# ON SOME PROBLEMS OF OPTIMIZATION IN SAMPLE SURVEYS

ABSTRACT

OF THE

THESIS

SUBMITTED FOR THE AWARD OF THE DEGREE OF

Doctor of Philosophy

IN

STATISTICS

BY

**NAJMUSSEHAR**

UNDER THE SUPERVISION OF

**DR. MOHAMMAD JAMEEL AHSAN**

DEPARTMENT OF STATISTICS & OPERATIONS RESEARCH
ALIGARH MUSLIM UNIVERSITY
ALIGARH (INDIA)

2002

# ABSTRACT

This thesis entitled "On some problems of optimization in sample surveys" is submitted to the Aligarh Muslim University, Aligarh, INDIA, to supplicate the degree of Doctor of Philosophy in Statistics. It embodies of research work carried out by me in the Department of Statistics and Operation Research, Aligarh Muslim University, Aligarh.

In the development of theory underlying statistical methods, one is frequently faced with optimization problems. Attempts have therefore been made to find optimization techniques that have wider applicability and can easily be implemented with the available computing power. One such technique that has the potential for increasing the scope of application of statistical methodology is mathematical programming. In this thesis an attempt has been made to formulate and solve some problems arising in sample surveys using classical optimization techniques such as Lagrange multipliers technique as well as using mathematical programming techniques.

This thesis consists of five chapters. Chapter-I provides an introduction to sample surveys with some basic results in Simple Random Sampling, Stratified Sampling, Non-Response and Double Sampling.

Chapter-II deals with the problem of allocation of a sample to strata in multivariate stratified sampling. In this chapter two new compromise allocations are proposed and compared with the already available

compromise allocations in sampling literature As assumed by Cochran (1977), here no assumption about the correlation between the different characteristics is made It has been shown through numerical illustrations that the proposed allocations are more precise than the already existing compromise allocations in sampling literature This chapter is based on my paper entitled "Allocation of a sample to strata The multivariate case" to be presented in the "National Seminar on Recent Development in Statistical Methods and Operation Research" organized by Department of Statistics, Dibrugarh University, Assam (India), during March 20-21,2003

In Stratified sampling the sampler has to decide about the sample sizes from various strata before drawing a sample In sampling literature this problem is known as the problem of allocation The equal, proportional and optimum allocations are well known allocations In practice any one type of allocation is selected according to the prevailing situation in the population and is applied to all strata However, there are practical situations in which the prevailing circumstances markedly differ from one group of strata to other Hence the use of the same allocation in all the strata may not be advisable In such situations it is proposed to divide the group of strata into non-overlapping and exhaustive subgroups according to some reasonable criterion The use of particular allocation is advised in a particular subgroup depending upon the characteristic of the subgroup Since different allocations are to be used in different subgroups, the proposed allocation is named as a "Mixed allocation" Chapter-III of this thesis discusses the "Mixed allocation in Stratified Sampling". It is assumed that the population mean is of interest. The problems of finding the mixed allocation for fixed cost and for fixed variance of the estimator of the population mean, based on a stratified sample are formulated and solved as nonlinear programming problems. The variance of the estimator under mixed allocation is worked out and compared with the variance under the overall optimum allocation. The relative increase in the variance due to the use of the mixed allocation is studied to decide whether a mixed allocation is advisable or not in a given situation. This chapter is based on my research paper entitled " Mixed Allocation in Stratified Sampling " to be presented in the International Conference on Statistics, Combinatorics and Related Areas Organized by Department of Statistics, University of Allahabad (India) going to be held during December 21-23, 2002.

In Chapter-IV the problem of optimum allocation in Double Sampling for stratification (DSS) with subsampling the non-respondents is formulated as a mathematical programming problem (MPP). When strata weights are not known, double sampling technique may be used to estimate them. A large simple random sample from the unstratified population is drawn and the units belonging to each stratum (in the sample) is obtained. A second stratified sample is then obtained from which a simple random subsample is drawn out of the previously selected units of the stratum. If the problem of non-response is there, then the subsamples are divided into respondents and non-respondents respectively. A second sub-sample of non-respondents units is selected out of non-respondents and information is obtained on second attempt.The objective of the problem is to find the optimum sizes of the subsamples of non-respondents. For this in the first phase of solution the optimum values of the sample sizes are obtained for which the variance of the estimated population mean for double sampling is minimum for a fixed sample size. In the second and the final phase of solution, the optimum values of subsamples of non-respondents are obtained for fixed total cost of the survey. A solution procedure using dynamic programming technique is developed to solve the resulting MPP. The computational details of the procedure are illustrated through a numerical example.

This chapter is based on my research paper entitled "Double sampling for stratification for subsampling the non-respondents" published in Aligarh Journal of Statistics (see Najmussehar and Abdul Bari (2002)).

