BAYES AND LEAST-SQUARES PROCEDURES IN SAMPLING FROM FINITE POPULATIONS

by

PRETTA LATHA MEHROTRA

Thesis submitted in fulfilment of the requirements for the degree of Master of Science

April 1984

To

Paps & Mi

ACKNOWLEDGEMENTS

The author wishes to express her sincere appreciation and thanks for the valuable guidance and supervision received from Dr. RM. Sekkappan during the course of her research. She also thanks Dr. Ng Vee Ming, her co-supervisor for his efforts and Puan Alauviyah Haji Ismail, for typing out this thesis.

TABLE OF CONTENTS

List of Notations

$U = \{1, 2, \ldots, N\}$	finite population
n, N	sample and population size respectively
$\underset{\sim}{Y} = (Y_1, \ldots, Y_N)$	response variable vector
$\bar{Y} = \frac{1}{N} \sum_{k=1}^{N} Y_k$	population mean of response variable
$\bar{X}_{j\cdot} = \frac{1}{N} \sum_{k=1}^{N} X_{jk}$, $j = 1,2,\ldots,p$	population mean of j-th auxiliary variable
$\bar{X}_{ij\cdot} = \frac{1}{N_i} \sum_{k=1}^{N_i} X_{ijk}$, $i = 1,2,\ldots,H$	i-th stratum population mean of j-th auxiliary variable
$S_y^2 = \sum_{k=1}^{N} (Y_k - \bar{Y})^2$	population sum of squares of Y
$S_{x_j}^2 = \sum_{k=1}^{N} (X_{jk} - \bar{X}_{j\cdot})^2$, $j = 1,2,\ldots,p$	population sum of squares of $X_{j\cdot}$
$S_{x_j y} = \sum_{k=1}^{N} (X_{jk} - \bar{X}_{j\cdot})(Y_k - \bar{Y})$, $j = 1,2,\ldots,p$	population sum of the product of Y and $X_{j\cdot}$
$R_{xy} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{S_{xy}}{S_x S_y}$	population correlation coefficient
$s = (1,\ldots,n)$	sequence of distinct units in sample, s
$s = (y_1,\ldots,y_n)$ $= (y : y \in s)$	sequence of sampled y values
S	the set of all possible subsets of U

$s_y^2 = \sum_{k=1}^{n} (y_k - \bar{y})^2$ — sample sum of squares of y

$s_{x_j}^2 = \sum_{k=1}^{n} (x_{jk} - \bar{x}_{j\cdot})^2$, $j = 1,2,\ldots,p$ — sample sum of squares of x_j.

$s_{x_j y} = \sum_{k=1}^{n} (x_{jk} - \bar{x}_{j\cdot})(y_k - \bar{y})$, $j = 1,2,\ldots,p$ — sample sum of product of y and x_j.

$\rho_{xy} = \frac{s_{xy}}{s_x s_y}$ — sample correlation coefficient of x and y

$\bar{x}_{ij} = \frac{\sum_{k=1}^{n_i} x_{ijk}}{n_i}$, $i = 1,2,\ldots,H$, $j = 1,2,\ldots,p$ — i-th stratum sample mean of j-th auxiliary variate

$\hat{\bar{Y}}$ — ξ-best linear unbiased predictor of mean of Y

$\hat{\beta}_j$, $j = 1,2,\ldots,p$ — estimate of regression coefficient

$\hat{\mu}_i$, $i = 1,2,\ldots,H$ — estimate of the effect of i-th stratum

$T_s^t = (1 \ldots\ldots 1)_{1xn}$

$T_{s_i}^t = (1 \ldots\ldots 1)_{1xn_i}$, $i = 1,2,\ldots,H$

$T_{\bar{s}}^t = (1 \ldots\ldots 1)_{1x(N-n)}$

$T_{\bar{s}_i}^t = (1 \ldots\ldots 1)_{1 \times (N_i - n_i)}$, $i = 1,2,\ldots,H$

$T_{s_t}^t = (T_{s_1}^t \mid \ldots\ldots \mid T_{s_H}^t)_{1 \times \sum_{i=1}^{H} n_i}$

$$\underset{\sim}{T}_{s_t}^{t} = (\underset{\sim}{T}_{s_1}^{t} \mid \ldots\ldots \mid \underset{\sim}{T}_{s_H}^{t})_{1 \times \sum_{i=1}^{H} (N_i - n_i)}$$

$$\underset{\sim}{m}_{\tilde{s}}^{t} = (N-n, \underset{\sim}{m}_{s_1}, \ldots, \underset{\sim}{m}_{s_p})$$

$$\underset{\sim}{m}_{s_j}^{t} = \sum_{k=1}^{N-n} (x_{jk} - \bar{X}_{j.}), \quad j = 1,2,\ldots,p$$

$$\underset{\sim}{m}_{\tilde{s}_i}^{t} = (N_i - n_i, \underset{\sim}{m}_{s_{i1}}, \ldots, \underset{\sim}{m}_{s_{ip}}), \quad i = 1,2,\ldots,H$$

$$\underset{\sim}{m}_{s_{ij}} = \sum_{k=1}^{N_i - n_i} (x_{ijk} - \bar{X}_{.j.}), \quad j = 1,2,\ldots,p$$

$$\underset{\sim}{m}_{\tilde{s}_t}^{t} = (N_1 - n_1, \ldots, N_H - n_H, \underset{\sim}{m}_{s_{t1}}, \ldots, \underset{\sim}{m}_{s_{tp}})$$

$$\underset{\sim}{m}_{s_{tj}} = \sum_{i=1}^{H} \sum_{k=1}^{N_i - n_i} (x_{ijk} - \bar{X}_{.j.}), \quad j = 1,2,\ldots,p$$

ABSTRAK

Di dalam tesis ini, kami akan pertimbangkan penganggaran dalam populasi-populasi terhingga dengan suatu pendekatan super-populasi berasaskan model, di mana tatacara-tatacara ramalan kuasa dua terkecil dan ramalan Bayesan digunakan. Perhatian yang lebih akan ditumpukan kepada pentaabiran Bayesan. Bab pertama meliputi suatu kajian ringkas am mengenai pendekatan populasi tetap 'klasik' dan pendekatan superpopulasi berasaskan model, kerangka rekabentuk tinjauan am yang digunakan serta suatu tinjauan semula bahan-bahan penulisan yang berkaitan dengan penyelidikan ini.

Di dalam Bab 2, kami kaji semula kesudahan-kesudahan Cassel, Sarndal & Wretman (1977) [Thm. 5.7. Pg. 127] dan Sekkappan (1983), dalam pendekatan ramalan kuasa dua terkecil untuk penerbitan peramal saksama linear terbaik-ξ bagi min populasi, $\bar{Y}$ dan min ralat kuasa dua terhadap taburan superpopulasi, ξ. Dengan mengguna kesudahan-kesudahan ini, kami tunjukkan secara berangka, untuk kes-kes tertentu, bagaimana penganggar-penganggar diperolehi. Di-tunjukkan juga, secara berangka,bahawa model regressi berganda yang dipertimbangkan akan suai dengan sebarang model regressi polinomial.

