldots,N$, where $Z = \sum_{i}^{N} z_i$, and information on x is collected. And a subsample of size n is selected from the initial sample with equal probabilities (wor) and information on y is collected.

Theorem 5.1 Under scheme I, if

$$\hat{Y} = (1/n) \sum^{n} (y_i/p_i) - (h/n) \sum^{n} (x_i/p_i) + (h/n') \sum^{n'} (x_i/p_i) \tag{5.1}$$

then $\hat{Y}$ is unbiased for Y , and

$$V(\hat{Y}) = S_1^2/n + (1/n - 1/n') (h^2S_1^2 - 2hdS_1S_2)$$

where

$$S_1^2 = \sum^{N} (y_i/p_i - Y)^2 p_i ,$$

$$S_2^2 = \sum^{N} (x_i/p_i - X)^2 p_i ,$$

$$d = [\sum^{N} (y_i/p_i - Y)(x_i/p_i - X) p_i]/S_1S_2$$

and h is a scalar. $V(\hat{Y})$ is a minimum when $h_{opt} = dS_1/S_2$, and

$$\min V(\hat{Y}) = S_1^2 (1-d^2)/n + S_1^2d^2/n' . \tag{5.2}$$

5.2 An Alternative Scheme

In order to compute $\hat{Y}$ from (5.1), we have to guess the value of h .

When this is difficult the following alternative scheme is suggested which avoids the necessity of h.

Scheme II

Select the initial sample, s', of size n' with probability proportional to z and collect information on x as in scheme I. And then select a subsample, s, of size n with probability proportional to x_i/z_i, with replacement.

Theorem 5.2 Under scheme II, if $Z = \sum^{N} z_i$, and

$$\hat{Y}^* = (Z/nn') \sum^{n} (y_i/x_i) \cdot \sum^{n'} (x_i/z_i) \tag{5.3}$$

then $\hat{Y}^*$ is unbiased for Y, and

$$V(\hat{Y}^*) = S_1^2/n' + S_3^2 (n'-1)/nn' \tag{5.4}$$

where

$$S_3^2 = \sum^{N'} (y_i X/x_i - Y)^2 \, x_i/X .$$

An unbiased estimate of $V(\hat{Y}^*)$ is provided by

$$v(\hat{Y}^*) = (\sum^{n} w_i^2/p_i^* - nn' \hat{Y}^{*2})/nn'(n'-1) + \sum^{n} (w_i/p_i^* - \sum^{n} w_i/np_i^*)^2/(n-1)nn'(n'-1) \tag{5.5}$$

where $w_i = y_i/p_i$ and $p_i^* = x_i/z_i \sum^{n'} (x_i/z_i)$.

Proof: Consider $\hat{Y}^* = (Z/nn') \sum^{n} (y_i/x_i) \cdot \sum^{n'} (x_i/z_i)$

$$= (\frac{Z}{nn'}) \sum^{n} \{(\frac{y_i}{z_i})(\frac{z_i}{x_i})\} \cdot \sum^{n'} (\frac{x_i}{z_i})$$

$$= \frac{1}{nn'} \sum^{n} \{(\frac{y_i}{p_i})(\frac{z_i}{x_i})\} \cdot \sum^{n'} (\frac{x_i}{z_i}), \quad \text{since } \frac{z_i}{Z} = p_i$$

$$= (1/nn') \sum^{n} (w_i/p_i^*)$$

where $w_i = y_i/p_i$ and $p_i^* = x_i/\{z_i \sum^{n'} (x_i/z_i)\}$.

Therefore,

$$E(\hat{Y}^*) = E_1E_2(\hat{Y}^*|s')$$

$$= E_1(\sum^{n'} w_i)/n' = E_1(\sum^{n'} y_i/p_i)/n' = \sum^{N} y_i = Y .$$

Hence $\hat{Y}^*$ is unbiased for Y . Next recall the standard result

$$V(\hat{Y}^*) = V_1E_2(\hat{Y}^*|s') + E_1V_2(\hat{Y}^*|s') .$$

Consider

$$V_1E_2(\hat{Y}^*|s') = [\sum^{N}(y_i/p_i - Y)^2 p_i]/n' = S_1^2/n' , \qquad (5.6)$$

and

$$V_2(\hat{Y}^*|s') = [\sum^{n'} (w_i/p_i^* - \sum^{n'} w_i)^2 p_i^*]/n'^2n$$

$$= (1/nn'^2)[\sum^{n'} w_i^2/p_i^* - (\sum^{n'} w_i)^2] .$$

Writing

$$\sum^{n'} w_i^2/p_i^* = Z^2 \sum^{n'} (y_i^2/x_iz_i) \sum^{n'} (x_i/z_i)$$

$$= \sum_{i=1}^{n'} \sum_{j\neq i}^{n'} (y_i^2/x_ip_i)(x_j/p_j) + \sum_{i=1}^{n'} y_i^2/p_i^2 ,$$

and $\sum^{n'} w_i = n' E_2(\hat{Y}^*)$ we have

$$E_1V_2(\hat{Y}^*|s') = \{n'(n'-1) \sum_{i=1}^{N} \sum_{j=1}^{N} (y_i^2x_j)/x_i$$

$$+ n' \sum_{i=1}^{N} (y_i^2/p_i) - n'^2Y - n'^2V_1E_2(\hat{Y}^*)\}/nn'^2$$

$$= \{n'(n'-1) \sum_{i=1}^{N} (Xy_i^2/x_i) + n' \sum_{i=1}^{N} (y_i^2/p_i) - n'^2Y^2$$

$$- n' \sum_{i=1}^{N} (y_i^2/p_i) + n'Y^2\}/nn'^2$$

$$= (n'-1) S_3^2/nn' . \qquad (5.7)$$

Hence from (5.6) and (5.7), $V(\hat{Y}^*) = S_1^2/n' + (n'-1)\, S_3^2/nn'$, which establishes (5.4).

Next we have to show that the expression given in (5.5) is an unbiased estimator of $V(\hat{Y}^*)$. Denote the first term on the right side of (5.5) by v_1 and the second by v_2 . Then .

$$E(v_2) = E(\sum^{n} w_i^2/p_i^* - nn'\hat{Y}^{*2})/nn'(n'-1)$$

$$= \{E(\sum^{n} w_i^2/p_i^*\, nn') - E(\hat{Y}^{*2})\}/(n'-1)$$

$$= \{\sum^{N} y_i^2/p_i - Y^2 - V(\hat{Y}^*)\}/(n'-1) \text{ , using } V(\hat{Y}^*) = E(\hat{Y}^{*2}) - [E(\hat{Y}^*)]^2$$

$$= \{S_1^2 - V(\hat{Y}^*)\}/(n'-1)$$

$$= S_1^2/n' - S_3^2/nn' \text{ .} \qquad (5.8)$$

Finally, $V_2(\hat{Y}^*|s')$ is estimated by

$$T = \sum^{n}(w_i/p_i^* - \sum^{n} w_i/np_i^*)^2/n(n-1)n'^2$$

and therefore T also estimates $E_1V_2(\hat{Y}^*|s') = (n'-1)S_3^2/nn'$. Hence

$$v_2 = n'T/(n'-1) \quad \text{estimates} \quad S_3^2/n \text{ .} \qquad (5.9)$$

From (5.8) and (5.9) it follows that $v_1 + v_2$ provides an unbiased estimator of $V(\hat{Y}^*)$. This completes the proof.

5.3 Efficiency of $\hat{Y}^*$

Consider

$$E = \min V(\hat{Y})/V(\hat{Y}^*) = \{(1-d^2)n' + d^2n\}/\{n + (n'-1)D\}$$

where $D = S_3^2/S_1^2$. This reflects the efficiency of $\hat{Y}^*$ over the estimator $\hat{Y}$.

In particular when y_i/x_i is a constant for all the population units, $S_3^2 = 0$ (hence $D = 0$) and $d = 1$ so that $E = 1$. In general denoting n'/n by r, the expression for E can be written as

$$E = \{(1-d^2)rn + d^2n\} / \{n + (rn-1)D\}$$

$$\approx \{(1-d^2)r + d^2\} / (1+rD) . \qquad (5.10)$$

The values of E (%) computed from (5.10) for typical values of r and D are given in Table 5.1.

Table 5.1: Values of E for typical values of r, D and d.

	r = 5					r = 10				
d \ D	0.9	0.8	0.7	0.6	0.5	0.9	0.8	0.7	0.6	0.5
0.1	1.17	1.63	2.03	2.37	2.67	1.35	2.12	2.79	3.38	3.88
0.2	-	1.22	1.52	1.78	2.00	-	1.41	1.86	2.25	2.58
0.3	-	-	1.22	1.42	1.60	-	1.06	1.40	1.69	1.94
0.4	-	-	1.01	1.19	1.33	-	-	1.12	1.35	1.55
0.5	-	-	-	1.02	1.14	-	-	-	1.13	1.30
0.6	-	-	-	-	1.00	-	-	-	-	1.11

- denotes values of E less than 1.

For small values of d and D, the estimator $\hat{Y}^*$ performs well. Precisely in this case guessing h_{opt} is difficult and $\hat{Y}^*$ can be used rather than $\hat{Y}$.

5.4 Cost Function

If c' and c denote the unit costs of collecting information on

x and y respectively, the total cost under scheme I or II would be

$$C = c'n' + cn \quad . \tag{5.11}$$

On the other hand if a straight probability proportional to z sample is taken for y (Scheme III) the sample size for the same total cost will be $n_o = (c'n' + cn)/c$ and variance of the estimator of the population total will be

$$S_1^2/n_o \quad . \tag{5.12}$$

Comparing (5.4) with (5.12), the condition that scheme II is better than scheme III under cost function (5.11) becomes

$$D < \left\{\frac{n}{n' - 1}\right\} \left\{\frac{c(n'-n) - c'n'}{c'n' + cn}\right\} \quad .$$

Roughly this reduces to

$$D < \frac{r(a-1) - a}{r^2 + ra} \quad , \text{ where } \quad a = c/c' \quad .$$

As an example, let $r = 10$, $a = 40$ then D should be less than 0.7, that is $S_3^2 < 0.7S_1^2$.