Chapter-V deals with the problem of optimum stratification. For stratified sampling to be efficient the strata should be as homogeneous as possible with respect to the main study variable. In other words the stratum boundaries are so chosen that the stratum variances are as small as possible. This could be done effectively when the frequency distribution of the main study variable is known. Usually this frequency distribution is unknown but it is possible to approximate it from the past experience and prior knowledge about the population. In this chapter the problem of optimum stratification and formulated as a Nonlinear Programming Problem (NLPP) assuming exponential frequency distribution of the main study variable. The formulated NLPP is separable with respect to the decision variables and is treated as a multistage decision problem. A procedure is developed using dynamic programming technique to work out the optimum stratum boundaries. These stratum boundaries are optimum in the sense that they minimize the sampling variance of the stratified sample mean under Neyman allocation. A computer program in 'Java SDK 2'is also developed for the procedure. This computer program is executed to work out the optimum strata boundaries for a given exponential distribution to provide a numerical example.

This chapter is based on my research paper entitled "The problem of optimum stratification for exponential study variable under Neyman allocation: A Mathematical Programming Approach" to be presented in the International Symposium on Optimization and Statistics to be held in the Department of Statistics and Operations Research, Aligarh Muslim University , Aligarh (India), during December 28-30, 2002.

A comprehensive list of references, arranged in alphabetical order is also provided at the end of the thesis.

# ON SOME PROBLEMS OF OPTIMIZATION IN SAMPLE SURVEYS

THESIS

SUBMITTED FOR THE AWARD OF THE DEGREE OF

**Doctor of Philosophy**

IN

**STATISTICS**

BY

**NAJMUSSEHAR**

UNDER THE SUPERVISION OF

**DR. MOHAMMAD JAMEEL AHSAN**



DEPARTMENT OF STATISTICS & OPERATIONS RESEARCH
ALIGARH MUSLIM UNIVERSITY
ALIGARH (INDIA)

2002

*Dedicated*

*To My*

*Parents*

DEPARTMENT OF
STATISTICS AND OPERATIONS RESEARCH
THE ALIGARH MUSLIM UNIVERSITY
ALIGARH – 202002 (INDIA)

Mohammad Jameel Ahsan
M.Sc., M. Phil , Ph D (Alig )
READER

Phone: (0571) 2701251 (O)
(0571) 2404221 (R)
email: mjahsan2001@yahoo.co.in

Dated: Dec. 20, 2002

# CERTIFICATE

I certify that the material contained in this thesis entitled "ON SOME PROBLEMS OF OPTIMIZATION IN SAMPLE SURVEYS", submitted by Miss Najmussehar for the award of the Degree of Doctor of Philosophy in Statistics is original.

This work has been done under my supervision. In my opinion the work is sufficient for consideration for the award of the Ph. D. degree in Statistics to the candidate.

(DR. MOHAMMAD JAMEEL AHSAN)
SUPERVISOR

# PREFACE

# PREFACE

This thesis entitled "On some problems of optimization in sample surveys" is submitted to the Aligarh Muslim University, Aligarh, INDIA, to supplicate the degree of Doctor of Philosophy in Statistics It embodies of research work carried out by me in the Department of Statistics and Operation Research, Aligarh Muslim University, Aligarh

In the development of theory underlying statistical methods, one is frequently faced with optimization problems Attempts have therefore been made to find optimization techniques that have wider applicability and can easily be implemented with the available computing power One such technique that has the potential for increasing the scope of application of statistical methodology is mathematical programming In this thesis an attempt has been made to formulate and solve some problems arising in sample surveys using classical optimization techniques such as Lagrange multipliers technique as well as using mathematical programming techniques

This thesis consists of five chapters Chapter-I provides an introduction to sample surveys with some basic results in Simple Random Sampling, Stratified Sampling, Non-Response and Double Sampling

Chapter-II deals with the problem of allocation of a sample to strata in multivariate stratified sampling In this chapter two new compromise allocations are proposed and compared with the already available compromise allocations in sampling literature As assumed by Cochran (1977), here no assumption about the correlation between the different characteristics is made It has been shown through numerical illustrations that the proposed allocations are more precise than the already existing compromise allocations in sampling literature This chapter is based on my paper entitled "Allocation of a sample to strata The multivariate case" to be presented in the "National Seminar on Recent Development in Statistical Methods and Operation Research" organized by Department of Statistics, Dibrugarh University, Assam (India), during March 20-21,2003

In Stratified sampling the sampler has to decide about the sample sizes from various strata before drawing a sample In sampling literature this problem is known as the problem of allocation The equal, proportional and optimum allocations are well known allocations In practice any one type of allocation is selected according to the prevailing situation in the population and is applied to all strata However, there are practical situations in which the prevailing circumstances markedly differ from one group of strata to other Hence the use of the same allocation in all the strata may not be advisable In such situations it is proposed to divide the group of strata into non-overlapping and exhaustive subgroups according to some reasonable criterion The use of particular allocation is advised in a particular subgroup depending upon the characteristic of the subgroup Since different allocations are to be used in different subgroups, the proposed allocation is named as a "Mixed allocation" Chapter-III of this thesis discusses the "Mixed allocation in Stratified Sampling" It is assumed that the population mean is of interest The problems of finding the mixed allocation for fixed cost and for fixed variance of the estimator of the population mean, based on a stratified sample are formulated and solved as nonlinear programming problems The variance of the estimator under mixed allocation is worked out and compared with the variance under the overall optimum allocation The relative increase in the variance due to the use of the mixed allocation is studied to decide whether a mixed allocation is advisable or not in a given situation This chapter is based on my research paper entitled " Mixed Allocation in Stratified Sampling " to be presented in the International Conference on Statistics, Combinatorics and Related Areas Organized by Department of Statistics, University of Allahabad (India) going to be held during December 21-23, 2002