Dengan mengikut suatu pendekatan Bayesan dalam Bab 3, kami dapatkan rumus-rumus untuk min dan varians posterior bagi min populasi terhingga dengan mengguna suatu model regressi heteros-kedastik linear yang mempunyai p pembolehubah bantu, dan itlakkan kesudahan-kesudahan Sekkappan (1969 I) untuk prior-prior am atau baur.

Di dalam Bab 4, model regressi linear diperluaskan kepada suatu populasi berstratum. Pendekatan ramalan Bayesan digunakan untuk menganggar min populasi terhingga mengikut suatu model

analisis kovarians berganda dengan ralat heteroskedastik bila terdapat maklumat mengenai lebih dari satu pembolehubah bantu untuk populasi berstratum. Kami pertimbangkan dua kes: bila parameter-parameter regressi yang berkaitan dengan cerapan-cerapan di dalam sebarang stratum adalah

1. sama untuk semua stratum
2. berbeza di antara stratum

Di sini kami tunjukkan bahawa kesudahan-kesudahan dari Sekkappan (1983) boleh diperolehi dari kesudahan-kesudahan Bayesan kami untuk prior baur.

Di dalam model regressi berganda dan model analisis kovarians berganda, sungguhpun anggaran-anggaran yang diperolehi menggunakan pendekatan Bayesan untuk prior-prior baur telah ditunjukkan berhampiran dengan kesudahan-kesudahan bagi pendekatan 'klasik' dan ramalan kuasa dua terkecil, varians bagi anggaran-anggaran akan berbeza dengan pendekatan yang digunakan.

Di dalam Bab 4, adalah dianggapkan bahawa saiz-saiz stratum, N_i ($i = 1,2,\ldots, H$) diketahui. Untuk keadaan-keadaan tertentu, sungguhpun saiz populasi terhingga, N diketahui, saiz stratum, N_i mungkin tidak diketahui. Kesudahan-kesudahan selari yang sepadan dengan penganggar-penganggar yang didapati dalam Bab 4 diterbitkan di dalam Bab 5 melalui pensampelan ganda dua untuk penstratuman. Kesudahan-kesudahan kami mengitlakkan kesudahan-kesudahan Sekkappan (1984) dan Bahadur Singh & Sedransk (1977). Penganggaran min populasi terhingga bila terdapat keadaan tanpa sambutan juga dibincangkan.

ABSTRACT

In this thesis, we shall consider estimation in finite populations under a model-based superpopulation approach using least-squares and Bayesian prediction procedures. Greater emphasis will be placed on the Bayesian inference. The first chapter covers a general preview of the 'classical' fixed population approach and the superpopulation model-based approach, the general survey design framework used and a literature review relevant to this study.

In Chapter 2, we review, respectively, the result of Cassel, Sarndal & Wretman (1977) [Thm. 5.7, Pg. 127] and Sekkappan (1983) under the least-squares prediction approach towards the derivation of the ξ-best linear unbiased predictor of the population mean, $\bar{Y}$ and its mean square error, with respect to the super-population distribution, ξ. Using these results, we shown numerically, for particular cases, how these estimates are obtained. It is also numerically shown that the multiple regression model considered will fit any polynomial regression model.

Follwing a Bayesian approach in Chapter 3, we obtain the formulae for the posterior mean and variance of the finite population mean using a linear heteroscedastic regression model with p auxiliary variables and hence generalize the results of Sekkappan (1982), Cassel, Sarndal & Wretman (1977) and Ericson (1969 I) for either general or diffuse priors.

In Chapter 4, the linear regression model is extended to a stratified population. Here, the Bayesian prediction approach is used to estimate the finite population mean under a multiple analysis of covariance model with heteroscedastic errors, when information on more than one auxiliary variable is available for stratified populations. We consider two cases: when the regression parameters

associated with the observations in any stratum are

1. the same for all strata
2. different for different stratum.

Here, we show that the results of Sekkappan (1983) can be obtained from our Bayesian results for a diffuse prior.

Under both the multiple regression model and the multiple analysis of covariance model, although the estimates obtained using the Bayesian approach for diffuse priors are shown to agree closely with the usual classical and least-squares prediction results, the mean square errors of these estimates under different approaches will differ.

In Chapter 4, it is assumed that the stratum sizes, N_i, ($i = 1,2, \ldots, H$) are known. However, in certain situations, although the finite population size, N is known, the stratum sizes, N_i maybe unknown. The parallel results corresponding to the various estimators obtained in Chapter 4 are derived in Chapter 5 using the double sampling technique for stratification. Our results include the results of Sekkappan (1984) and Bahadur Singh & Sedransk (1977) as particular cases. The estimation of the finite pouplation mean when there is non-response is also discussed.

CHAPTER 1

INTRODUCTION

1.1 Preliminaries

The estimation procedure in finite population sampling can be studied through two approaches, namely the Fixed Population Approach and the Superpopulation approach. Under the fixed population approach, which is the 'traditional' one in classical survey sampling, with each population unit is associated a fixed but unknown real number, that is, the value of the interested variable. However, following the superpopulation approach, with each population unit is associated a random variable for which a stochastic structure is specified and the actual value associated with a population unit is treated as the outcome of this random variable.

In this thesis, we shall deal with estimation in finite population sampling using Bayesian and Least-squares prediction procedures under the superpopulation approach. The prediction procedures under the superpopulation approach allows the superpopulation distribution, ξ an essential role in inference. A relevant point to remember in statistical analysis under this approach is that superpopulation models should not be equated with only Bayesian models.

Under the Bayesian approach, ξ which specifies the joint distribution of $(Y_1, \ldots, Y_N)$ is taken as a prior distribution from which the posterior distribution of the unobserved coordinates of $(Y_1, \ldots, Y_N)$, given the sample, s, is derived. Alternatively, under the least-squares prediction approach, ξ maybe assumed to contain unknown parameters which

must first be estimated using tools of classical inference, such as the least-squares method, the method of maximum likelihood, etc.

Definition (i): $\hat{\bar{Y}}$ is called the ξ-unbiased predictor of $\bar{Y}$, if any only if, for a given super-population distribution, ξ,

$$E_\xi(\hat{\bar{Y}} - \bar{Y}) = 0, \text{ for all } s \in S$$

Definition (ii): $\hat{\bar{Y}}$ is called a $p\xi$-unbiased predictor of $\bar{Y}$, if and only if, for a given sampling design, p and a superpopulation distribution, ξ,

$$E_\xi E(\hat{\bar{Y}} - \bar{Y}) = 0, \text{ for all } s \in S$$

In our study, we shall deal with the model-based approach, that is,the approach where with ξ as the essential element for inference, the sample, s is treated as given and the design, p that produced s is of minor interest. Therefore, our main interest will be to choose $\hat{\bar{Y}}$, for a given s, to minimize $E_\xi(\hat{\bar{Y}} - \bar{Y})^2$ and beyond this, averaging with respect to p is secondary. This implies that, under the model-based approach, the problem of finding a predictor, $\hat{\bar{Y}}$, which is good for any sample actually obtained is more important than just a good strategy $(p, \bar{Y})$.