Numerical illustration. In order to compare the efficiency of $\hat{Y}^*$ with that of $\hat{Y}$ computed from (5.1) with h_{opt}, the data relating to 34 villages given in Murthy (1967) were treated as making up the population, where y, x and z respectively denote area under wheat in 1964, in 1963 and cultivated area in 1961. For this population the values of the required parameters are given in Table 5.2. The values of E (%) for typical values of [illegible] given in Table 5.3.

Table 5.2: Parameters of the population in the illustration

Parameter	Value
S_1^2	4421.81×10^3
S_3^2	826.06×10^3
S_2^2	2165.27×10^3
d	0.6404
D	0.1868

Table 5.3: Values for E for typical r for the population in the illustration

r	E (%)
2	116
3	140
4	159
5	174
6	186

It is seen that E increases as subsampling rate decreases. Therefore the proposed scheme will be advantageous especially when the information on y is expensive as compared to information on x .

CHAPTER 6.

USE OF A POSITIVE AND NEGATIVE VALUED AUXILIARY VARIATE IN SURVEYS

[Auxiliary variates in surveys are occasionally positive and negative valued. This introduces difficulties in ratio and product methods of estimation and in pps sampling. Two simple methods of dealing with the situation are outlined: stratification by sign and transformation by simple translation. For illustration purpose srswor and ppswr designs are assumed.]

6.1 Introduction

In Chapters 2 to 5 it was mostly assumed that the auxiliary variate was non-negative since survey variates are generally so. But exceptions do occur. Rates of change, elasticities of supply and demand, and variates arising as differences, for example, saving (= income-expenditure) and profit (= revenue - cost) are some cases in point. Such variates can not be used directly as (i) the auxiliary variate in ratio and product methods of estimation, or (ii) the measure of size in pps sampling. The former is due to that the population mean and/or the sample mean of the variate may be close to zero which could easily happen when the distribution of the variate in the population is nearly symmetric.

The other difficulty is because of the non-positive values. This chapter suggests (i) a stratification of the population according to the sign of the value of the auxiliary variate, and as an alternative (ii) a change of origin for the auxiliary variate to make it positive valuedand a corresponding change for the study variate in order to have proportionality between the values of the two variates.

6.2 The Stratification Approach

The values x_i, y_i are assumed to be real, but they may be positive, negative or zero. The x_i are known. Consider a stratification of the population according to the sign of x_i. That is, unit i is allotted to stratum 1 if $x_i \geq 0$; otherwise to stratum 2. The sample size n is allocated to these strata, for example, by proportional allocation. Let n_h denote the size of the sample and $\bar{x}_h$, $\bar{y}_h$ denote the sample means for stratum h under a srswor scheme. Assume that the selections from the two strata are independent. Let $\bar{X}_h$, $\bar{Y}_h$ be the means of x and y in stratum h. Then an estimator of Y_h, the stratum total for y, is

$$t_h = N_h \bar{y}_h \, (\bar{X}_h / \bar{x}_h) .$$

And Y may be estimated by

$$\hat{Y}_1 = t_1 + t_2 . \tag{6.1}$$

This is the usual 'separate' ratio estimator. If the n_h are at least moderate then the mse of t_h is given approximately by

$$M(t_h) = N_h(N_h - n_h)(S_{yh}^2 + R_h^2 S_{xh}^2 - 2R_h S_{yxh})/n_h \tag{6.2}$$

where S_{yh}^2 , S_{xh}^2 , S_{yxh} are variances and covariance in stratum h and $R_h = \bar{Y}_h/\bar{X}_h$. And

$$M(\hat{Y}_1) = M(t_1) + M(t_2) \ . \tag{6.3}$$

It is known that a ratio estimator t_h performs well if $(B_h/R_h) > 1/2$ where B_h is the coefficient of regression of y on x in stratum h . It is assumed above that this is in fact the case. On the other hand if $(B_h/R_h) < -1/2$ a product-type estimator $t_h = N_h\bar{y}_h(\bar{x}_h/\bar{X}_h)$ may be formed in that stratum. And if B_h/R_h is in $[-1/2, 1/2]$ the simple estimator $N_h\bar{y}_h$ itself may be used as t_h . Of course, the expression for $M(t_h)$ in (6.2) should be modified accordingly. Thus the stratification approach allows a choice among ratio, product and simple estimators in each of the two strata. This is handy when the nature and extent of dependence of y on x differ in the two cases: $x_i \geq 0$ and $x_i < 0$.

If stratification by sign is difficult, a strategy of post stratification might be appropriate, particularly as the problem of zero stratum sample sizes is unlikely to occur.

6.3 A Change of Origin

As an alternative to stratification consider the following approach. Suppose that a srswor of n units is drawn from the entire population, the value of x and y obtained for these units and the sample means $\bar{x},\bar{y}$ computed. Then as an estimator of Y the ratio method uses

$$\hat{Y}_r = N\bar{y}(\bar{X}/\bar{x}) \ , \tag{6.4}$$

while the product method uses

$$\hat{Y}_p = N\bar{y}(\bar{x}/\bar{X}) \ . \tag{6.5}$$

The estimator $\hat{Y}_r$ is an improvement over the simple estimator $N\bar{y}$ if $(B/R) > 1/2$, where B is the coefficient of regression of y on x and $R = \bar{Y}/\bar{X}$. The estimator $\hat{Y}_p$ may be used when $B/R < -1/2$. And when B/R is in $[-1/2, 1/2]$ the estimator $N\bar{y}$ itself is preferred. However, it is apparent from (6.4) and (6.5) that the use of $\hat{Y}_r$ or $\hat{Y}_p$ should be avoided when $\bar{X}$ and/or $\bar{x}$ is likely to be close to zero. This can easily be the case when x assumes both positive and negative values.

In this context consider the transformation

$$w = x + c \ , \quad z = y + \Theta \ . \tag{6.6}$$

The scalar c is chosen such that w is a positive valued variate. For instance, let x_1 be the smallest x-value and assume that it is negative. Then c must satisfy $|x_1| < c < \infty$. Next, the choice of Θ is made so as to control the mse of the estimator. This is discussed in Sec. 6.4.

Now let $\bar{w} = \bar{x} + c$, $\bar{z} = \bar{y} + \Theta$, $\bar{W} = \bar{X} + c$, $\bar{Z} = \bar{Y} + \Theta$. Then the usual ratio estimator of the total Z is

$$\hat{Z}_r = N\bar{z}\ (\bar{W}/\bar{w}) = N(\bar{y}+\Theta)\{(\bar{X}+c)/(\bar{x}+c)\} \ , \tag{6.7}$$

and hence Y may be estimated by

$$\hat{Y}_2 = \hat{Z}_r - N\Theta \ . \tag{6.8}$$

The transformations (6.6), being only changes of origin, leave the

variances, covariances and correlations unchanged. Also the standard theory for ratio estimators applies to $\hat{Z}_r$, except that in the place of R we now have $R_1 = (\bar{Y}+\Theta)/(\bar{X}+c)$. Thus the bias and mse of $\hat{Y}_2$ as an estimator of Y are, up to second order moments

$$B(\hat{Y}_2) = (N-n)\ (R_1 S_x^2 - S_{xy})/n(\bar{X}+c)\ , \tag{6.9}$$

and

$$M(\hat{Y}_2) = N(N-n)(S_y^2 + R_1^2 S_x^2 - 2R_1 S_{yx})/n\ , \tag{6.10}$$

where S_y^2 is the population variance of y, etc.

In order to use $\hat{Y}_2$ in practice we expect that it is more precise than $N\bar{y}$. This will be the case when $(B/R_1) > 1/2$. Also the introduction of the parameter Θ will be justified only if the estimator is more precise than that without Θ, that is than that with $\Theta = 0$. It can be shown that these two considerations require that Θ lies

$$\text{between } -\bar{Y} \text{ and } D - \bar{Y}\ , \tag{6.11}$$

and

$$\text{between } 0 \text{ and } D - 2\bar{Y}\ , \tag{6.12}$$

where $D = 2B\ (\bar{X}+c)$. The second condition ensures also a simultaneous reduction in the absolute bias relative to the $\Theta = 0$ case. For a given c, the value of Θ minimizing $M(\hat{Y}_2)$ is provided by $R_1 = B$ which implies $\Theta_{opt} = (D/2) - \bar{Y}$. In this case $\hat{Y}_2$ is almost unbiased for Y. And

$$M(\hat{Y}_2) = N(N-n)\ S_y^2\ (1 - \rho^2)/n\ .$$

This mse is the same as the large sample approximation to the mse of

the linear regression estimator. In the special case $y_i = Bx_i$ for all i, $\Theta_{opt} = Bc$ and $\hat{Z}_r$ reduces to $NB(\bar{X}+c)$ so that $\hat{Y}_2 = NB\bar{X}$. Hence $\hat{Y}_2$ estimates $Y = NB\bar{X}$ without any error just like $\hat{Y}_r$ in this situation.

6.4 The Choice of Θ

The optimum Θ is $B(\bar{X}+c) - \bar{Y}$. Here $\bar{X}$ and c are known, but generally B and $\bar{Y}$ are not. A geometric interpretation of Θ_{opt} is that it is the distance between the x and w axes such that the population regression of z on w is through the origin (See fig. 6.1). Past experience and data from a pilot study may be used in finding approximations to Θ_{opt}, as commonly done in such situations. Also since Θ is needed only at the estimation stage, the knowledge of $\bar{y}$ and a scatter diagram for at least a part of the sample data on y and x may be helpful in this regard. In general (6.11) and (6.12) specify two intervals in which Θ should lie in order that $\hat{Y}_2$ is more precise than $N\bar{y}$ or $\hat{Y}_2$ with $\Theta = 0$. The midpoint of these intervals is Θ_{opt}, and the widths are $|D|$ and $|D-2\bar{Y}|$ respectively.

If $(B/R_1) \leq -1/2$, a product-type estimator

$$\hat{Y}_3 = N(\bar{y}+\Theta)\{(\bar{x}+c)/(\bar{X}+c)\} - N\Theta \qquad (6.13)$$

may be used. Here $\Theta_{opt} = -B(\bar{X}+c) - \bar{Y}$.