In Chapter-IV the problem of optimum allocation in Double Sampling for stratification (DSS) with subsampling the non-respondents is formulated as a mathematical programming problem (MPP) When strata weights are not known, double sampling technique may be used to estimate them A large simple random sample from the unstratified population is drawn and the units belonging to each stratum (in the sample) is obtained A second stratified sample is then obtained from which a simple random subsample is drawn out of the previously selected units of the stratum If the problem of non-response is there, then the subsamples are divided into respondents and non-respondents respectively A second sub-sample of non-respondents units is selected out of non-respondents and information is obtained on second attempt The objective of the problem is to find the optimum sizes of the subsamples of non-respondents For this in the first phase of solution the optimum values of the sample sizes are obtained for which the variance of the estimated population mean for double sampling is minimum for a fixed sample size In the second and the final phase of solution, the optimum values of subsamples of non-respondents are obtained for fixed total cost of the survey A solution procedure using dynamic programming technique is developed to solve the resulting MPP The computational details of the procedure are illustrated through a numerical example

This chapter is based on my research paper entitled "Double sampling for stratification for subsampling the non-respondents" published in Aligarh Journal of Statistics (see Najmussehar and Abdul Bari (2002))

Chapter-V deals with the problem of optimum stratification For stratified sampling to be efficient the strata should be as homogeneous as possible with respect to the main study variable In other words the stratum boundaries are so chosen that the stratum variances are as small as possible This could be done effectively when the frequency distribution of the main study variable is known Usually this frequency distribution is unknown but it is possible to approximate it from the past experience and prior knowledge about the population In this chapter the problem of optimum stratification and formulated as a Nonlinear Programming Problem (NLPP) assuming exponential frequency distribution of the main study variable The formulated NLPP is separable with respect to the decision variables and is treated as a multistage decision problem A procedure is developed using dynamic programming technique to work out the optimum stratum boundaries These stratum boundaries are optimum in the sense that they minimize the sampling variance of the stratified sample mean under Neyman allocation A computer program in 'Java SDK 2' is also developed for the procedure This computer program is executed to work out the optimum strata boundaries for a given exponential distribution to provide a numerical example

This chapter is based on my research paper entitled "The problem of optimum stratification for exponential study variable under Neyman allocation A Mathematical Programming Approach" to be presented in the International Symposium on Optimization and Statistics to be held in the Department of Statistics and Operations Research, Aligarh Muslim University , Aligarh (India), during December 28-30, 2002

A comprehensive list of references, arranged in alphabetical order is also provided at the end of the thesis

# ACKNOWLEDGEMENT

# ACKNOWLEDGEMENT

I feel great pleasure at the completion of this research work and would like to show deep indebtness and sincere gratitude to all who helped me in the completion of this work

I would like to express my indebtness and sincere gratitude to my supervisor **Dr. Mohammad Jameel Ahsan,** Reader, Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh, for his invaluable guidance, and constant encouragement throughout the course of this research work I must appreciate him for his patience, great involvement and sympathetic behavior, which enabled me to complete this work within the stipulated time

I am thankful to Prof Sirajur Rehman, Chairman, Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh for providing me necessary facilities in the department and for extending his kind cooperation at every stage of the pursuit of this research work I am also thankful to Prof S U Khan, Prof M Z Khan, Prof A H Khan, Dr S N Alam Dr S M Arshad and Dr Rafiqullah Khan for their encouragement and best wishes

My sincere thanks are due to Dr Abdul Bari for his valuable help, support, words of encouragement and well wishes during the pursuit of this research work in general and specially for the matter presented in the Chapter-IV of this thesis

I am also thankful to Dr. Irshad Ahmad Khan for his valuable help and suggestions.

With a deep sense of gratitude I would also like to thank Dr. Mohammad Golam Mostafa Khan, of the Department of Mathematics and Computing Science, The University of the South Pacific, Suva, Fiji, for his kind cooperation and help through emails that enabled me to present this research work in time. My sincere thanks are also due to Mr. Hasan Chawdhury for his great help in writing the computer programs presented in this manuscript.

I am also thankful to Dr. Haseeb Athar, Mr. Vaseem Ismail, Mr. Abdul Razaak , Ms. Ruhi, Mohd. Firoz, Dr. Nadeem Siddiqui, Mr. Irfan Siddiqui, Ms. Rana and Ms. Beenish for their best wishes.

My sincere thanks are due to all teaching and non-teaching staff members, for their kind help and cooperation.

Finally a very special acknowledgement is for my parents for their constant support, encouragement and patience at all the times.

Najmussehar.