Since the predictor $\hat{\bar{Y}}_{\bar{s}}$ of $\bar{Y}_{\bar{s}}$, the mean of the unobserved coordinates of the response variable, is related to the predictor $\hat{\bar{Y}}$ of the overall mean $\bar{Y}$ through

$$\hat{\bar{Y}} = \frac{1}{N}\left[n\bar{y} + (N - n)\hat{\bar{Y}}_{\bar{s}}\right] \quad \ldots\ldots\ldots \quad (1.1.1)$$

then, $\hat{\bar{Y}}$ is ξ-unbiased as a predictor of $\bar{Y}$, if and only if, for every $s \in S$, $\hat{\bar{Y}}_s$ is ξ-unbiased as a predictor of $\bar{Y}_s$.

A more detailed theory on the least squares prediction approach and the Bayesian approach is given in Chapter 2, Section 2.2, and Chapter 3, Section 3.2, respectively.

The Bayesian inference under quadratic loss consists simply of deriving the posterior expectation of $\bar{Y}$, given the data $(s, \underset{\sim}{y})$. Hence, we arrive at the Bayes' estimate

$$E[\bar{Y} \mid (s, \underset{\sim}{y})] = \frac{\sum_{i \in s} y_i + \sum_{i \in s} E(Y_i \mid (s, \underset{\sim}{y}))}{N} \qquad \ldots\ldots\ldots \quad (1.1.2)$$

1.2 Sample Survey Model

The basic model for survey sampling in the problem of estimation is presented here. A given population, U consists of N units indexed by $k = 1,2,\ldots,N$. We denote by Y_k the real variate associated with the unit k, $(k = 1,2,\ldots,N)$ of the population. Then $\underset{\sim}{Y} = (Y_1, \ldots, Y_N)$ is a point on the Euclidean space, R_N. Any function of R_N is called a parametric function, example, the parametric function $\bar{Y}$ is said to be the population mean if

$$\bar{Y} = \frac{1}{N} \sum_{k=1}^{N} Y_k \text{ , for all } \underset{\sim}{Y} \in R_N$$

Our interest will be to estimate $\bar{Y}$ on the basis of the units sampled from the population U and the observed values y_k $k = 1,2,\ldots,n$, associated with them, that is, on the basis of $(s, y_k : k \in s)$ where s is a subset of U, drawn from U with a given sampling design, p.

A sampling design, $p(s)$ is any function of S, the set of all possible subsets s of U, such that $p(s) \geq 0$ and $\Sigma\, p(s) = 1$, for $s \in S$.

1.3 Likelihood Function

Given the sampling design, (S, p) and the population vector, $\underset{\sim}{Y} = (Y_1, \ldots, Y_N)$, the probability of obtaining the data $(s, y_k : k \in s)$ can be shown as

$$\text{Prob}[(s, y_k : k \in s) \,|\, Y, (S, p)] = \begin{cases} p(s), & \text{if } Y_k = y_k, \text{ for all } k \in s \\ 0, & \text{elsewhere} \end{cases} \qquad \ldots\ldots\ldots(1.3.1)$$

Now, with the parameter space, R_N and parameter vector $\underset{\sim}{Y}$, the likelihood function, ℓ for $\underset{\sim}{Y}$ given $(s, y_k : k \in s)$ is

$$\ell(\underset{\sim}{Y} \,|\, (s, y_k : k \in s)) = \begin{cases} p(s), & \text{if } \underset{\sim}{Y} \in R_N(y_k : k \in s) \\ 0, & \text{if } \underset{\sim}{Y} \notin R_N(y_k : k \in s) \end{cases} \qquad \ldots\ldots\ldots(1.3.2)$$

where $\underset{\sim}{Y} \in R_N(y_k : k \in s)$, if and only if

$Y_k = y_k$, for all $k \in s$

and $R_N(y_k : k \in s) \subset R_N$

1.4 Literature Review

The concept of generalizing from a small 'section' of the population to the 'whole' has been used rather subjectively in various population studies. In order to ascertain how objective the generalization is, two questions which arise are:

(i) how to select the 'section'

(ii) how to form the generalization

Basically, the answers to these questions would be to find a combination of selection and estimation techniques which minimize the cost incurred while at the same time optimize the precision of the inference.

During the 1930's, simple ratio and regression estimation were introduced and Cochran (1942) consolidated the theories on this estimation. If we make inferences about a response variable, ignoring the availability of other correlated auxiliary variables, we will sacrifice a lot of information and so Olkin (1958) extended the ratio estimation for finite populations to a multivariate model where he assumed the population means of the auxiliary variables to be known and nonzero. Further, he considered a stratified population for a univariate case and dealt with the two particular ratio estimators, namely

(i) the separate ratio estimate

(ii) the combined ratio estimate.

Cochran (1963) compiled the results of Olkin (1958) as well as the estimation under some linear regression models following the fixed population approach. He obtained the combined and separate ratio and regression estimates in a stratified population and adopted the double sampling technique when the population mean of the auxiliary variable is unknown.

When the population means of the auxiliary variables are known, Shukla (1965) derived the multivariate regression estimate. For the case when the population means of auxiliary variables are unknown, Shukla (1965) obtained a double sampling regression estimate.

The inference problem began to take an approach where the finite population was assumed to be generated from an infinite superpopulation. Through prior knowledge, if this superpopulation was known to be of a specific parametric form, example a gamma or normal distribution, then the values of the units not found in the selected sample could be predicted through this distribution of the superpopulation. This is known as predictive statistical inference. Ericson (1969 I & 1969 II) and Scott & Smith (1969) have used this idea in a Bayesian framework for inference.

A non-Bayesian prediction approach utilizing the least-squares method, under some multiple linear regression models in finite populations were considered by Royall (1970), (1971) and Brewer (1963). Cassel, Sarndal and Wretman (1977) derived a ξ-best linear unbiased predictor of the population mean under a multiple regression model using a model-based approach where ξ is the superpopulation distribution. Holt & Smith (1979) used the method of least-squares for prediction in a one way analysis of variance model. Under a multiple analysis of covariance model, Sekkappan (1983) generalised the results of Royall (1970), Brewer (1963) and Holt & Smith (1979) to obtain the ξ-best linear unbiased predictors of $\bar{Y}$ and their respective mean square errors for the combined and separates regression estimates.

Under the Bayesian approach, Ericson (1969 I) considered estimation under a linear regression model in finite populations. Ericson (1969 I) also gave properties of the posterior distributions for some special cases obtained by assuming the conditional density of a random variate, that is, $f(y_k|\theta)$,

$k = 1,2,\ldots,N$ where θ is a real or vector valued parameter, to be a normal with either known or unknown variance and by taking $dF(\theta)$ to be either a natural conjugate or a uniform prior. In the same paper, Ericson extended his results to the case involving the use of concomitant measurements and showed that under this Bayesian model, several ratio and regression estimators arise as the means of posterior distributions. Later, in Ericson (1969 II), the Bayesian superpopulation approach was extended for stratification. Sekkappan (1981 I) generalized the Bayesian models considered in Ericson (1969 I) for K characteristics, which was extended to a stratified population in Sekkappan (1981 II). For an analysis of covariance model, Sekkappan (1982) obtained a regression estimate of the population mean when the regression parameters are different in different strata.