6.5 PPS Sampling

Again consider the transformation (6.6), which in the first place makes w a positive valued variate. Therefore it can be used as a

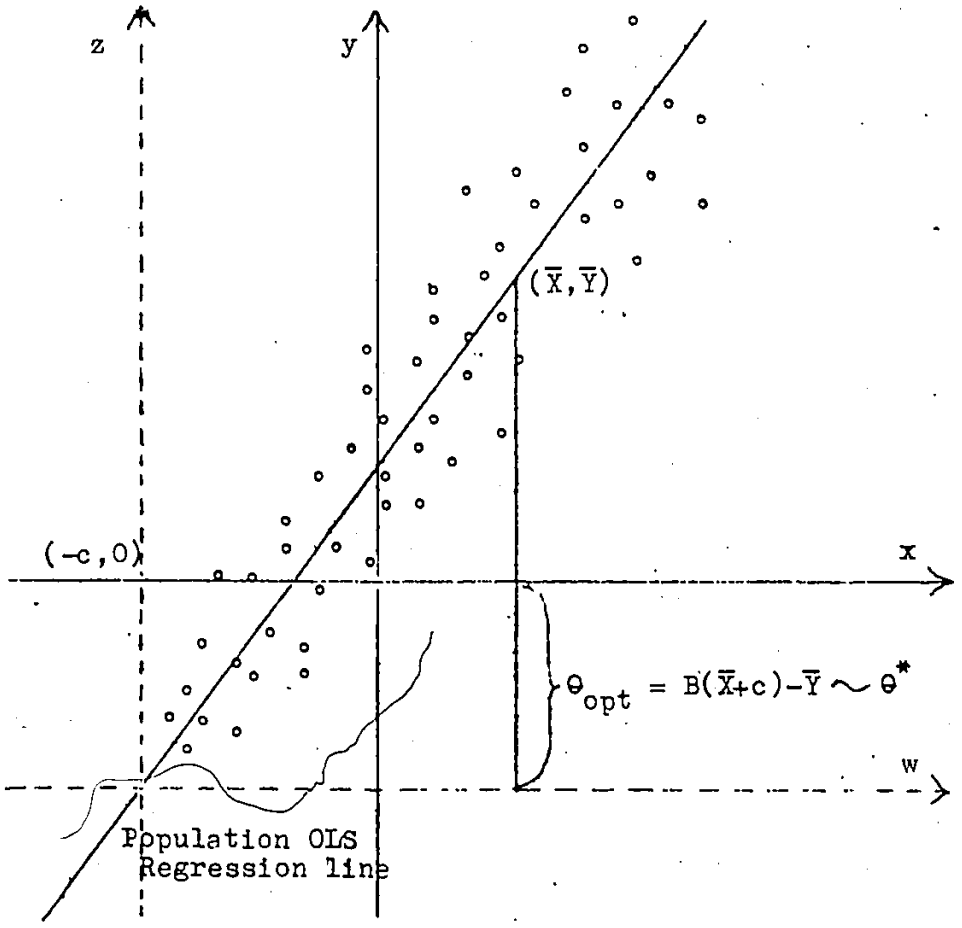


Fig. 6.1 Geometric interpretation of Θ_{opt}

measure of size. With this suppose a ppswr sample of n units is drawn. Then an unbiased estimator of Z is

$$\hat{Z}_{pps} = (W/n) \sum^{n} (z_i/w_i)$$

and hence an unbiased estimator of Y is

$$\hat{Y}_4 = (W/n) \sum^{n} (z_i/w_i) - N\Theta$$

$$= \{N(\bar{X}+c)/n\} \sum^{n} \{(y_i+\Theta)/(x_i+c)\} - N\Theta \; . \qquad (6.14)$$

And

$$V(\hat{Y}_4) = \frac{N(\bar{X}+c)}{n} \sum^{N} \{(y_i+\Theta)^2/(x_i+c)\} - \frac{N^2}{n} (\bar{Y}+\Theta)^2 \; .$$

It is natural to expect that $\hat{Y}_4$ should be more precise than either $N\bar{y}$ or $\hat{Y}_4$ with $\Theta = 0$. For a given c, the former requires (Raj, 1968, p.50)

$$\sum^{N} (w_i - \bar{W}) z_i^2/w_i > 0 \qquad (6.15)$$

while the latter needs

$$\Theta \text{ to lie between } 0 \text{ and } 2\Theta^* \; . \qquad (6.16)$$

where

$$\Theta = \bar{W} \tilde{W} (R_o - \bar{R}_o)/(\bar{W} - \tilde{W}) \qquad (6.17)$$

with $R_o = \bar{Y}/\bar{W}$, $\bar{R}_o = \sum^{N} (y_i/Nw_i)$, $\tilde{W} = N/\sum^{N}(1/w_i)$.

The condition (6.15) implies that w and z^2/w are positively correlated. The other condition (6.16) for Θ is the same as (4.7) for 'a' . And $V(\hat{Y}_4)$ is minimized for given c when $\Theta = \Theta^*$. When $y_i = A + Bx_i$ for all i , Θ^* reduces to Θ_{opt} of the previous sections. Thus when the regression of y on x is linear, Θ^* can be interpreted as an approximation

to Θ_{opt} . Refer figure 6.1.

If the choice of Θ is felt to be difficult, one may use the stratification suggested in Sec. 6.2, and use $|x_i|$ as the measure of size provided no $x_i = 0$. The strata totals for y may be estimated separately and then added to give an estimate of Y.

CHAPTER 7.

SCOPE FOR FURTHER WORK

7.1 General Points

1. A product-type estimator has been suggested either as a dual or as an alternative to the commonly used ratio estimator in Chapters 2 and 3. The conditions for the optimality of the latter are known. The corresponding conditions for the product estimator based on harmonic means are given in Section 2.6. However, the optimality of a product estimator based on arithmetic means is to be investigated.

2. The performance of different estimators proposed in Chapter 2 to 6 may be studied by imposing on y a superpopulation dependent on x. Specific assumptions regarding the distribution of x may be made. For example, we may postulate that

$$y_i = A + Bx_i + e_i \qquad (7.1)$$

where $E(e_i|x_i) = 0$, and the x_i have independent and identical gamma distributions. For instance, following Srivenkataramana's (1978) work, Chaudhuri and Adhikary (1979) have examined the use of variate transformations in improving the efficiencies of some sampling strategies involving selections with varying probabilities. They have assumed regression models with and without specified forms of distributions for auxiliary variates.

3. The different estimators suggested in Chapters 2 to 6 may be generalized to incorporate auxiliary information on several variates. In

order to be practicable, these methods should be simple for application. The generalization suggested in Sec. 3.4 of forming a weighted combination of estimators, each using auxiliary information on a single x, may not be always effective.

4. The use of auxiliary information in estimating proportions and other features of interest may be investigated. Some work in this regard has been reported. For example Wynn (1976), Rao (1977) and Das (1979).

5. The Midzuno-Sen scheme was employed in Sections 4.5 to 4.8 to demonstrate a change of origin for the study variate for improving the efficiency of the Horvitz-Thompson estimator. Generalizations applying to any ppswor scheme may be attempted. One line of approach is the following. Consider probability proportional to x sampling wor with $p_i = x_i/X$, $i = 1,\ldots,N$, as initial probabilities where $X = \sum^{N} x_i$. Let the first and second order inclusion probabilities be denoted by π_i and π_{ij} $(i \neq j)$. Then π_i can be written as

$$\pi_i = p_i + p_i' \; ; \; i = 1,\ldots,N \; , \tag{7.2}$$

where p_i' is the probability of selecting i^{th} unit in the second or subsequent selections. The familiar Horvitz-Thompson estimator of Y is

$$\hat{Y}_{HT} = Z' E \; , \tag{7.3}$$

where $Z' = (y_1/\pi_1,\ldots,y_N/\pi_N)$, $E' = (d_1,\ldots,d_N)$ and $d_i = 1$ if unit i is included in the sample and zero otherwise. The variance of $\hat{Y}_{HT}$ is

$$V(\hat{Y}_{HT}) = Z'WZ \; , \tag{7.4}$$

where $W = (w_{ij})_{N\times N}$ with $w_{ii} = \pi_i(1-\pi_i)$; $w_{ij} = \pi_{ij} - \pi_i\pi_j$; $i \neq j = 1,\ldots,N$.

If y_i is proportional to π_i for all i then the estimator $\hat{Y}_{HT}$ reduces to a constant and hence $V(\hat{Y}_{HT}) = 0$. But from (7.2) we note that if p_i' is not proportional to p_i for at least one i , then y_i is not proportional to π_i even if $y_i = Bx_i$ for all i . In this context consider the transformation

$$y_i^* = y_i + bp_i' \; ; \; i = 1,\ldots,N \tag{7.5}$$

where b is a scalar to be chosen.

Then an unbiased estimator of Y is

$$\hat{Y}_{HT}^* = Z^{*'}E - b\sum_{i}^{N} p_i' = Z^{*'}\cdot E - b\,(n-1) \; , \tag{7.6}$$

since for any scheme $\sum_{i}^{N} p_i' = \sum_{i}^{N} (\pi_i - p_i) = n - 1$.

Here $Z^{*'} = (y_1^*/\pi_1,\ldots,y_N^*/\pi_N)$. And

$$V(\hat{Y}_{HT}^*) = Z^{*'}\, W\, Z^* = Z'WZ + 2bZ'WP + b^2P'WP \; , \tag{7.7}$$

where $P' = (p_1'/\pi_1,\ldots,p_N'/\pi_N)$. This variance is a minimum when

$$b_{opt} = -\,(Z'WP)/(P'WP) \; , \tag{7.8}$$

and

$$\min.V(\hat{Y}_{HT}^*) = Z'WZ - (Z'WP)^2/(P'WP) \; . \tag{7.9}$$

Comparing (7.4) with (7.7), the restriction on b in order that $\hat{Y}_{HT}^*$ is at least as precise as $\hat{Y}_{HT}$ is

$$0 \leq b \leq 2b_{opt} \; . \tag{7.10}$$

The choice of b may be made on the lines suggested in Sec. 4.7.

6. The auxiliary variates occasionally assume negative and positive values, as pointed out in Chapter 6. Two problems introduced by this were mentioned and two ways of overcoming these problems were outlined. Other methods of tackling these problems may be developed.