Dated: 20.12.02

**(NAJMUSSEHAR)**
Department of Statistics & Operations Research
Aligarh Muslim University
Aligarh 202002, India.

# CONTENTS

**CHAPTER-2**

**SOME NEW COMPROMISE ALLOCATIONS IN MULTIVARIATE STRATIFIED SAMPLING DESIGNS**

## CHAPTER-III

## MIXED ALLOCATION IN STRATIFIED SAMPLING



## CHAPTER-IV

## DOUBLE SAMPLING FOR STRATIFICATION WITH SUBSAMPLING THE NONRESPONDENTS: A DYNAMIC PROGRAMMING APPROACH

## CHAPTER-V

### THE PROBLEM OF OPTIMUM STRATIFICATION UNDER NEYMAN ALLOCATION: A MATHEMATICAL PROGRAMMING APPROACH

# CHAPTER-I

## INTRODUCTION

# CHAPTER-I

# INTRODUCTION

## 1A: SAMPLING

### 1A.1 INTRODUCTION TO SAMPLING

Sampling is used in all kinds of investigations in our every day life. It is the selection and study of a part of an aggregate material to represent the whole. Deming (1950) describes sampling as "The science and art of controlling and measuring reliability of useful statistical information through the theory of probability".

A sampling method is a scientific and objective procedure of selecting units from a population and provides a sample that is representative of the population.

The sampling procedures discussed in this thesis are procedures of random sampling or probability sampling. All random sampling procedures satisfy the following properties.

(i) A set of distinct samples are available which the procedure is capable of selecting if applied to a specific population.

(ii) Each possible sample has assigned to it known probability of selection.

(iii) A selection procedure is available in which each sample receives its assigned probability of selection.

(iv) The method of constructing the estimate from the sample must lead to a unique value for a specified sample. (See Cochran (1977)).

Other sampling procedures that do not possess the above properties are called non-random on non-probability sampling procedures are out of the scope of this thesis.

## 1A.2 USES OF SAMPLING

Sample surveys are widely in use in all most all walks of life in a variety of ways all over the world. The objective of a sample survey may be to obtain some measure with respect to the characteristic of the whole population under study. For example for national planning and socio-economic development the governments need information about agricultural production, utilization of land and water resources, industrial production, unemployment, labor force, whole sale and retail prices of various commodities, income and expenditure per household, number of literate persons and school going children, health status of people etc, which can be obtained efficiently through sample surveys.

Sampling methods are also used in census or complete enumerations. In fact, except for certain basic information, all other data in a census are collected on sampling basis, which results in much earlier publication of the census report and substantial savings in terms of money and time.

Sampling methods are used extensively in business and industry to increase operational efficiency. Market research is heavily dependent on sample surveys. Manufacturers and retailers can have an idea of the reactions of people to new products, their complaint about old products and the reasons for preferring one product to another, through sample surveys.

Sampling methods are also used in experimental investigations. For example in determining the quality of milk, response of fertilizers to various crops, the composition of the soil etc.

## 1A.3 SAMPLING DESIGNS

Various random sampling procedures that can be applied to the population under study according to the aims and objectives of the survey and the nature and variation in the population are also termed as sampling designs. The commonly used sampling designs are:

(i) Simple Random Sampling (SRS)

(ii) Stratified Sampling

(iii) Cluster Sampling

(iv) Systematic Sampling

(v) Two-Stage Sampling

(vi) Sub-Sampling or Multistage Sampling etc.

A sampling procedure or design may be carried out with replacement or without replacement. In with replacement (WR) sampling the selected unit is replaced before the next draw, whereas in without replacement (WOR) sampling the unit once selected is not considered for further draws. In a WR sample a population unit may appear more than once, while in a WOR sample all the selected units are distinct. Obviously a WOR sample contains more information about the population as compared to a WR sample. In this thesis the discussions are limited to WOR sampling.

Some times the population characteristic under study is strongly correlated to another characteristic called the auxiliary characteristic. The data on this auxiliary characteristic is either available or can be easily collected for all the units in the population. This auxiliary information (data on auxiliary characteristic) may be used to improve the quality (precision) of the estimates of the population parameters obtained from a sample. Some methods that uses the auxiliary information are:

(i) Ratio Method

(ii) Regression Method

(iii) Double Sampling or Two-Phase Sampling Method

Out of the sampling designs pointed out in this section the first two namely Simple Random Sampling and Stratified Sampling are the most commonly used sampling designs. In the following two sections the basic results of these sampling design are stated.

## 1A.4 SIMPLE RANDOM SAMPLING WITHOUT REPLACEMENT (SRSWOR)

It is the simplest method of random sampling. It is a method of selecting $n$ units out of N such that every possible distinct sample of size $n$ has an equal chance of being selected.

Let $y_i$ be the value of the characteristic under study for the ith unit of the population/ sample.

Further, let

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i \qquad \text{the population mean}$$

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \overline{Y}\right)^2 \quad \text{the population mean square}$$

$$\overline{y} = \frac{1}{n}\sum_{i-1}^{n} y_i \quad \text{the sample mean}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 \quad \text{the sample mean square}$$

The following are the basic results in SRSWOR.

(1) The sample mean $\overline{y}$ is an unbiased estimate of the population mean $\overline{Y}$.

(2) The sample mean square $s^2$ is an unbiased estimate of the population mean square $S^2$.