Various attempts have been made to identify methods of obtaining a high degree of response since it is evident that the problem of nonresponse exists in most surveys. Bahadur Singh and Sedransk (1977) considered nonresponse in the context of a two phase sampling design, assuming post stratification. This result was generalized by Sekkappan (1984) under a multiple analysis of covariance model when the regression parameters are different for different strata. Earlier, Draper & Guttman (1968 I) and Mohd. Zubair Khan (1976) showed that for a stratified population, a two phase Bayesian sampling scheme will improve the Bayes' estimates of unknown parameters in a regression model.

1.5 Summary of Results

This thesis is primarily an extension of the work of Bahadur Singh & Sedransk (1977) and Sekkappan (1981 I), (1982), (1984) for finding the Bayes' estimate for finite population mean. The work also involves an extension of Royall (1970, 1971), in the sense that we have studied some aspects of the estimation problem with numerical examples using the least-squares prediction approach. Each chapter has its own summary at the beginning.

CHAPTER 2

LEAST-SQUARES PREDICTION APPROACH

2.1 Summary

In this chapter, we will review the results of Cassel, Sarndal & Wretman (1977) [Thm. 5.7, Pg. 127] and Sekkappan (1983) under the multiple regression model (2.3.1) and the multiple analysis of covariance models (2.6.1) and (2.8.1), respectively, regarding the derivation of the ξ-best linear unbiased (ξ-BLU) predictor of $\bar{Y}$ and its mean square error (MSE) with respect to the superpopulation distribution, ξ. Based on the results of Cassel, Sarndal & Wretman (1977), the details towards obtaining the ξ-BLU predictor of $\bar{Y}$ when $p = 2$ and $p = 1$ together with their respective MSE are given. The classical ratio estimator will be deduced by assuming certain constraints in the model (2.3.1). As a special case, the model (2.3.1) is shown to fit any polynomial regression model. For the particular cases of $p = 2$ and $p = 1$, we provide numerical derivations of the ξ-BLU predictor of $\bar{Y}$ and its MSE, based on the results of Cassel, Sarndal & Wretman (1977) and Sekkappan (1983).

2.2 Introduction

In the superpopulation approach, the vector of population values $\underset{\sim}{y} = (y_1,\ldots,y_N)$ is assumed to be the realised outcome of a vector of random variables $\underset{\sim}{Y} = (Y_1,\ldots,Y_N)$. The joint distribution of $Y_1,\ldots,Y_N$ will be denoted by ξ, about which certain features are assumed known. For a superpopulation model, a specified set of conditions define a class of distributions to which the superpopulation distribution, ξ is

assumed to belong. In the model-based approach, we distinguish two main approaches to inference under superpopulation assumptions:

1. models leading to the use of classical inference tools which we deal with in this chapter
2. models leading to the use of Bayesian inference tools which we deal with in the Chapters 3,4 and 5.

In the classical inference version, the observed y_k are first used to make inference about a possibly unknown parameter vector $\underset{\sim}{\theta}$ indexing ξ. Finally, ξ is used to predict the mean $\bar{Y}_s$ of the unobserved units.

In this chapter, we shall consider the classical inference of the prediction approach to a multiple regression model and a multiple analysis of covariance model.

2.3 Multiple Regression Estimate

The class of probability measures, ξ, on R_N is such that $Y_1, \ldots, Y_N$ are independently distributed and

$$E(Y_k) = \mu + \beta_1(X_{1k} - \bar{X}_{1.}) + \ldots + \beta_p(X_{pk} - \bar{X}_{p.})$$

$$V(Y_k) = \lambda_k \sigma^2$$

where

$\beta_1, \ldots, \beta_p$ and σ^2 are unknown

$X_{1k}, \ldots, X_{pk}, \lambda_k$ is a set of known numbers for every $k(k = 1,2, \ldots, N)$.

Then, the model for a finite population would be

$$\underset{\sim}{Y}_{Nx1} = Z_{Nx(p+1)}\underset{\sim}{\beta}_{(p+1)x1} + \underset{\sim}{e}_{Nx1} \quad \ldots\ldots\ldots\ldots \quad (2.3.1)$$

where

$$\underset{\sim}{Y} = (Y_1, \ldots, Y_N)^t$$

$$Z = \begin{pmatrix} 1 & (X_{11} - \bar{X}_{1.}) & \ldots & (X_{p1} - X_{p.}) \\ \vdots & \vdots & & \vdots \\ 1 & (X_{1N} - \bar{X}_{1.}) & \ldots & (X_{pN} - X_{p.}) \end{pmatrix}$$

$$\underset{\sim}{\beta} = (\mu, \beta_1, \ldots, \beta_p)^t$$

The vector of random disturbances, $\underset{\sim}{e}$, is distributed with a vector of zero means and a variance covariance matrix $\sigma^2\eta^{-1}_{NxN}$ where,

$$\eta^{-1}_{NxN} = \begin{pmatrix} \lambda_1 & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & \lambda_N \end{pmatrix} \quad \ldots\ldots\ldots \quad (2.3.2)$$

with $\lambda_k = g(x_{1k}, \ldots, x_{pk})$, $k = 1,2,\ldots,N$

Given a sample of size n, the model maybe partitioned as

$$\begin{pmatrix} \underset{\sim}{Y}_{snx1} \\ \hline \underset{\sim}{Y}_{\bar{s}N-nx1} \end{pmatrix} = \begin{pmatrix} Z_{snxp+1} \\ \hline Z_{\bar{s}N-nxp+1} \end{pmatrix} \underset{\sim}{\beta} + \begin{pmatrix} \underset{\sim}{e}_{snx1} \\ \hline \underset{\sim}{e}_{\bar{s}N-nx1} \end{pmatrix}$$

$$\ldots\ldots\ldots \quad (2.3.3)$$

with $\operatorname{cov}(\underset{\sim}{e}) = \sigma^2 \left(\begin{array}{c|c} \eta^{-1}_{nxn} & \underset{\sim}{0}_{nxN-n} \\ \hline \underset{\sim}{0}_{N-nxn} & \eta^{-1}_{N-nxN-n} \end{array}\right)$