7. The auxiliary information may be incomplete in several respects. For example, with multiple auxiliary variates, it is quite likely that the values of these are not available for some of the population units. Some genuine techniques for utilizing the available information to the extent possible may be developed. Some efforts in this direction are in Han (1973) and Singh (1977).

8. The auxiliary information is generally used for improving the estimates of population mean, total, ratio etc. The possibility of using such information for obtaining better estimates of variance of these estimates may be investigated.

7.2 The Predictive Approach

Suppose s denotes a sample of n units from the finite population under consideration. Then any sampling procedure divides the population into a completely known (in respect of the character y being observed), and a completely unknown part. If we write

$$Y = \sum_{i \in s} y_i + \sum_{i \notin s} y_i = Y^{(1)} + Y^{(2)} \qquad (7.11)$$

then the first term $Y^{(1)}$ of (7.11) is known. Therefore, essentially a predictor of $Y^{(2)}$ is needed for estimating Y.

With the above premises Srivenkataramana and Tracy (1979) have suggested four methods of estimating the population total Y, given a simple random sample of n units without replacement from that population. Two of these methods are suited for the case of positive correlation between y and an auxiliary variate x, and the other two for the case of negative correlation.

Let $X^{(1)}, X^{(2)}$ denote the totals of the x-variate values for the sample and nonsample units respectively and

$$\hat{Y} = NY^{(1)}/n \ , \ \hat{X} = NX^{(1)}/n \tag{7.12}$$

be the (unbiased) simple expansion estimators of Y and X respectively, constructed from the sample. Then the suggested estimators are the following.

(i) Case of positive correlation

$$\hat{Y}_1 = Y^{(1)} + (\hat{Y} - Y^{(1)})X^{(2)}/gX = Y^{(1)}(1 + X^{(2)}/fX), \tag{7.13}$$

$$\hat{Y}_2 = Y^{(1)} + (\hat{Y} - Y^{(1)})X/\hat{X} = Y^{(1)}(1 + gX/X^{(1)}). \tag{7.14}$$

(ii) Case of negative correlation

$$\hat{Y}_3 = Y^{(1)} + (\hat{Y}-Y^{(1)})gX/X^{(2)} = Y^{(1)}(1 + g^2X/fX^{(2)}), \tag{7.15}$$

$$\hat{Y}_4 = Y^{(1)} + (\hat{Y}-Y^{(1)})\hat{X}/X = Y^{(1)}(1 + gX^{(1)}/f^2X). \tag{7.16}$$

Here $f = n/N$ is the sampling fraction and $g = 1-f$.

The expressions on the extreme right in (7.13) and (7.14) exhibit the near duality of the estimators $\hat{Y}_1$ and $\hat{Y}_2$. Here the unknown $Y^{(2)}$ has

been predicted to be $(\hat{Y} - Y^{(1)})$ modified by the factor $X^{(2)}/gX$ or $X/\hat{X}$. Similar is the case with $\hat{Y}_3$ and $\hat{Y}_4$. The guidelines for choosing among the estimators $\hat{Y}$, $\hat{Y}_1$, $\hat{Y}_2$ and the usual ratio estimator $\hat{Y}_r$ are in Table 7.1. And the guidelines in the case of negative correlation for choosing among $\hat{Y}$, $\hat{Y}_3$, $\hat{Y}_4$ and the product estimator $\hat{Y}_p$ are in Table 7.2.

Table 7.1: Guidelines in the positive correlation case

$f \leq \frac{1}{2}$		$f > \frac{1}{2}$ (or: $g = 1 - f < \frac{1}{2}$)	
Interval for $k = V_{11}/V_{02}$	Preferred estimator	Interval for k	Preferred estimator
$0 \leq k \leq f/2$	$\hat{Y}$	$0 \leq k \leq g/2$	$\hat{Y}$
$f/2 < k \leq \frac{1}{2}$	$\hat{Y}_1$	$g/2 < k \leq \frac{1}{2}$	$\hat{Y}_2$
$\frac{1}{2} < k \leq 1 - f/2$	$\hat{Y}_2$	$\frac{1}{2} < k \leq 1 - g/2$	$\hat{Y}_1$
$1 - f/2 < k$	$\hat{Y}_r$	$1 - g/2 < k$	$\hat{Y}_r$

Table 7.2: Guidelines in the negative correlation case

$f \leq \frac{1}{2}$		$f > \frac{1}{2}$ (or: $g < \frac{1}{2}$)	
Interval for $k' = -V_{11}/V_{02}$	Preferred estimator	Interval for k'	Preferred estimator
$0 \leq k' \leq f/2$	$\hat{Y}$	$0 \leq k' \leq g/2$	$\hat{Y}$
$f/2 < k' \leq \frac{1}{2}$	$\hat{Y}_3$	$g/2 < k' \leq \frac{1}{2}$	$\hat{Y}_4$
$\frac{1}{2} < k' \leq 1 - f/2$	$\hat{Y}_4$	$\frac{1}{2} < k' \leq 1 - g/2$	$\hat{Y}_3$
$1 - f/2 < k'$	$\hat{Y}_p$	$1 - g/2 \leq k'$	$\hat{Y}_p$

If we call the values of $k(k')$ in the four different intervals specified in columns 1 and 3 of Tables 7.1 and 7.2 to be very low, moderately low, moderately high and very high respectively, a quick summary of the results is as follows. When the sampling fraction is less than or equal to $1/2$ and the value of $k(k')$ is (i) very low it is preferable to use the simple estimator $\hat{Y}$, (ii) moderately low it is better to use $\hat{Y}_1(\hat{Y}_3)$, (iii) moderately high it is efficient to use $\hat{Y}_2(\hat{Y}_4)$ and (iv) very high it is good to use $\hat{Y}_r(\hat{Y}_p)$ as an estimator of the population total Y. In the contrary case of $f > 1/2$ the roles of $\hat{Y}_1(\hat{Y}_3)$ and $\hat{Y}_2(\hat{Y}_4)$ are interchanged.

7.2.1 Use of multiauxiliary information

One fairly general approach is the following. Assume that the auxiliary information is available on p variates $x_1,\ldots,x_p$. Let ρ_{ot} be the coefficient of correlation between y and x_t, C_t be the coefficient of variation of x_t in the population and $k_t = \rho_{ot}C_y/C_t : t = 1,\ldots,p$. Suppose we can make a good guess of the sign and magnitude of k_t for each t. Assume $f \leq 1/2$. We partition the set S of all the x-variates into seven subsets S_r, $t = 0,1,\ldots,6$ as follows: S_0 consists of all x_t for which

$$-f/2 \leq k_t' \leq f/2 ;$$

$S_1 : f/2 < k_t \leq 1/2$; $S_2 : 1/2 < k_t \leq 1 - f/2$;

$S_3 : 1 - f/2 < k_t$; $S_4 : -1/2 \leq k_t < -f/2$;

$S_5 : f/2 - 1 \leq k_t < -1/2$; $S_6 : k_t < f/2 - 1$.

Denote $S_1 \cup S_2 \cup S_4 \cup S_5$ by A. Thus A is the subset of all x-variates having moderately low or moderately high positive or negative values of k_t. In fact $A : f/2 < |k_t| \leq (1-f/2)$. Let X_t, $X_t^{(2)}$, $\hat{X}_t$ denote respectively the population total, total for the non-sample units and the simple expansion estimator of the population total for the variate x_t, $t = 1,\ldots,p$.

We may consider the following estimator of Y based on a without replacement simple random sample of size n from the population.

$$\hat{Y}^* = \hat{Y} \sum_{S_0} W_t + Y^{(1)} \sum_{S_1} W_t (1 + X_t^{(2)}/fX_t)$$

$$+ Y^{(1)} \sum_{S_2} W_t (1 + gX_t/X_t^{(1)}) + \hat{Y} \sum_{S_3} W_t X_t/\hat{X}_t$$

$$+ Y^{(1)} \sum_{S_4} W_t (1 + g^2 X_t/fX_t^{(2)})$$

$$+ Y^{(1)} \sum_{S_5} W_t (1 + gX_t^{(1)}/f^2 X_t) + \hat{Y} \sum_{S_6} W_t \hat{X}_t/X_t \qquad (7.17)$$

where $W = (W_1,\ldots,W_p)$, $\sum_1^p W_t = 1$, is a weight function and $\sum_{S_0}$ indicates summation over all x-variates belonging to S_0 and so on. For computation purposes $\hat{Y}^*$ can be recast as

$$\hat{Y}^* = \hat{Y}(\sum_{S_0} W_t + f \sum_A W_t + \sum_{S_1} W_t X_t^{(2)}/X_t + g \sum_{S_2} W_t X_t/\hat{X}_t$$

$$+ \sum_{S_3} W_t X_t/\hat{X}_t + g^2 \sum_{S_4} W_t X_t/X_t^{(2)} + g \sum_{S_5} W_t \hat{X}_t/X_t + \sum_{S_6} W_t \hat{X}_t/X_t) . \qquad (7.18)$$

The approximate bias and mse of the estimators and the best weight function have been given in Srivenkataramana and Tracy (1979).

7.2.2 Mixing estimators

Following the above work, Vos (1980) has considered mixing of direct, ratio and product method estimators. He introduces two classes of estimators:

$$\hat{Y}_{M1} = W\,\hat{Y} + (1-w)\,\hat{Y}_p\,, \tag{7.19}$$

$$\hat{Y}_{M2} = W\,\hat{Y} + (1-w)\,\hat{Y}_r\,. \tag{7.20}$$

It is shown that, for a suitable choice of the weights, each of these estimators has smaller variance than the direct as well as the product or ratio method estimators.

More generally one may consider pooled estimators of the type

$$\hat{Y}_W = w(t)\,T_1 + \{1 - w(t)\}T_2 \tag{7.21}$$

where T_1, T_2 are any two estimators of Y and $w(t)$ is a function of the statistic t used to test a relevant hypothesis; e.g. $H_o : B = B_o$ against the alternative $H_1 : B \neq B_o$ where B is the coefficient of regression of y on x in the population. Grimes and Sukhatme (1980) have examined the special case restricting the choice of $w(t)$ to functions of the type

$$w(t) = 1 \quad \text{if} \quad |t| \leq t_o$$
$$= 0 \quad \text{if} \quad |t| > t_o$$

and with difference and regression estimators as T_1, T_2 respectively.