(3) The sampling variance of the sample mean $\overline{y}$ is

$$V(\overline{y}) = \left(\frac{1}{n} - \frac{1}{N}\right)S^2.$$

(4) $v(\overline{y}) = \left(\frac{1}{n} - \frac{1}{N}\right)s^2$ is an unbiased estimate of $V(\overline{y})$.

## 1A.5 STRATIFIED SAMPLING

It is the most popular and widely used sampling design. In this sampling procedure the population is divided in non-overlapping and exhaustive groups of units. These groups are called strata. Independent WOR simple random samples are then drawn from each stratum. Let there be L strata. The following symbols refer to the stratum h (h=1,2,...,L):

| | |
|---|---|
| $N_h$ | number of units in the stratum (the stratum size) |
| $n_h$ | number of units in the sample (the sample size) |
| $y_{hi}$ | value obtained from the ith unit |
| $W_h = \frac{N_h}{N}$ | stratum weight |
| $\bar{Y} = \frac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}$ | stratum mean |
| $S_h^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(y_{hi} - \bar{Y}_h)^2$ | stratum variance |

Also let

$$\bar{Y} = \frac{1}{N}\sum_{h=i}^{L}\sum_{i=1}^{N_h} y_{hi}$$

$$= \sum_{h=1}^{L} W_h \bar{Y}_h \qquad \text{the over all population mean}$$

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \qquad \text{the stratified sample mean}$$

The following are the basic results.

(1) $\bar{y}_h$ and $s_h^2$ are the unbiased estimates of $\bar{Y}_h$ and $S_h^2$ respectively.

(2) The sampling variance of $\bar{y}_h$ *is*

$$V(\bar{y}_h) = (\frac{1}{n_h} - \frac{1}{N_h})S_h^2.$$

(3) An unbiased estimate of $V(\bar{y}_h)$ is

$$v(\bar{y}_h) = (\frac{1}{n_h} - \frac{1}{N_h})s_h^2.$$

(4) $\bar{y}_{st}$ is an unbiased estimate of $\bar{Y}$.

(5) The sampling variance of $\bar{y}_{st}$ is

$$V(\bar{y}_{st}) = \sum_{h=1}^{L}(\frac{1}{n_h} - \frac{1}{N_h})W_h^2 S_h^2.$$

(6) An unbiased estimate of $V(\bar{y}_{st})$ is

$$v(\bar{y}_{st}) = \sum_{h=1}^{L}(\frac{1}{n_h} - \frac{1}{N_h})W_h^2 s_h^2.$$

The use of stratified sampling design involves the following four decision-making design operations.

(1) The choice of the stratification variable.

(2) The choice of the number L of strata.

(3) The choice of the stratum boundaries.

(4) The choice of the size $n_h$ of the sample from the hth stratum.

The discussion of all the operations is beyond the scope of the thesis. However, the last two operations are discussed in detail in the subsequent chapters of this thesis.

A chapter of this thesis is devoted to the use of double sampling to deal with the problem of non-response. In the next two sections these topics are introduce in brief.

## 1A.6 DOUBLE SAMPLING

As given in section 1.3, a number of sampling techniques use the auxiliary information. Ratio and regression methods require knowledge of the population mean $\bar{X}$ of the auxiliary variable $x$. If the auxiliary variable $x$ is to be used as the stratification variable its frequency distribution must be known. When such information is not available the technique of double sampling (or two-phase sampling) may be used to obtain the auxiliary information. In double sampling a large preliminary sample is taken in which the auxiliary variable $x$ alone is measured and a reasonably good estimate of the auxiliary information required is obtained.

The double sampling may be appropriate when the cost of measuring the auxiliary variable is significantly low as compared to the main variable.

### 1A.7 ERRORS IN SURVEYS

In sample surveys there will always be a difference between the population value and its estimate. This error is due to the sampling itself, that is due to the none-enumeration of the entire population and thus is called sampling error.

Other errors in the surveys arising in the collection, processing, compiling and analysis of the data are called non-sampling errors.

Non-sampling errors can further be classified into response error and non-response error. Errors of measurement on a unit due to the use of faulty or biased measuring device and errors introduced in editing coding and tabulating the results are called response errors. Whereas the error due to the failure in measuring some units selected in the sample, which results in an incomplete sample data, are called non-response error.

### 1A.8 NON-RESPONSE

The non-response refers to the failure to measure some of the units selected in the sample. In surveys it is commonly experienced that complete data from the sampling units is often not obtainable for various reasons. For example, in an opinion survey, the selected family might have shifted to some other place, selected person might have died. In mailed questionnaire, some of the selected addresses may be wrong or they do not reply. Such a problem of incomplete sample data is known as the problem of non-response in sampling literature.

One way to deal with the problem of non-response is to assume that the population consists of two strata, one of the respondents, on which the information is available and the other of the non-respondents, on which the information is not available at first attempt. A sub sample is drawn out of the sampled units falling in non-respondents stratum and a second and extensive attempt is made to obtain information on these units. The information obtained on the two attempts is then pooled to construct the required estimate.