Hence, after selecting the sample, s, of size n, the model for the observed values is

$$\underset{\sim}{Y}_s = Z_s \underset{\sim}{\beta} + \underset{\sim}{e}_s \qquad \ldots\ldots\ldots (2.3.4)$$

where

$$E(\underset{\sim}{e}_s) = 0$$

$$D(\underset{\sim}{e}_s) = \sigma^2 \eta_s^{-1}$$

By the least-square method, the generalised least-squares estimate of $\underset{\sim}{\beta}$ is

$$\hat{\underset{\sim}{\beta}} = (Z_s^t \eta_s Z_s)^{-1}(Z_s^t \eta_s \underset{\sim}{Y}_s) \qquad \ldots\ldots\ldots (2.3.5)$$

and $$D(\hat{\underset{\sim}{\beta}}) = (Z_s^t \eta_s Z_s)^{-1} \sigma^2 \qquad \ldots\ldots\ldots (2.3.6)$$

$$\hat{\sigma}^2 = \frac{1}{n-(p+1)} \left[\underset{\sim}{Y}_s^t \eta_s \underset{\sim}{Y}_s - \hat{\underset{\sim}{\beta}}^t Z_s^t \eta_s \underset{\sim}{Y}_s \right] \ldots (2.3.7)$$

Under the multiple regression model (2.3.1), the ξ-BLU predictor of the population mean $\bar{Y}$, for any design p, is given by

$$\hat{\bar{Y}} = \frac{1}{N} \left[\sum_{k \epsilon s} y_k + \sum_{k \epsilon \bar{s}} \hat{Y}_k \right]$$

$$= \frac{1}{N} \left[n\bar{y} + T_{\bar{s}}^t Z_{\bar{s}} \hat{\underset{\sim}{\beta}} \right] \qquad \ldots\ldots\ldots (2.3.8)$$

where $$\hat{Y}_{\bar{s}} = Z_{\bar{s}} \hat{\underset{\sim}{\beta}}$$

and $$T_{\bar{s}}^t = (1 \ldots 1)_{1xN-n}$$

We can simplify (2.3.8) further to obtain

$$\hat{\bar{Y}} = \frac{1}{N} \left[n\bar{y} + (N-n) \underset{\sim}{m}_{\bar{s}}^t \hat{\underset{\sim}{\beta}} \right] \qquad \ldots\ldots\ldots (2.3.9)$$

since

$$T_{\bar{s}}^{t} Z_{\bar{s}} = \left(N-n, \sum_{k=1}^{N-n} (x_{1k} - \bar{X}_{1.}), \ldots, \sum_{k=1}^{N-n} (x_{pk} - \bar{X}_{p.})\right)$$

$$= (N-n, m_{\bar{s}1}, \ldots, m_{\bar{s}p})$$

$$= m_{\tilde{s}}^{t}$$

with

$$m_{\bar{s}j} = \sum_{k=1}^{N-n} (x_{jk} - \bar{X}_{j.})$$

$$= \sum_{k=1}^{N-n} x_{jk} - (N-n)\bar{X}_{j.}$$

$$= N \bar{X}_{j.} - n \bar{x}_{j.} - (N-n)\bar{X}_{j.}$$

$$= n(\bar{X}_{j.} - \bar{x}_{j.}) \qquad \text{for } j = 1,2,\ldots,p$$

The mean square error of $\hat{\bar{Y}}$, with respect to the superpopulation distribution ξ, is

$$MSE(\hat{\bar{Y}}) = E_{\xi}[\hat{\bar{Y}} - \bar{Y}]^2$$

$$= \frac{1}{N^2}\left[T_{\bar{s}}^{t} D(\hat{Y}_{\bar{s}}) T_{\bar{s}}\right]$$

$$= \frac{1}{N^2}\left[T_{\bar{s}}^{t} E\{(\hat{Y}_{\bar{s}} - Y_{\bar{s}})(\hat{Y}_{\bar{s}} - Y_{\bar{s}})^{t}\}T_{\bar{s}}\right]$$

$$= \frac{1}{N^2}\left[T_{\bar{s}}^{t} E\{(Z_{\bar{s}}\hat{\beta}_{\tilde{}} - Z_{\bar{s}}\beta_{\tilde{}} - e_{\tilde{\bar{s}}})(Z_{\bar{s}}\hat{\beta}_{\tilde{}} - Z_{\bar{s}}\beta_{\tilde{}} - e_{\tilde{\bar{s}}})^{t}\}T_{\bar{s}}\right]$$

$$= \left[\frac{1}{N^2}\left[T_{\bar{s}}^{t} E\{Z_{\bar{s}}(\hat{\beta}_{\tilde{}} - \beta_{\tilde{}})(\hat{\beta}_{\tilde{}} - \beta_{\tilde{}})^{t} Z_{\bar{s}}^{t} + e_{\tilde{\bar{s}}} e_{\tilde{\bar{s}}}^{t}\}T_{\bar{s}}\right]\right.$$

$$= \frac{1}{N^2}\left[T_{\bar{s}}^{t}\{Z_{\bar{s}} \operatorname{cov}(\hat{\beta}_{\tilde{}}) Z_{\bar{s}}^{t} + \operatorname{cov}(e_{\tilde{\bar{s}}})\}T_{\bar{s}}\right]$$

$$= \frac{\sigma^2}{N^2}\left[T_{\bar{s}}^{t} Z_{\bar{s}} (Z_s^{t} \eta_s Z_s)^{-1} Z_{\bar{s}}^{t} T_{\bar{s}} + T_{\bar{s}}^{t} \eta_{\bar{s}}^{-1} T_{\bar{s}}\right]$$

$$= \frac{\sigma^2}{N^2}\left[\underset{\sim}{m}_{\bar{s}}^{t}(Z_s^{t} \eta_s Z_s)^{-1} \underset{\sim}{m}_{\bar{s}} + \sum_{k=1}^{N-n} \lambda_k\right] \quad \ldots\ldots\ldots \quad (2.3.10)$$

2.4 Special Cases

Case I: Suppose $\lambda_1 = \lambda_2 = \ldots = \lambda_N = 1$, then η^{-1} becomes an identity matrix. Hence, in model (2.3.1)

$$D(\underset{\sim}{e}) = \sigma^2 \; I_{NxN}$$

The estimates (2.3.9) and (2.3.10) will still hold true but

$$\underset{\sim}{\hat{\beta}} = (Z_s^{t} Z_s)^{-1} Z_s^{t} \underset{\sim}{Y}_s \quad \ldots\ldots\ldots \quad (2.4.1)$$

and

$$\hat{\sigma}^2 = \frac{1}{n-(p+1)}\left[\underset{\sim}{Y}_s^{t} \underset{\sim}{Y}_s - \underset{\sim}{\hat{\beta}}^{t} Z_s^{t} \underset{\sim}{Y}_s\right]$$