They call the resulting estimator $\hat{Y}_W$ the sometimes regression estimator. The more general cases are to be investigated.

7.3 Orthogonal Auxiliary Variates

The case of two auxiliary variates x_1, x_2 and a difference method of estimation are assumed for illustration. Raj (1965a) has suggested an estimator of the form

$$\hat{Y}_d = \hat{Y} - w_1 k_1 (\hat{X}_1 - X_1) - w_2 k_2 (\hat{X}_2 - X_2) \tag{7.22}$$

which has, under srswor , variance given by

$$V(\hat{Y}_d) = \frac{N(N-n)}{n} \{S_{00} + w_1^2 k_1^2 S_{11} + w_2^2 k_2^2 S_{22} - 2w_1 k_1 S_{01} - 2w_2 k_2 S_{02} + 2w_1 w_2 k_1 k_2 S_{11}\} \tag{7.23}$$

where S_{00} is the population variance of y ; etc. It is clear that a positive S_{12} inflates the sampling variance (7.23). Therefore transformations which diffuse the correlation between the auxiliary variates will be helpful. Thus we may consider replacing the set (x_1, x_2) of auxiliary variates by variates which are orthogonal or nearly so. For example, replace (x_1, x_2) by $(x_1, x_2 - B_{21}x_1)$. The implication of this idea for the case of two or more auxiliary variates and practicable ways of effecting the transformation are to be studied.

PART TWO : PRACTICE

A SURVEY OF RURAL TRANSITION IN KARNATAKA, INDIA.

1980 - 81

CHAPTER - 8

BACKGROUND AND SURVEY DESIGN

8.1 Introduction

The rural life in India has been undergoing a visible transition to modern ways during the recent years. In the post-independence era (1947 -), particularly of late, a number of levelling forces have been operative on the Indian village scene. For example, villages are now uniformly administered by statutory *Panchāyats* (Village councils) elected through universal adult franchise. Schools, hospitals, post offices and cooperative societies are found at least in the vicinity of almost every village. Transport and communication have been greatly accelerated through roads, railways, post and telegraph, newspapers and radio. Illiteracy is being steadily reduced. Statute law has replaced customary law. Absentee landlordism is almost extinct. The land tenure of tenants has become secure. There is electricity supply to almost all villages in many of the states. A few case studies which examine some of these aspects have been made once in a while. However there is scope and need for conducting more socio-economic studies of the villages. The present survey was carried out in the Karnataka State in South India, with the objective of identifying the major factors causing the village transition and assessing the magnitude of change in the rural parts. Household amenities, farming practices, individual life-style and outlook, and transition of the village as a unit are studied. Since no comparable earlier study exists for Karnataka, it is hoped that this will serve as a benchmark.

8.2 The Land and the People

The Karnataka State is situated in the south-western part of India. The total area is 191,791 sq km. According to the 1981 census the population is 37 million, showing a 26.4% increase over the 1971 figure of 29.3 million. In terms of area and population Karnataka is the eighth among the states and union territories of India. An idea of the largeness of the population may be obtained by noting that it is greater than that of Canada (23.7 million).

The Karnataka State came into existence on November 1, 1956 under the States' Reorganization Act (1956). Initially it was known as Mysore State. The name was changed to Karnataka State on November 1, 1973 to signify the fact that the state comprises the Kannada land of olden times which was spread over different administrative units before the reorganization. Accordingly we note five regions in the state: (i) Old-Mysore Region; (ii) Kodagu Region; (iii) Madras-Karnataka Region; (iv) Bombay-Karnataka Region and (v) Hyderabad-Karnataka Region. Kannada is the official language of the state, but English continues to be used for official purposes and as medium of instruction in higher education. Several other Indian languages are spoken by segments of the population. For administrative purposes Karnataka has been divided into 19 districts. Table 8.1 gives some related details.

About 76% of the people in Karnataka live in rural areas spread over 29,553 villages of which 26,826 are inhabited. Thus Karnataka is a land of villages. The rural economy is mainly based on agriculture and allied activities like animal husbandry and fishing. About 55% of the villages

Table - 8.1: Districts, their area, population and its density in Karnataka†

	Region	District	Area (sq km)	1981 Population (million)	Density
I.	Old-Mysore	Bangalore	8005	4.9	615
		Chikkamagalur	7201	0.9	126
		Chitra Durga	10852	1.8	164
		Hassan	6814	1.4	198
		Kolar	8223	1.9	231
		Mandya	4961	1.4	285
		Mysore	11954	2.6	216
		Shimoga	10553	1.7	157
		Tumkur	10598	2.0	186
II.	Kodagu	Kodagu	4102	0.5	112
III.	Madras-Karnataka	Dakshina Kannada	8441	2.4	281
IV.	Bombay-Karnataka	Belgaum	13415	3.0	222
		Dharwad	13738	2.9	214
		Uttara Kannada	10291	1.1	104
V.	Hyderabad-Karnataka	Bellary††	9885	1.5	150
		Bidar	5448	1.0	182
		Bijapur	17069	2.4	141
		Gulbarga	16224	2.1	128
		Raichur	14017	1.8	127

† Source: Census of India, 1981, Series 9 Karnataka, Provisional population totals.

†† Though Bellary was in Madras State before 1953, it is included in region V because of its proximity to the other districts in this region and distance from the district in region III.

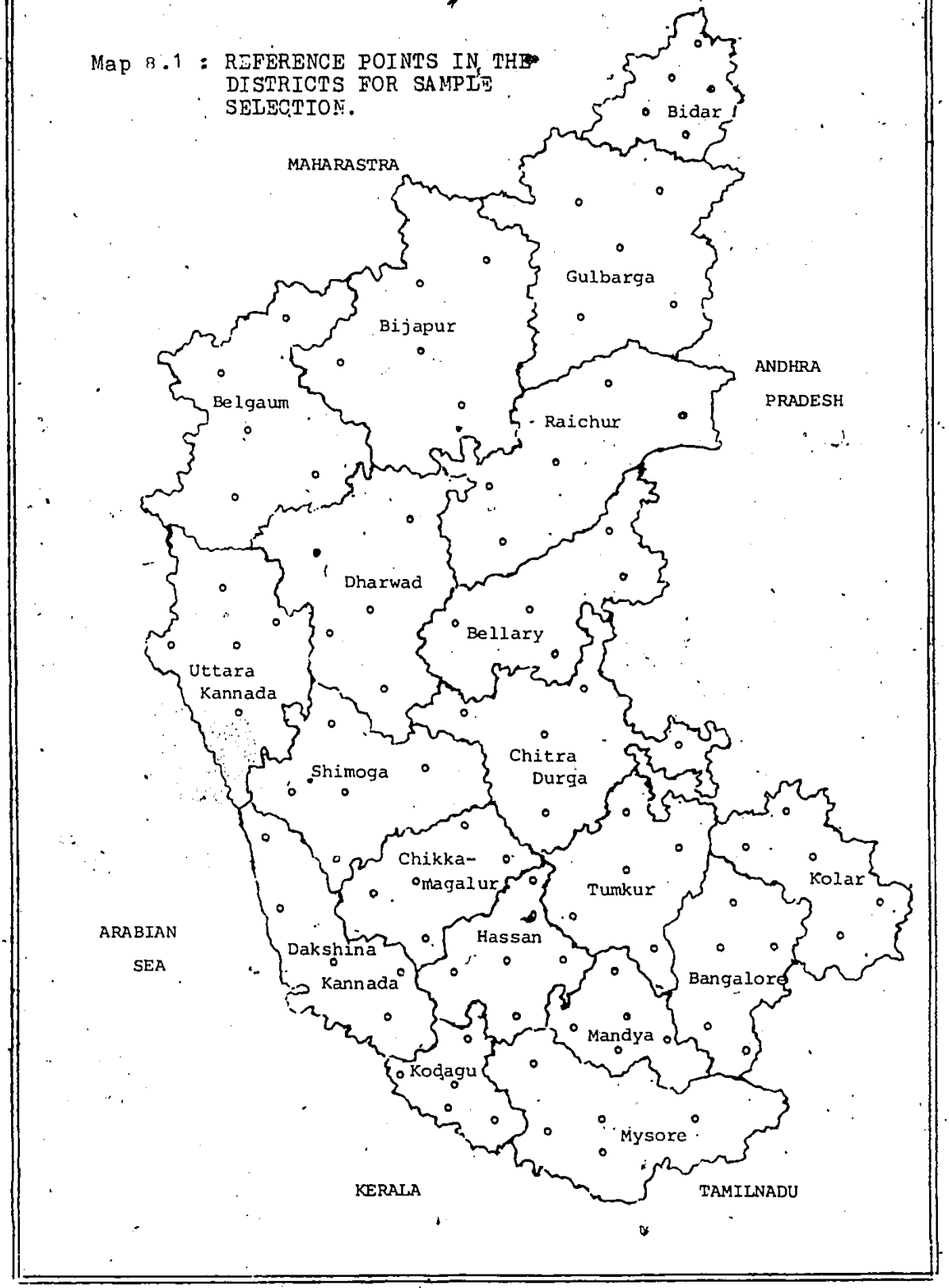
are small, each having less than 500 inhabitants, 39% are medium sized (having 500 to 2000 inhabitants) while the others are large.

8.3 Sample Selection

The units of interest in the survey are individuals, households, farms and villages. A restricted random selection ensuring an uniform physical coverage of the state was considered desirable, since differences exist between one region and another. Therefore, it was decided to include several rural points from each district, in the sample. Accordingly the following sample design was adopted.

In each district the larger towns were identified and grouped into five sets: one comprising all those at the centre of the district and the other four in the four directions from the centre - east, south, west and north. From each set one town was selected at random as the reference point. This ensured that the five points were well spread out in the district (see Map 8.1). A pair of investigators was assigned to each district. The pair was familiar with the district alloted to it and had a good knowledge of the local language. Around each reference point the investigators were to visit at least two villages separated by a minimum of 10 km. Houses of all types - huts, small houses and large houses were included. In order to assess individual transition, two persons were selected at random from each selected household, excluding children aged below 10 years. In all, information from 184 villages, 431 households, 379 farms, and 823 individuals was recorded. Initially, the more objective method of listing the villages near a reference point and selecting two from them and then listing the households in the selected villages for picking the sample households was tried. But this was seen to be too time consuming since ready lists were rarely available and

Map 8.1 : REFERENCE POINTS IN THE DISTRICTS FOR SAMPLE SELECTION.



was abandoned in favour of the above, though less objective, alternative.