## 1B: OPTIMIZATION

### 1B.1 INTRODUCTION

Optimization is the act of obtaining the best result under given circumstances. The efforts required or the benefits desired in any practical situation can often be expressed as a function of some decision variables. The ultimate goal of such decision is either to maximize the benefit desired or to minimize the loss or cost incurred or efforts required. Mathematically, optimization is the maximization or minimization of a function of several variables. These variables may be unconstrained or subjected to certain constraints in the form of equations or/and inequalities. There is no single method available for solving all optimization problems. A number of optimization methods are developed for solving different types of optimization problems. The constrained optimization techniques are also known as mathematical programming methods or techniques.

## 1B.2 A BRIEF HISTORICAL SKETCH

The existence of optimization methods can be traced back to the days of Newton, Lagrange and Cauchy. But in spite of these early contributions very little progress has been made until the middle of the nineteenth century, when the high-speed digital computers made the implementation of the optimization procedures possible and stimulated further research on new methods.

Constrained optimization or mathematical programming has developed rapidly during and after World War II as a new field of study dealing with applications of the scientific method of business operations and management decision-making. Mathematical programming problems can be broadly classified as (i) Linear Programming Problems (LPP) when all the involved functions are linear and (ii) Nonlinear Programming Problems (NLPP), when all the involved functions are not linear.

In 1947 the United States Air Force team SCOOP (Scientific Computation of Optimum Programs) started intensive research on some optimum resource allocation problem that led to the development of the famous simplex method by George B. Dantzig for solving a linear programming problem (LPP). The simplex method is an iterative procedure, which yields an exact optimal solution in a finite number of steps. But the method was not available until it was published in the Cowles Commission Monograph No. 13 in 1951.

One of the earliest enterprises undertaken by the exponents of mathematical programming grew out of the problems involved in the war mobilization program of the 1940's. The problems of planning and co-coordinating among various project and optimum allocation of limited resources to obtain the desired result were emerged as the basic problems.

Kuhn, H.W. and Tucker, A.W. (1951) derived the necessary conditions for the optimal solution of a constrained optimization or mathematical programming problem. These conditions (popularly known as K-T conditions) laid the foundation of a great deal of later research and development in the area of non-linear programming.

No single technique (like simplex method for solving LPP) is available till date for solving NLPP. However different methods are available for solving some special types of NLPP. Beale (1959) developed a method for solving convex quadratic programming problem (CQPP). Wolfe (1959), using the K-T conditions, transformed the CQPP into equivalent LPP with an additional non-linear restriction to which simplex method could be applied. Other authors who gave the technique for solving QPP are Markowitz (1956), Hilderth (1957), Houthaker (1960), Lemke (1962), Van de Panne and Whisnton (1964a, 1964b, 1966), Graves (1967), Fletcher (1971), Aggarwal (1974a, 1974b), Finkbeiner and Kall (1978), Arshad, Khan and Ahsan (1981). Ahsan, Khan and Arshad (1983), Todd (1985), Fukushima (1986), Yuan (1991), Wei (1992), Benzi (1993), Anstreicher, Den Hertog and Terlaky (1994) and Several others.

Rosen (1960, 1961), Kelly (1960), Goldfarb (1969), Du, Wu and Zhag (1990), Lai, Gao, and He (1993) developed Gradient projection methods for solving NLPP with linear and nonlinear constants. This is an iterative procedure in which at each step we move from one feasible solution to another in such a way that the value of the objective function is improved.

A linear fractional programming technique was proposed by Charnes and Cooper (1962). The algorithms for solving non-linear fractional programming were developed by Dinkelbach (1967) and Mangasrian (1969).

Geometric programming provides a systematic method for formulating and solving the class of optimization problems that tend to appear mainly in engineering designs. This technique was first developed by Duffin, Peterson and Zener (1967). Ermer (1971) used geometric programming for optimization of the constrained machinery economic problem. His work was further extended by Dembo (1982), Kortanek and Hoon (1992), Yeh (1993) and several others.

Dantzig (1959), Charnes and Cooper (1959, 1960) developed stochastic programming techniques. Some other authors who worked on stochastic programming are Shapiro (1990), Weintraub and Vera (1991), Flam and Schult (1993), Schoen (1994) and Bahn et al. (1995) etc.

A technique known as goal programming for solving multi-objective linear and non-linear programming problems was developed by Charnes and Cooper (1977). Other authors who made contribution for solving multiobjective linear and non-linear programming problems are Sherali (1982), Roy and Wallenius (1992), Arbel (1993, 1994), Bit, Biswal and Alam (1993) and Okada (1993) etc.

Dynamic programming technique, based on the principle of optimality, was developed by Richard Bellman (1957). This technique is applicable to mathematical programming problems having some special features. Several others who contributed significantly to this area are Bellman and Dreyfus (1962), Wachs (1989), Li (1990), Li and Haimes (1990), Wang (1990a, 1990b) Wang and Xing (1990), Lin (1994), Badinelli (2000) etc.

Developments of new techniques for solving mathematical programming are still going on. To cover all of them is beyond the scope of this thesis.

In this thesis dynamic programming technique is used for solving some of the optimization problems arising in sampling. The following section gives a brief account of the dynamic programming technique.