Case II: Suppose p = 2 in Case I, then

$$Z = \begin{pmatrix} 1 & (X_{11} - \bar{X}_{1.}) & (X_{21} - \bar{X}_{2.}) \\ \vdots & \vdots & \vdots \\ 1 & (X_{1N} - \bar{X}_{1.}) & (X_{2N} - \bar{X}_{2.}) \end{pmatrix}$$

$$\underset{\sim}{\beta} = (\mu, \beta_1, \beta_2)^{t}$$

Given the sample and observed variate values $y_k (k = 1,2,\ldots,n)$, the estimates of μ, β_1, β_2 can be easily found through (2.4.1) where

$$(Z_s^t Z_s)^{-1} = \frac{1}{\left|Z_s^t Z_s\right|}\begin{pmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{pmatrix}$$

$$= \begin{pmatrix} K_{11}' & K_{12}' & K_{13}' \\ K_{21}' & K_{22}' & K_{23}' \\ K_{31}' & K_{33}' & K_{33}' \end{pmatrix}$$

with the determinant

$$|Z_s^t Z_s| = n s_{x_1}^2 s_{x_2}^2 (1 - \rho_{x_1x_2}^2)$$

and

$$K_{11}' = \frac{1}{n} + \frac{(\bar{x}_{2.} - \bar{X}_{2.})^2}{(1-\rho_{x_1x_2}^2)s_{x_2}^2} + \frac{(\bar{x}_{1.} - \bar{X}_{1.})^2}{(1-\rho_{x_1x_2}^2)s_{x_1}^2}$$

$$- \frac{2\rho_{x_1x_2}^2 (\bar{x}_{1.} - \bar{X}_{1.})(\bar{x}_{2.} - \bar{X}_{2.})}{s_{x_1x_2}(1 - \rho_{x_1x_2}^2)}$$

$$K_{12}' = K_{21}' = \frac{\rho_{x_1x_2}^2 (\bar{x}_{2.} - \bar{X}_{2.})}{s_{x_1x_2}(1-\rho_{x_1x_2}^2)} - \frac{(\bar{x}_{1.} - \bar{X}_{1.})}{s_{x_1}^2 (1-\rho_{x_1x_2}^2)}$$

$$K_{13}' = K_{31}' = \frac{\rho_{x_1x_2}^2 (\bar{x}_{1.} - \bar{X}_{1.})}{s_{x_1x_2}(1-\rho_{x_1x_2}^2)} - \frac{(\bar{x}_{2.} - \bar{X}_{2.})}{s_{x_2}^2 (1 - \rho_{x_1x_2}^2)}$$

$$K_{22}' = \frac{1}{s_{x_1}^2 (1 - \rho_{x_1x_2}^2)}$$

$$K_{23}' = K_{32}' = \frac{-\rho_{x_1x_2}^2}{s_{x_1x_2}(1-\rho_{x_1x_2}^2)}$$

$$K'_{33} = \frac{1}{s^2_{x_2}(1 - \rho^2_{x_1x_2})}$$

On substituting and simplyfying (2.4.1), we obtain

$$\hat{\mu} = \bar{y} - \hat{\beta}_1(\bar{x}_{1.} - \bar{X}_{1.}) - \hat{\beta}_2(\bar{x}_{2.} - \bar{X}_{2.}) \quad \ldots\ldots\ldots (2.4.2)$$

$$\hat{\beta}_1 = \frac{s_{x_1y}}{s^2_{x_1}(1 - \rho^2_{x_1x_2})} - \frac{\rho^2_{x_1x_2}\, s_{x_2y}}{s_{x_1x_2}(1 - \rho^2_{x_1x_2})} \quad \ldots\ldots\ldots (2.4.3)$$

$$\hat{\beta}_2 = \frac{s_{x_2y}}{s^2_{x_2}(1 - \rho^2_{x_1x_2})} - \frac{\rho^2_{x_1x_2}\, s_{x_1y}}{s_{x_1x_2}(1 - \rho^2_{x_1x_2})} \quad \ldots\ldots\ldots (2.4.4)$$

noting the fact,

$$Z_s^t \underset{\sim}{Y}_s = \left(\sum_{k\epsilon s} y_k,\ s_{x_1y} + (\bar{x}_{1.} - \bar{X}_{1.}) \sum_{k\epsilon s} y_k,\ s_{x_2y} + (\bar{x}_{2.} - \bar{X}_{2.}) \sum_{k\epsilon s} y_k \right)^t$$

and $s^2_{x_1}$, $s^2_{x_2}$, s_{x_1y}, s_{x_2y}, $s_{x_1x_2}$ and $\rho^2_{x_1x_2}$ are defined in the list of notations.

Then, the estimate (2.3.9) becomes

$$\hat{\bar{Y}} = \frac{1}{N}\left\{ n\bar{y} + (N-n)\left[\hat{\mu} + \frac{\hat{\beta}_1 n(\bar{X}_{1.} - \bar{x}_{1.})}{N-n} + \frac{\hat{\beta}_2 n(\bar{X}_{2.} - \bar{x}_{2.})}{N-n} \right] \right\}$$

which after substituting (2.4.2), (2.4.3) and (2.4.4) will become

$$\hat{\bar{Y}} = \bar{y} + \hat{\beta}_1(\bar{X}_{1.} - \bar{x}_{1.}) + \hat{\beta}_2(\bar{X}_{2.} - \bar{x}_{2.}) \quad \ldots\ldots\ldots (2.4.5)$$

Following this, (2.3.10) becomes

$$MSE(\hat{\bar{Y}}) = \frac{\sigma^2}{N^2}\left[(N-n) + \underset{\sim}{m}_{\bar{s}}^t (Z_s^t Z_s)^{-1} \underset{\sim}{m}_{\bar{s}} \right] \quad \ldots\ldots\ldots (2.4.6)$$

with $\underset{\sim}{m}_{\bar{s}}^t = (N-n, n(\bar{X}_{1.} - \bar{x}_{1.}), n(\bar{X}_{2.} - \bar{x}_{2.}))$

After substituition, factorization and simplification, (2.4.6) becomes,

$$MSE(\hat{\bar{Y}}) = \sigma^2 \frac{N-n}{Nn} + \frac{(\bar{x}_{1.} - \bar{X}_{1.})^2}{s^2_{x_1}(1-\rho^2_{x_1x_2})} + \frac{(\bar{x}_{2.} - \bar{X}_{2.})^2}{s^2_{x_2}(1-\rho^2_{x_1x_2})} - \frac{2\rho^2_{x_1x_2}(\bar{x}_{1.} - \bar{X}_{1.})(\bar{x}_{2.} - \bar{X}_{2.})}{s_{x_1x_2}(1-\rho^2_{x_1x_2})} \quad \ldots\ldots (2.4.7)$$

Example

In jute fibre crops for estimating mean yield of fibre per plant, two auxiliary variables have been measured, namely, height and base diameter, which are correlated with the yield of fibre. Population consisting of 50 jute plants (Capsulanes) sown at the Jute Agricultural Research Institute Farm, Barrackpore in the year 1962 - 63 were taken and data collected is given in Table 2.1 below. This example is taken from Shukla, G.K. (1965).