The data were collected during December 1980 and January 1981. During these months of the year the field conditions are generally favourable in the sense of being free from floods or other extreme climatic conditions. The investigators filled up a specially designed questionnaire (Refer Appendix), which was pretested.

8.4 The Questionnaire

This had six sections - basic information, house, farm, outlook, individual and village transitions. The majority of the questions was of multiple choice type. The questionnaire was in English. A translation was provided in the provincial language, Kannada, to facilitate understanding by a larger section of the people.

The information on household and farm was obtained by setting up a dialogue with that member who was able to respond most clearly. He (she) was allowed to be assisted by the others, since the information was on the household or farm and not on the individual. Direct observation by the interviewer was used for factors like housetypes and condition of surroundings. A household was defined as a group of people living together and eating from the same kitchen.

Information on outlook and individual transitions was obtained from the individuals selected. Information on village transition was obtained by talking to person(s) well informed about the village, like the village *Panchāyat* President or Accountant and also through direct observation.

CHAPTER - 9

FINDINGS OF THE SURVEY

9.1 Quality of Data

The investigators in the survey were the staff and students of the Department of Statistics of Bangalore University. To begin with, they were familiarized with the objectives of the project. A small scale pilot study was also conducted. In the main survey the investigators actually visited the rural points in the sample and obtained quite a bit of information through direct observation. Thus the nonsampling errors are expected to be minimal.

The Indian villagers generally mistake the investigators to be revenue officials of the government, collecting information for tax purposes and hence hesitate to talk to them. In order to overcome this problem the field workers carried with them an identification card and a statement that the project was for an academic purpose. On reaching the village they contacted the *Panchāyat* President or an important local person and acquainted him of their objective. He usually sent word around the village, instructing the villagers to cooperate. The investigators knew the language spoken by the local populace. All this helped to win the confidence of the respondents. This reduced the nonresponse and improved the quality of the data. When the completed questionnaires arrived in Bangalore, a specially trained staff scrutinized them according to instructions prepared beforehand and coded the relevant entries. The data were then transferred to punched cards and tabulated.

9.2 Results of the Survey

The data from the survey are used to examine: (i) housetypes; (ii) household conveniences; (iii) household change; (iv) outlook transition; (v) individual transition; (vi) farm transition; and (vii) village transition in the rural parts of Karnataka. Since the responses were mainly categorical, the results are mostly expressed as percentages. In each case the sample size and the method used to obtain the data are stated.

9.2.1 Housetypes

The statements are based on the data obtained from the 431 houses in the sample, through direct observation by the investigators. There are three main types of dwelling units that one sees in the villages of Karnataka - independent houses (58%), huts (26%) and part of a building (16%). Generally the walls are constructed with clay puddle (56%) or locally made brick and stone (40%). Sometimes we come across the use of bamboo poles or wooden planks (4%) for this purpose. The use of flat or pan tiles for the roof is most common (54%). Straw, grass, Palmyrah or coconut palm is used instead in the huts. The use of reed and mud (16%), asbestos or galvanized steel sheets (3%) is also occasionally seen. The flooring is made of mud plaster (60%) or brick and stone. The percentage of houses having own wells as source of water is about 25, while the majority (71%) depend on community wells. About 60% of the houses have their own *angāla* (open-air courtyard).

About 58% of the houses have 3 or fewer rooms for an average of 8 members. Therefore per force the same room is to be used for a number of different purposes. Ventilation is often poor (36%) in the case of the older structures.

9.2.2 Household conveniences

The conclusions here are partly based on direct observation by the interviewer and partly on the information obtained by setting up a dialogue with a member of the household who was able to respond most clearly. All the 431 households in the sample are covered. It is noted that a large number (62%) of the households have a separate space as kitchen. Aluminum (45%) and earthen (41%) cooking vessels are the most common. In the upper class families copper and brass vessels were common earlier, now making room for steelware. In the poorer sections, the change is from earthenware to aluminum vessels. Kerosene lamps continue to be the main source of light (95%), but there is a progressive switch-over to electric lamps. This is aided by the fact that almost all the villages in the state are supplied with electricity. More than 50% of the houses have some furniture.

One of the most significant changes is the trend with respect to medical care. People have changed over from native medicines to government hospitals (69%) or private medical practitioners (27%). Improved roads and bus transport to urban centres have provided access to the hospitals. However medical aid is generally sought only in serious cases. The situation of veterinary care is similar. The habit of seeing movies for recreation is noted in the villages near urban centers. Tours are undertaken generally as *yatré* (pilgrimage) rather than for sightseeing.

9.2.3 Household change

The respondents were asked to what extent their family had changed in respect to education, dress pattern, household amenities, medical aid,

traditions and customs and the status of women during the past decade. The percentage distribution of the responses from the 431 families is shown in Table 9.1.

The percentage of families reporting moderate or significant change with respect to the different factors is quite high. An exception is the case of traditions and customs where half the number of households reported no change. The trends in respect to education and medical aid are particularly noteworthy.

Table 9.1 Percentage distribution according to extent of change in the families (n = 431).

Factor	Extent of change		
	Nil	Moderate	Significant
Education	24	47	29
Dress pattern	28	55	17
Household amenities	40	47	13
Medical aid	23	60	17
Traditions & customs	50	36	14
Status of women	39	47	14

9.2.4 Outlook transition

A typical villager has religious and caste beliefs and has at least a fair awareness of the village-level political situation (77%), though less so at the provincial (53%) and national (44%) levels. The outlook on certain other factors is summarized in Table 9.2.

A high percentage of people in favour of women's education, equal rights for women, family planning and a high percentage against the dowry system in marriage is noted.

Table 9.2 Percentage distribution by outlook on certain factors in Karnataka (n = 431).

Factor	Outlook		
	For	Against	Indifferent
Religious & caste beliefs	58	29	13
Untouchability	25	58	17
Dowry system	15	67	18
Women's education	77	14	9
Equal rights for women	71	17	12
Family planning	70	16	14
Superstitions	32	39	29

9.2.5 Individual transition

Two individuals were selected at random from each selected household for studying this aspect. Children below 10 were excluded. There were 470 male and 353 female respondents. Their distribution by education level is in Table 9.3.

A typical individual in the village is traditionally dressed (56% for males, 69% for females) or shows a change towards modern type (36% for males, 24% for females) of dress. Men traditionally wear dhoties, while women wear saris. The practices as regards the use of certain items in daily life are summarized by the percentages in Table 9.4.

Table 9.3 Distribution by education level and sex of respondent

Education level	Percentage	
	Male	Female
Nil	28	62
Up to grade 5	19	13
Grades 6 through 10	37	22
College	16	3

Table 9.4 Use of certain items in daily life by individuals

Item	Male (%)		Female (%)	
	No	Yes	No	Yes
Footwear	42	58	61	39
Cosmetics	78	22	66	34
Toilet soap	35	65	41	59
Toothpaste/powder	76	24	62	38
Hair oil	57	43	58	42
Wrist watch	54	46	85	15
Safety razor	53	47	N.A.	
Vanity bag	N.A.		82	18

It is interesting to note that 42% of the men and 61% of the women do not use any footwear. Also only 65% of the men and 59% of the women use toilet soap. The 66% of women using no cosmetics or the 82% using no vanity bag are indicative of a traditional type of rural women folk.

9.2.6 Farm transition

A typical (57%) farm is a small landholding (less than 5 acres), growing food crops only (53%), like paddy and wheat, and an additional 41% growing both food and cash crops (like cotton or sugarcane). The mode of ploughing land is mainly by animal-drawn wooden or metal plough (96%). The use of tractors is just 4%. The farmer depends heavily (66%) on rain for irrigation; 15% have water pumps and another 15% have water supply through canal systems.

The practice of using certain important factors in agriculture is indicated by the percentages in Table 9.5.

Table 9.5 Percentages showing the use of certain factors in farming in Karnataka (n = 379).

Factor	Extent of use		
	Never	Occasional	Frequent
Compost	26	18	56
Chemical fertilizers	24	38	38
Chemical insecticides	32	38	30
Hybrid & high yielding varieties	45	31	24
Banking facility	46	34	20
Farm cooperative	45	36	19

These percentages are almost equally distributed among the three columns. Exceptions are the case of compost, where usage is frequent, and, to some extent, the case of banking facility, where the situation is not very satisfactory. The mode of storing farm produce is quite poor (39%) and is just satisfactory in another 57% of the cases. Again,

the mode of marketing the produce is unsatisfactory in about half the number of cases. Only a moderate change in farming practices during the last 10 years was reported in 57% of the cases, with another 23% reporting no change at all.

9.2.7 Village transition

A village in Karnataka has, on the average, about 200 households with a population of about 750. The average distances of certain facilities are given in Table 9.6. By and large these distances are moderate, considering the importance of the facility for a village, except for railway stations and banks.

Table 9.6 Average distance of facilities from a village in Karnataka (n = 184)

Facility	Average distance (km)
Railway station	33.4
Bus station	2.8
Tar road	1.9
Post office	1.4
Elementary school	0.5
High school	4.1
College	12.4
Government hospital	5.3
Doctor's shop	4.0
Fair price shop	2.7
Bank	5.9
Cinema	7.9

The condition of primary and secondary schools is found to be good or tolerable in 83% of the villages (Table 9.7). The government-sponsored adult education programme does not seem to work satisfactorily, with more than 60% of the villages having very little activity of this type. The primitive practices of barter, bonded labour and untouchability are on their way out (Table 9.8).

Table 9.7 Condition of certain facilities for villages in Karnataka (Percentage of villages, n = 184).

Facility	Condition		
	Good	Tolerable	Bad
School	33	50	17
College	13	50	37
Adult education	10	29	61
Medical aid	16	51	33
Bus transport	40	33	27
Electricity	46	28	26

Table 9.8 Prevalence of certain practices in villages in Karnataka (Percentage of villages, n = 184).