## 1B.3 DYNAMIC PROGRAMMING TECHNIQUE

The problems requiring sequential decision-making at different stages may be called multistage decision problems. The problem of making a set of optimal decisions may be formulated as an MPP. The dynamic programming technique is a procedure, which can handle the problem of optimal decision-making at various stages of a multistage decision problem. The general nature of the MPP that can be handled by this technique may be described as follows.

(i) The MPP can be treated as a multistage decision problem. At each stage the value(s) of one or more decision variables are to be determined.

(ii) The MPP must have the same structure at every stage irrespective of the number of stages.

(iii) At every stage the values of the decision variables and the objective function must depend on a specified set of parameters describing the state of system. These parameters are called the state parameters.

(iv) Same set of state parameters must describe the state of the system irrespective of the number of stages.

(v) The decision at any stage must have no effect on the decisions to be made at the remaining stages except in changing the values of the state parameters.

In solving an MPP by dynamic programming technique we start with a one-stage problem, moving on to a two-stage problem, to a three-stage problem and so on until all stages are included. The final solution is obtained by adding the nth (final) stage to the solution of (n-1) stage. For this a relation between the two successive stages is defined. This relation is called the "Recurrence Relation" of dynamic programming.

The computational efficiency of the dynamic programming technique as compared to the complete enumeration is very impressive. But unfortunately the computational efforts involved in solving an MPP by dynamic programming technique multiply incredibly fast with the increase in the number of state parameters (number of constraints). The number of state parameters is called the dimensionality of the MPP. The problem of handling the great bulk of computation in dynamic programming technique is termed as the "Problem of Dimensionality" or the "Curse of Dimensionality" to dynamic programming.

Bellman and Dreyfus (1962) suggested a procedure to reduce the dimensionality of the problem.

However, as far as the problems discussed in this thesis are concerned dimensionality poses no threat to the convergence of computational procedures developed using dynamic programming technique.

## 1B.4 APPLICATIONS OF OPTIMIZATION TECHNIQUES

During the last five decades attempts have been made to develop suitable and efficient optimization techniques that can be easily implemented with the available computing power to solve various optimization problems. The early applications of optimization techniques were limited to problems involving military operations. Later on they are widely used in dealing with the optimization problems in almost every walk of life. In recent past the optimization or the mathematical programming techniques (as they are popularly known) are successfully used in solving a variety of constrained optimization problems arising in Planning, Business, Industry, Economics, Commerce, Biological and Medical Services, Agriculture, Environmental Protection, Artificial Intelligence, Space Research, Engineering, Information Technology, Statistics etc etc.

## 1B.5 OPTIMIZATION TECHNIQUES IN STATISTICS

According to C.R. Rao (See Arthanari and Dodge (1981)) all statistical procedures are, in the ultimate analysis, solutions to suitably formulated optimization problems. Whether it is designing a scientific experiment or planning a large scale survey for collection of data, or choosing a stochastic model to characterize observed data, or drawing inference from available data, such as estimation, testing of hypothesis and decision making, one has to choose an objective function and minimize or maximize it subject to given constraints on unknown parameters and inputs such as the cost involved. The classical optimization methods based on differential calculus are too restrictive and are either inapplicable or difficult to apply in many situations that arise in statistical work. This together with the lack of suitable numerical algorithms for solving optimizing equations has placed several limitations on the choice of objective functions and constraints and led to the development and use of some inefficient statistical procedures.

Attempts have therefore been made during the last five decades to find other optimization techniques that have wider applicability and can be easily implemented with the available computing power. One such technique that has the potential for increasing the scope for application of efficient statistical methodology is mathematical programming. Although endowed with a vast literature, this method has not come into regular use in statistical practice mainly because of lack of good expositions integrating the techniques of mathematical programming with statistical concepts and procedures.

A few successful applications of optimization or mathematical programming techniques to the problems arising in statistical analysis are given below.

Jesen (1969), Rao (1971), Buhler et al (1975), Littschwager and Wang (1978) in cluster analysis.

Foody and Heydayat (1977), Arthanari and Dodge (1978), Whitaker, Thriggs and John (1990) in construction of BIB designs.

Barankin (1951), Dantzig and Wald (1951) Francis and Wright (1969), Kraft (1970), Meeks and Francis (1973), Pukelshein (1978), Kabe (1989), Ozturk (1991) in testing of statistical hypothesis.

Neauhardt, Bradely and Henning (1973) in optimal design of multifactor experiments.

Chakraborthy (1986, 1988, 1990, 1991), Gosh (1989), Seidel (1991), Crowder (1992) in quality control.

Tillman, Hwang and Kuo (1977) in reliability theory etc etc.

## 1B.6 OPTIMIZATION TECHNIQUES IN SAMPLING

The basic need of present day society is the need of reliable data to understand better the world in which we live. Such data can only be collected through sample surveys. The fundamental problem in sample surveys is to choose a sampling design that either gives the maximum precision within available budget or minimizes the cost of survey for a prefix level of tolerance regarding the precision. Thus the base of sample survey methodology is an optimization problem. The cost of the sample survey and the precision of estimates are function of sample size. Thus the problem of deriving statistical information on population characteristics based on sample data can be formulated as an optimization problem. In stratified sampling the problem of determining the optimum number of strata, the problem of fixing optimum strum boundaries, the problem of obtaining optimum allocations to sample sizes from various strata are optimization problems that can be formulated and solved as mathematical programming problems.