Table 2.1

Sr.No.	Plant Hgt. (ft) x_1	Base diameter (cm) x_2	Fibre weight (g) y	Sr.No.	Plant Hgt. (ft) x_1	Base diameter (cm) x_2	Fibre weight (g) y
1	7.08	1.3	5.0	26	6.83	1.5	7.0
2	7.00	1.3	5.5	27	7.17	1.7	7.5
3	7.08	1.4	5.5	28	6.83	1.4	6.5
4	5.67	1.3	3.5	29	7.50	1.8	9.0
5	5.75	1.2	4.0	30	6.25	1.2	5.0
6	6.08	1.3	6.0	31	6.67	1.5	6.0
7	6.83	1.2	4.5	32	6.42	1.4	4.5
8	7.58	1.4	6.5	33	6.25	1.7	4.5
9	6.33	1.5	5.0	34	6.58	1.3	5.0
10	6.17	1.3	4.5	35	7.25	1.2	6.0
11	6.75	1.4	5.5	36	7.33	1.3	6.5
12	6.25	1.3	4.5	37	5.33	1.5	4.5
13	5.92	1.3	3.5	38	6.92	1.4	6.0
14	6.00	1.3	4.5	39	6.75	1.4	5.0
15	7.25	1.5	7.0	40	7.25	1.5	7.0
16	5.50	1.3	4.5	41	5.58	1.6	6.5
17	6.83	1.2	5.5	42	7.42	1.5	7.5
18	6.67	1.2	4.5	43	7.08	1.4	7.0
19	6.83	1.2	5.5	44	7.00	1.7	6.5
20	7.33	1.5	7.5	45	7.33	1.4	7.0
21	5.25	1.4	3.5	46	7.58	1.6	7.5
22	6.17	1.4	4.5	47	6.42	1.7	5.5
23	7.25	1.7	9.5	48	6.83	1.6	7.0
24	7.17	1.6	6.5	49	6.00	1.5	4.0
25	7.08	1.2	5.0	50	6.17	1.5	4.5

$\bar{X}_{1.}$ = 6.65 ft. $\bar{X}_{2.}$ = 1.42 cm. $\bar{Y}$ = 5.69 gm.

To illustrate the method of estimation discussed above, the population units in Table 2.1 are used to select a sample of size of size n = 20, in order to estimate the mean yield of fibre per plant given information on two auxiliary variables correlated to the response variable. The sample data, given in Table 2.2 below, is selected using the random numbers tables.

Table 2.2

Sr. No.	x_1	x_2	y	Sr. No.	x_1	x_2	y
5	5.75	1.2	4.0	29	7.50	1.8	9.0
7	6.83	1.2	4.5	31	6.67	1.5	6.0
9	6.33	1.5	5.0	34	6.58	1.3	5.0
11	6.75	1.4	5.5	35	7.25	1.2	6.0
14	6.00	1.3	4.5	36	7.33	1.3	6.5
16	5.50	1.3	4.5	38	6.92	1.4	6.0
17	6.83	1.2	5.5	41	5.58	1.6	6.5
20	7.33	1.5	7.5	42	7.42	1.5	7.5
26	6.83	1.5	7.0	44	7.00	1.7	6.5
27	7.17	1.7	7.5	49	6.00	1.5	4.0

From the data above, we obtain the following sample values

$$\bar{x}_{1.} = 6.68 \qquad \bar{x}_{2.} = 1.43 \qquad \bar{y} = 5.925$$

$$s^2_{x_1} = 7.3779 \qquad s^2_{x_2} = 0.622$$

$$s_{x_1 y} = 11.31 \qquad s_{x_2 y} = 3.195$$

$$s_{x_1 x_2} = 0.502$$

$$\rho^2_{x_1 x_2} = \frac{s^2_{x_1 x_2}}{s^2_{x_1} s^2_{x_2}} = 0.0549$$

The estimates of β_1, β_2 and μ are obtained from (2.4.2), (2.4.3) and (2.4.4), respectively, as

$$\hat{\beta}_1 = \frac{11.31}{7.3779(1-0.0549)} - \frac{0.0549(3.195)}{0.502(1-0.0549)} = 1.2523$$

$$\hat{\beta}_2 = \frac{3.195}{0.622(1-0.0549)} - \frac{0.0549(11.31)}{0.502(1-0.0549)} = 4.1263$$

$$\hat{\mu} = 5.925 - 1.2523(0.03) - 4.1263(0.01) = 5.8462$$

Hence, the estimate of the mean, given by (2.4.5) is

$$\hat{\bar{Y}} = 5.925 + 1.2523(6.65 - 6.68) + 4.1263(1.42 - 1.43)$$

$$= 5.8462$$

whose mean square error, given by (2.4.6), is

$$MSE(\hat{\bar{Y}}) = \sigma^2 \left[\frac{30}{1000} + \frac{(0.03)^2}{7.3779(1-0.0549)} + \frac{(0.01)^2}{0.622(1-0.0549)} - \frac{2(0.0549)(0.03)(0.01)}{0.502(1-0.0549)} \right]$$

$$= \sigma^2(0.0302)$$

where

$$\hat{\sigma}^2 = \frac{1}{n-3} \left[\sum_{k=1}^{n} y_k^2 - \hat{\mu} \sum_{k=1}^{n} y_k - \hat{\beta}_1 \sum_{k=1}^{n} y_k(x_{1k} - \bar{X}_1) - \hat{\beta}_2 \sum_{k=1}^{n} y_k(x_{2k} - \bar{X}_2) \right]$$

$$= \frac{1}{17} \left[736.75 - 5.8462(118.5) - 1.2523(14.8650) - 4.1263(4.38) \right]$$

$$= 0.4286$$

Therefore,

$$MSE(\hat{\bar{Y}}) = 0.0129$$

Case III: Suppose p = 1 in Case I, then

$$Z = \begin{pmatrix} 1 & \cdots & 1 \\ (X_1 - \bar{X}) & \cdots & (X_N - \bar{X}) \end{pmatrix}^t$$

$$\underset{\sim}{\beta} = (\mu, \beta_1)^t$$

The estimate (2.4.1) will become

$$\hat{\underset{\sim}{\beta}} = \begin{pmatrix} n & n(\bar{x} - \bar{X}) \\ n(\bar{x} - \bar{X}) & \sum_{k=1}^{n} (x_k - \bar{X})^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{k=1}^{n} y_k \\ \sum_{k=1}^{n} y_k(x_k - \bar{X}) \end{pmatrix}$$

$$= (\bar{y} - \hat{\beta}_1(\bar{x} - \bar{X}), \hat{\beta}_1)^t \qquad \cdots\cdots\cdots\cdots (2.4.8)$$

after simplification,

where $$\hat{\beta}_1 = \frac{\sum_{k=1}^{n} (x_k - \bar{x}) y_k}{\sum_{k=1}^{n} (x_k - \bar{x})^2}$$

Hence, the estimate (2.3.9) will be

$$\hat{\bar{Y}} = \frac{1}{N}\left[n\bar{y} + (N-n)\left[\bar{y} - \hat{\beta}_1(\bar{x} - \bar{X}) + \frac{\hat{\beta}_1 n(\bar{X} - \bar{x})}{N-n}\right\}\right]$$

$$= \bar{y} + \hat{\beta}_1 (\bar{X} - \bar{x})$$

noting the fact, $Z_s^t \underset{\sim}{Y}_s = \left(\sum_{k\epsilon s} y_k, \sum_{k\epsilon s} (x_k - \bar{X}) y_k\right)^t$

$$= \left(\sum_{k\epsilon s} y_k, s_{xy} + (\bar{x} - \bar{X}) \sum_{k\epsilon s} y_k\right)^t$$

The estimate (2.4.9) is known in classical survey sampling as the regression estimator which is also obtained by Cassel, Sarndal & Wretman (1977), Example 2, Section 5.6.