Practice	Prevalence		
	Nil	Not much	High
Barter system	48	38	14
Bonded labour	60	29	11
Untouchability	41	40	19

About 59% of the villages exhibit an overall moderate change in village life-style during the last decade, while 33% showed a substantial change (Table 9.9).

Table 9.9 Extent of change in village life during 1971-81 in Karnataka (n = 184)

Extent	Percentage
Nil	8
Moderate	59
Great	33

9.3 Factors Causing Transition

An attempt was made to order the following factors according to their influences on village life-style:

(a) Schooling

(b) Roads and means of communication

(c) Rural electricity

(d) Government programmes like community development

(e) Newspapers and radio.

The village was considered as a unit and persons well informed about the village were asked to state the impact of the above factors on rural transition in general. Their responses are summarized by the percentages in Table 9.10.

Roads and means of communication have contributed the most towards village transition, schooling comes next and then we have newspapers and radio, government programmes and rural electricity. The observed influence of these is briefly outlined next.

Table 9.10 Impact of certain factors on rural transition (Percentage of responses, n = 184).

Factor	Impact		
	Nil	Moderate	Significant
Schooling	13	54	33
Roads & means of communication	18	46	36
Rural electricity	32	43	25
Government programmes	30	43	27
Newspapers & radio	18	52	30

Roads passing through or nearby a village have a visible impact on the life around. This can be easily seen by comparing the life-style in a village adjacent to a road with that in a village about 10 km away from the nearest road.

In the former, a bunch of shops on or near the road where people engage themselves in petty trades is common. Bullock carts and bicycles ply around transporting people and goods. This wider accessibility sets up a network and makes well-mixed economic activity possible.

Children make use of the road for reaching the school, walking along it or taking the bus that may ply a couple of times during the day. This helps them beat the rivulets putting up hurdles during the rainy months. Farmers and worker groups also find similar utility from the roads. This has greatly improved the income from farm and dairy produce and also wage rates.

Easy access to the cities allows the village folk to observe frequently and directly the life-style in urban areas. This has resulted in such

style being copied in terms of dress, food habits and practices. Visits to cinema and coffee houses change the attitudes of rustic people.

With an activity spot like the road-side bus stand or intersection of roads a reversal in the mode of availability of certain services and facilities is seen. The village barber, for example, used to go from door to door to make his service available. These days he prefers to set up a petty shop and wait for the customers to walk in. Similar observations can be made regarding the seller of bangles, earthen pots or fish.

On the other hand, in a village far removed from roads, the activities are geographically very much restricted. The hills, hillocks, rivers and rivulets delimit the area of activity of a person, setting up natural barriers to mobility. This is especially true of rivers during the rainy months.

The ability to read and write greatly increases the sources of information for a person, and improves his capacity to examine a situation critically. The quality of his life improves and he develops his own values in life. The general picture in Karnataka is one of pronounced increase in literacy rate over the years. It was 25.4% in 1961, 31.5% in 1971 and now it is 38.4%. This trend has been maintained by both males and females. However there is enormous scope for improvement in female literacy. Among men, literacy improved from 41.6% in 1971 to 48.6% in 1981, while the corresponding percentages for women were only 21.0 and 27.8. The increase in literacy is the result of the expansion of primary and secondary education in the state, free education up to secondary level and the cumulative effect of investments made during the

past decades.

Newspapers and radio supply the people with current information constantly. Special programs and features meant for the villagers are included. These media assist in the quick and continuous spread of relevant information as a routine matter.

Special government programs, like community development, food for work and adult education, have helped to fill the gaps in rural development. They have also helped people to realize the benefit of cooperatives and other collective efforts.

Almost all villages in Karnataka are supplied with electricity. This has improved irrigation and has replaced some manual work by machine work. Electricity is also being used progressively for lighting.

9.4 Some Limitations

The present study was a multipurpose one and covered a large geographical area. Because of an inadequate network of roads and poor transport facilities, certain rural pockets of land were inaccessible. Further the standard sampling techniques and methods of data collection used in the developed countries cannot be applied in the rural Indian situation. Here the lists on which sampling frames can be based are inadequate. The majority of the people are illiterate and have a suspicion of outsiders, the women folk particularly shying away from strangers. The present study was planned with an awareness of these handicaps. Accordingly the survey design was adapted to suit the field conditions. Also since care was taken to deputize well trained interviewers, knowing the local language and other peculiarities, the quality of the data was reasonable. The results obtained are in general agreement with other comparable results,

like that from the census. The study sheds light on the ways of life in a society about which investigations of this type have been rare. It also serves as a benchmark and has implications on the larger rural Indian context.

9.5 Follow-up Studies

A study of transition essentially tries to measure the extent of change that has taken place during a period of time. Therefore it is necessary to have comparable studies at the two points of time concerned. In the present case, no comparable earlier study exists. Thus it may be interesting to use the present study as a benchmark and conduct a similar study after some time, say five years, and assess the transition that occurs meanwhile. On a smaller scale, such studies may be conducted in greater depth by considering individual districts. Case studies of a few villages can also be taken up.

The Karnataka State was under the British before 1947. The neighbouring union territory of Goa on the west coast was under the Portuguese while, on the east coast, Pondicherry was under the French. A comparison of rural transition in Karnataka with that in Goa and Pondicherry will be worthwhile.

QUESTIONNAIRE

BANGALORE UNIVERSITY

BANGALORE

DEPARTMENT OF STATISTICS

QUESTIONNAIRE

A SURVEY ON RURAL TRANSITION

IN KARNATAKA, INDIA

1980-81

(WITH ASSISTANCE FROM I D R C, CANADA)

A Survey on Transformation of Rural People to Modern ways of Life in Karnataka, India.

Serial No .....

Stratum No.....

Interviewers i) .....

ii) .....

Date.....

I. BASIC INFORMATION

1. Village .....
2. Taluk .....
3. District .....
4. Household size .....
5. Joint family ? Yes/No

6. Particulars of the members :

Name	Sex M/F	Age	Relation to the Head	Level of education	Currently studying Yes/No
i)					
ii)					
iii)					
iv)					
v)					
vi)					
vii)					
viii)					
ix)					
x)					
xi)					
xii)					
xiii)					
xiv)					
xv)					

7. Average monthly household income Rs.
8. Average monthly household expenditure Rs.
9. Two main sources of income i) ii)
10. Number of couples in the household

3

II. HOUSE TRANSITION

1. House — i) Own — ii) not own

2. Plinth area (Sq. Ft) .....

3. House type — i) hut — ii) part of a building — iii) Independent house — iv) flat

4. Roof type — i) straw, grass, palmyrah or coconut palm — ii) reed and mud — iii) flat or pan tiles — iv) asbestos sheets — v) galvanized steel sheets — vi) R. C. C.

5. Wall type — i) bamboo poles etc. — ii) clay puddle — iii) wooden plank — iv) brick or stone

6. Floor type — i) mud — ii) brick, stone — iii) cement

7. Condition of structure — i) bad — ii) moderately good — iii) good

8. Period since built (years) _____

9. Period since last major repair undertaken (years) _____

10. Condition of surroundings — i) bad — ii) moderately good — iii) good

11. Number of living rooms _____

12. Number of other rooms _____

13. Ventilation — i) bad — ii) tolerable — iii) good

14. Kitchen — i) not separate — ii) separate

15. Bathing place — i) no specific arrangement — ii) sharing with others — iii) own — iv) near water pond, river etc.

16. Type of lavatory i) no arrangement ii) services
iii) septic tank iv) drainage

17. Source of water for house i) community well (or bore well) etc.
ii) own well iii) water supply

18. Main cooking vessels i) earthen ii) alluminium iii) brass, copper or bronze
iv) stainless steel

19. Main cooking fuel i) fire wood, dung cake
ii) charcol, saw dust iii) kerosene
iv) gas v) electricity

20. Main lighting fuel i) kerosene ii) electricity iii) others

21. Household furniture (numbe
i) chairs ..... ii) cots .....
iii) benches ..... iv) tables .....
v) almairahs ..... vi) stools .....

22. Vehicles i) cart yes/no ii) bicycle yes/no
iii) scooter (moped) yes/no iv) car (van) yes/no

23. Other articles i) clock (watch) yes/no ii) radio yes/no
iii) pen (pencil) yes/no

24. Main transport mode within village limits
i) walking ii) cart
iii) bicycle iv) public transport
v) rickshaw vi) scooter (moped) vii) car

25. Medical care i) native medicine ii) govt. hospital
iii) private doctor

26. Distance to the nearest medical centre in kms.....

27. Do you take medical help only in serious cases ? yes/no

28. Veterinary care i) native medicine ii) govt, hospital
iii) private doctor

29. Distance to the nearest veterinary centre in kms .....

30. Do you take veterinary help only in serious cases ?

Subscription to periodicals

31.	dailies	a) English	Yes/No
		b) Others	Yes/No
32	magazines	a) English	Yes/No
		b) Others	Yes/no

33. Does the family regularly listen to radio news ? Yes/No

34. Does the family go to movies regularly ? Yes/No

35. Does the family go on picnics or tours frequently ? Yes/No

36. Does the family like a city type of living ? Yes/No

To what extent did your household change with respect to the following, during the last years ?

	Item	Extent		
37.	Education	i) nil	ii) Moderate	iii) significant
38.	Dress pattern	i)	ii)	iii)
39.	Household amenities	i)	ii)	iii)
40.	Medical aid	i)	ii)	iii)
41.	Traditions and customs	i)	ii)	iii)
42.	Status of women	i)	ii)	iii)

6

III. FARM TRANSITION

1. Livestock (number) i) cows and oxen..... ..... ii) buffaloes.....
iii) sheep and goat..... iv) poultry .....

2. Area under cultivation i) Nil ii) ..... acres

3. Land i) own ii) not own

4. Two main items grown i) .... ii) .....

5. Mode of ploughing land i) animal drawn wooden plough
ii) animal drawn metal plough
iii) tractor

6. Main mode of irrigation i) rain and natural water
ii) lift system iii) tube well
iv) water pump v) canal

Use of the following

	Compost	i)	never	ii)	occasional	iii)	frequently	If (i) specify reasons .....
8.	Chemical fertilizers	i)		ii)		iii)		.....
9.	Chemical insecticides	i)		ii)		iii)		.....
10.	Hybrid and high yielding varieties	i)		ii)		iii)		.....
11.	Banking facility	i)		ii)		iii)		.....
12.	Farm co-operative	i)		ii)		iii)		.....

13. Mode of storing the produce

i) unsatisfactory ii) satisfactory iii) good

14. Mode of marketing the produce

i) unsatisfactory ii) satisfactory iii) good

15. To what extent your farming practices have changed during the last ten years ?

i) nil ii) moderate iii) significant

(Sections IV and V for selected individuals)

IV. OUTLOOK TRANSITION

Political awareness

1.	Local	i) good	ii) fair	iii) poor
2.	Provincial	i)	ii)	iii)
3.	National	i)	ii)	iii)
4.	Religious and caste beliefs	i) for	ii) against	iii) indifferent
5.	Untouchability	i)	ii)	iii)
6	Dowry system	i)	ii)	iii)
7.	Women's education	i)	ii)	iii)
8	Equal rights for women	i)	ii)	iii)
9.	Family planning	i)	ii)	iii)
10.	Superstitions	i) strong	ii) moderate	iii) nil

8

V. INDIVIDUAL TRANSITION

Item	Respondent A	Respondent B
1. Name	.....	.....
2. Sex	M/F	M/F
3. Age (years)	.....	.....
4. Level of education	.....	.....
5. Type of dress		
i) traditional		
ii) modern		
iii) urbanized		
iv) Ultramodern	i) ii) iii) iv)	i) ii) iii) iv)
Use of the following		
6 Safety razor (for men only)	yes/no	yes/no
7. Foot wear		
i) nil ii) chappals		
iii) shoes	i) ii) iii)	i) ii) iii)
8. Cosmetics	yes/no	yes/no
9. Toilet soap	yes/no	yes/no
10. Tooth paste/powder	yes/no	yes/no
11. Hair oil	yes/no	yes/no
12. Wrist watch	yes/no	yes/no
13. Vanity bag (for women only)	yes/no	yes/no

9

VI. VILLAGE TRANSITION

1. District.....
2. Taluk .....
3. Village.....
4. No. of Household in the village.....
5. Population of the village.....

Distance from the nearest following places*

Place	Distance (k.ms.)	Place	Distance (k.ms.)
6. Railway station	.....	16 Veterinary hospital	.....
7. Bus station	.....	17. Fertilizer depot.	.....
8. Tar road	.....	18. Repair place for pump sets etc,	.....
9. Post office	.....	19 Flour mill	.....
10. Elementary school	.....	20. Fair price shop (ration shop)	.....
11. High school	.....	21. Shopping centre	.....
12. College	.....	22. Bank	.....
13. Govt. hospital	.....	23. Cinema	.....
14. Doctor's shop	.....	24. Library & reading room	.....
15. Primary health centre	.....		

The condition of the following facilities for the village

25. School	i) good	ii) tolerable	iii) bad
26. College	i)	ii)	iii)
27. Adult education	i)	ii)	iii)

REFERENCES

Brewer, K.W.R. (1963). Ratio estimation in finite populations: Some results deducible from the assumption of an underlying stochastic process. Aust. J. Statist., 5, 93-105.

Chaudhuri, A. and Adhikary, A.K. (1979). Improving on the efficiencies of standard sampling strategies through variate transformations in estimators for finite population means. Preprint, Indian Statistical Institute, Calcutta.

Cochran, W.G. (1942). Sampling theory when the sampling units are of unequal sizes. J. Amer. Statist. Ass., 37, 199-212.

Cochran, W.G. (1977). Sampling Techniques, 3rd ed., New York: John Wiley and Sons.

Das, A.K. (1979). On the use of auxiliary information in estimating proportions. Tech. Report 10, Indian Statistical Institute, Calcutta.

Goodman, L.A. (1960). On the exact variance of products. J. Amer. Statist. Ass., 55, 708-13.

Grimes, J.E. and Sukhatme, B.V. (1980). A regression-type estimator based on preliminary test of significance. J. Amer. Statist. Ass., 75, 957-962.

Han, C.P. (1973). Double sampling with partial information on auxiliary variables. J. Amer. Statist. Ass., 68, 914-18.

Hansen, M.H. and Hurwitz, W.N. (1943). On the theory of sampling from finite populations. Ann. Math. Statist., 14, 333-62.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. J. Amer. Statist. Ass., 47, 663-85.

Kish, L. (1965). Survey sampling. New York: Wiley.

Midzuno, H. (1952). On the sampling system with probability proportionate to the sums of sizes. Ann. Inst. Statist. Math., 3, 99-107.

Murthy, M.N. (1964). Production method of estimation. Sankhyā, A26, 69-74.

Murthy, M.N. (1967). Sampling theory and methods. Calcutta: Statistical Publishing Society.

Murthy, M.N. and Nanjamma, N.S. (1959). Almost unbiased ratio estimates based on interpenetrating subsample estimates. Sankhya, 21, 381-92.

Narasimha Prasad, N.G. and Srivenkataramana, T. (1980). A modification of the Horvitz-Thompson estimator under the Midzuno sampling scheme. Biometrika, 67, 709-11.

Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. J.R. Statist. Soc., 97, 558-625.

Olkin, I. (1958). Multivariate ratio estimation for finite populations. Biometrika, 45, 154-65.

Quenouille, M.H. (1956). Notes on bias in estimation. Biometrika, 43, 353-60.

Raj, D. (1965). On sampling over two occasions with probability proportionate to size. Ann. Math. Statist., 36, 327-330.

Raj, D. (1965a). On a method of using multi-auxiliary information in sample surverys. J. Amer. Statist. Ass., 60, 270-77.

Raj, D. (1968). Sampling theory. New York: McGraw-Hill.

Rao, J.N.K. (1969). Ratio and regression estimators. In New Developments in survey sampling. Eds. N.L. Johnson and H. Smith, pp. 213-34. New York: Wiley.

Rao, J.N.K., Hartley, H.O. and Cochran, W.G. (1962). On a simple procedure of unequal probability sampling procedure without replacement. J. Roy. Statist. Soc. B, 24, 482-91.

Rao, P.S.R.S. and Mudholkar, G.S. (1967). Generalized multivariate estimator for the mean of finite populations. J. Amer. Statist. Ass., 62, 1009-12.

Rao, T.J. (1977). Estimation of change in proportions. Bull. Inter. Stat. Inst., 47(4), 418-21.

Reddy, V.N. (1978). A study on the use of prior knowledge on certain population parameters in estimation, Sankhya, C40 Pt. 1, 29-37.

Reddy, V.N. and Rao, T.J. (1977). Modified pps method of estimation. Sankhya, C39, 185-97.

Robson, D.S. (1957). Applications of multivariate polykays to the theory of unbiased ratio-type estimation, J. Amer. Statist. Ass., 52, 511-22.

Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. Biometrika, 57, 377-87.

Royall, R.M. and Herson, J. (1973). Robust estimation in finite populations, I. J. Amer. Statist. Ass., 68, 880-9.

Sampford, M.R. (1978). Prediction estimation and internal congruency. In Contributions to survey sampling and applied statistics. Ed. H.A. David, pp. 29-39, New York: Academic Press.

Singh, R. (1977). A note on the use of incomplete multiauxiliary information in sample surveys. Austral. J. Statist., 19, 105-107.

Srivastava, V.K., Shukla, N.D.and Bhatnagar, S. (1981). Unbiased Product Estimators. Metrika, 28, 191-96.

Srivenkataramana, T. (1978). Change of origin and scale in ratio and difference methods of estimation in sampling. Can. J. Stat.. 6, 79-86.

Srivenkataramana, T.. (1980). A dual to ratio estimator in sample surveys. Biometrika, 67, 199-204.

Srivenkataramana, T. and Srinath, K.P. (1982). Change of scale in difference estimation in sampling. J. Ind. Soc. Agri. Stat., 34(1).

Srivenkataramana, T. and Tracy, D.S. (1979). On ratio and product methods of estimation in sampling. Statist. Neerl. 33, 37-49.

Srivenkataramana, T.and Tracy, D.S. (1980a). An alternative to ratio method in sample surveys. Ann. Inst. Statist. Math., 32, 111-120.

Srivenkataramana, T. and Tracy, D.S. (1980b). A change of origin after pps sampling. Metron, 37(1-2), 175-181.

Srivenkataramana, T. and Tracy, D.S. (1980c). Extending product method of estimation to positive correlation case in surveys. Aust. J. Statist. 23, 95-100.

Sukhatme, P.V. and Sukhatme, B.V. (1970). Sampling theory of surveys with applications, Iowa: Iowa State University Press.

Vos, J.W.E. (1980). Mixing of direct, ratio, and product method estimators. Statist. Neerl., 34, 209-18.

Wynn, H.P. (1976). An unbiased estimator under SRS of a small change in a proportion. The Statistician, 25(3), 225-28.

Yates, F. (1960). Sampling methods for censuses and surveys, 3rd ed., London: Griffin.

Yates, F. and Grundy, P.M. (1953). Selection with replacement from within strata with probability proportional to size. J.R. Statist. Soc., B15, 253-62.

VITA AUCTORIS

1945	Born on the 18th of May at Mangalore, India.
1959	Matriculated from Anandashrama High School, Dakshina Kannada, India.
1963	Graduated with a B.Sc. degree from Government College, Mangalore, affiliated to University of Mysore, India.
1965	Received M.Sc. degree in Statistics from Karnataka University, Dharwar, India. Appointed Lecturer in Statistics, Department of Collegiate Education, Karnataka, India.
1969	Appointed Assistant Professor in Statistics, Bangalore University, India.
1975	Won Bangalore University Research Award in Statistics.
1976	Admitted to Graduate Programme at the University of Windsor, Canada.
1978	Won the Pierre Robillard Award of the Statistical Society of Canada.
1979	Received M.Sc. degree in Mathematics from the University of Windsor, Canada. Advanced to Ph.D. candidacy. Won a Ph.D. Thesis Research Award of the International Development Research Centre of Canada.