The mean square error of $\hat{\bar{Y}}$ (2.4.9) will be

$$MSE(\hat{\bar{Y}}) = \sigma^2 \frac{(N-n)^2}{N^2}\left[\underset{\sim}{m}_s^t (Z_s^t Z_s)^{-1} \underset{\sim}{m}_s + \frac{1}{N-n}\right] \quad \ldots\ldots\ldots \quad (2.4.10)$$

where $\underset{\sim}{m}_s^t = \left(1, \frac{n(\bar{X} - \bar{x})}{N-n}\right)$

Formula (2.4.10) can be further simplified to obtain

$$MSE(\hat{\bar{Y}}) = \sigma^2 \left[\frac{N-n}{Nn} + \frac{(\bar{X} - \bar{x})^2}{s_x^2}\right]$$

Example

We use the same data given in Table 2.2, but for our interest, let the response variate, y be related to only one auxiliary variable, namely, plant height. Therefore, for a sample of size n = 20, the data given in Table 2.3 will be used

Table 2.3

Sr.No.	x	y	Sr.No.	x	y
5	5.75	4.0	29	7.50	9.0
7	6.83	4.5	31	6.67	6.0
9	6.33	5.0	34	6.58	5.0
11	6.75	5.5	35	7.25	6.0
14	6.00	4.5	36	7.33	6.5
16	5.50	4.5	38	6.92	6.0
17	6.83	5.5	41	5.58	6.5
20	7.33	7.5	42	7.42	7.5
26	6.83	7.0	44	7.00	6.5
27	7.17	7.5	49	6.00	4.0

We obtain the following sample values

$$\bar{x} = 6.680 \qquad \bar{y} = 5.925$$

$$s_x^2 = 7.3779$$

$$s_{xy} = 11.31$$

From these values, we obtain the estimates of

$$\hat{\beta}_1 = \frac{s_{xy}}{s_x^2} = \frac{11.31}{7.3779} = 1.5330$$

and from (2.4.8)

$$\hat{\mu} = 5.925 - 1.5330(6.68 - 6.65) = 5.8790$$

The estimate of the mean, as given by (2.4.9), is

$$\hat{\bar{Y}} = 5.925 + 1.5330(6.65 - 6.68)$$

$$= 5.8790$$

whose mean square error, as given by (2.4.11), is

$$MSE(\hat{\bar{Y}}) = \sigma^2 \left[\frac{30}{1000} + \frac{(0.03)^2}{7.3779}\right]$$

$$= \sigma^2 \ (0.0312)$$

where $\hat{\sigma}^2 = \frac{1}{n-2}\left[\sum_{k=1}^{n} y_k^2 - \hat{\mu}\sum_{k=1}^{n} y_k - \hat{\beta}_1 \sum_{k=1}^{n} y_k(x_k - \bar{x})\right]$

$= \frac{1}{18}\left[736.75 - 5.8790(118.5) - 1.5530(14.865)\right]$

$= 0.9611$

Therefore,

$$MSE(\hat{\bar{Y}}) = 0.03$$

Case IV: Suppose we let $\mu = \beta_1\bar{X}_{1.} + \beta_2\bar{X}_{2.} + \ldots + \beta_p\bar{X}_{p.}$ in model (2.3.1), then the model becomes a regression through the origin, that is

$$\underset{\sim}{Y}_{Nx1} = Z^*_{Nxp}\ \underset{\sim}{\beta}^*_{p+1} + \underset{\sim}{e}_{Nx1} \qquad \ldots\ldots\ldots\ldots \quad (2.4.12)$$

where

$$Z^* = \begin{pmatrix} X_{11} & \ldots\ldots & X_{p1} \\ \vdots & & \\ X_{1N} & \ldots\ldots & X_{pN} \end{pmatrix}$$

$$\underset{\sim}{\beta}^* = (\beta_1, \ldots, \beta_p)^t$$

The results (2.3.9) and (2.3.10) still hold good for model (2.4.12).

Case V: In Case IV, if η^{-1} is an identity matrix, that is, $\lambda_k = 1$ for $k = 1,2,\ldots, N$, the model (2.4.12) becomes a linear homoscedastic model.

The estimate (2.3.9) becomes

$$\hat{\bar{Y}} = \frac{1}{N}\left[n\bar{y} + (N-n)\ \underset{\sim}{m}_s^{*t}\ \hat{\underset{\sim}{\beta}}^*\right] \qquad \ldots\ldots\ldots \quad (2.4.13)$$

where $\underset{\sim}{m}_s^{*t} = (m^*_{s1}, \ldots, m^*_{sp})$

with $m^*_{\bar{s}_j} = \dfrac{N\bar{X}_{j.} - n\bar{x}_{j.}}{N-n}$

and $\hat{\underset{\sim}{\beta}}^* = (Z_s^{*t} Z_s^*)^{-1} Z_s^{*t} \underset{\sim}{Y}_s$

while (2.3.10) becomes

$$MSE(\hat{\bar{Y}}) = \frac{\sigma^2}{N^2}\left[(N-n)^2 \underset{\sim}{m}^{*t}_{\bar{s}} (Z_s^{*t} Z_s^*)^{-1} \underset{\sim}{m}^*_{\bar{s}} + (N-n)\right] \quad \ldots\ldots\ldots \quad (2.4.14)$$

where $\hat{\sigma}^2 = \dfrac{1}{n-p}\left[\underset{\sim}{Y}_s^t \underset{\sim}{Y}_s - \hat{\underset{\sim}{\beta}}^{*t} Z_s^{*t} \underset{\sim}{Y}_s\right]$

Case VI: Suppose p= 1 in Case IV, and $\lambda_k = g(x_k) = x_k$, $k = 1,2,\ldots,N$. Then

$$\eta^{*-1} = \begin{pmatrix} x_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & x_N \end{pmatrix}$$

and the estimate (2.3.8) becomes

$$\hat{\bar{Y}} = \frac{1}{N}\left[n\bar{y} + \sum_{k=1}^{N-n} x_k \cdot \hat{\beta}_1\right]$$

$$= \frac{1}{N}\left[n\bar{y} + \hat{\beta}_1(X - n\bar{x})\right] \quad \ldots\ldots\ldots \quad (2.4.15)$$

where $\hat{\beta}_1 = (Z_s^{*t} \eta_s^* Z_s^*)^{-1} Z_s^{*t} \eta_s^* \underset{\sim}{Y}_s$

$$= \frac{\bar{y}}{\bar{x}} \quad \ldots\ldots\ldots \quad (2.4.16)$$

Using (2.4.16) and (2.4.15), we get

$$\hat{\bar{Y}} = \frac{\bar{y}}{\bar{x}} \bar{X} \quad \ldots\ldots\ldots \quad (2.4.17)$$

which is the classical ratio estimator of the population mean.