In multivariate surveys where more than one characteristic are to be measured on each and every unit of the selected sample the problem of working out optimum sample size (or sizes in case of stratified sampling) can be formulated as a multi objective optimization problem.

When two or more sample surveys are conducted on the same population, the same population unit may be assigned different probabilities for different surveys. In such situations we may want to maximize the expected number of common units in the selected sample for different surveys for the given probabilities of selection. This is called integration of surveys. Thus the problem of optimum integration of surveys is also an optimization problem.

Some successful applications of optimization techniques in the problems arising in sample surveys are due to:

Stock and Frankel (1939), Ghosh (1958), Aoyama (1963), Kokan (1963), Folk and Anle (1965), Ericson (1967), Kokan and Khan (1967), Kish (1967), Chatterjee (1966, 1967, 1968, 1972), Murthy (1967), Raj (1969), Chaddha et al (1971), Ahsan (1975,1978), Ahsan and Khan (1977,1982),Cochran (1977), Omule (1985), Bethal (1989), McCallion (1992), Sheela and Unnithan (1992), Kreinbrock (1993), Rahim and Currie (1993), Jahan et.al. (1994), Mandowara (1994), Jahan and Ahsan (1995), Csenki (1997), Khan et al (1997, 2002a), Clark and Steel (2002), Bretthauer, Ross and Shetty (1999), etc in optimum allocation of sample sizes.

Dalenius and Gurneym (1951), Dalenius and Hodge (1959), Ghosh (1963), Sethi (1963), Hartley (1965), Herleker (1967), Serfling (1967), Buhler, Aachen and Mannhein (1975), Singh

(1977), Unnithan (1978), Jarque (1981), Khare (1987), Miles et al (1987), Miles and Robert (1989), Rahim and Jocelyn (1994), Chernayak and Starytskyy (1998), Chernayak and Chornous (2000), Jahan et al (2001), Khan et.al. (2002b) etc in optimum stratification.

Dalenius (1957), Cochran (1963), Serfling (1968), Khan et al (1995) etc in determining optimum number of strata.

Alldredge and Amstrong (1974) in estimation of overlap size created by interlocking sampling process.

Kefitz (1951), Lahiri (1954), Murthy (1967), Raj (1969), Arthanari and Dodge (1981), Mitra and Pathak (1984), Aragon and Pathak (1990), Fahim and Pathak (1992) etc in optimum integration of surveys.

# CHAPTER-II

## SOME NEW COMPROMISE ALLOCATIONS IN MULTIVARIATE STRATIFIED SAMPLING DESIGNS

# CHAPTER II

## SOME NEW COMPROMISE ALLOCATIONS IN MULTIVARIATE STRATIFIED SAMPLING DESIGNS

### 2.1 INTRODUCTION

In stratified random sampling the value of the sample sizes for various strata are to be chosen in advance. In sampling literature, the problem of selecting the sample sizes for various strata is termed as an allocation problem. The sample sizes may be chosen to minimize the variance of the estimate for a fixed total cost of the survey or to minimize the total cost of the survey for a given precision of the estimate. Equal, proportional and optimum allocations are well known in sampling literature.

When several characteristics (say 'p') are to be measured on each selected unit of the sample, the problem of optimum allocation becomes more complicated because there is no single optimality criterion through which we can attack the allocation problem. In such situations we need a suitable compromise criterion to workout a usable allocation which is optimum in some sense for all characteristics. This allocation may be called a compromise allocation because it is based on a compromise criterion.

In this chapter the already existing compromise allocations are discussed and two new compromise allocations are proposed that are more precise than their existing counter parts.

## 2.2 STRATIFIED SAMPLING

Let a population of size $N$ is be divided into $L$ strata. The following symbols refer to stratum $h$ $(h= 1,2,\ldots,L)$.

| | |
|---|---|
| $N_h$ | stratum size (number of units in the stratum) |
| $n_h$ | sample size (number of units selected in the sample) |
| $y_{hi}$ | value obtained for the ith units |
| $W_h = \dfrac{N_h}{N}$ | stratum weight |
| $\overline{Y}_h = \dfrac{\sum_{i=1}^{N_h} y_{hi}}{N_h}$ | stratum mean |
| $\overline{y}_h = \dfrac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$ | sample mean |
| $S_h^2 = \dfrac{\sum_{i=1}^{N_h}\left(y_{hi} - \overline{Y}_h\right)^2}{N_h - 1}$ | stratum variance |

If the estimation of the population mean per unit $\overline{Y} = \dfrac{\sum_{h=1}^{L}\sum_{i=1}^{N_h} y_{hi}}{N} = \sum_{h=1}^{L} W_h \overline{Y}_h$ is of interest then it is well known that the stratified sample mean $\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$ serves as an unbiased estimate of $\bar{Y}$ with a sampling variance:

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} - \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{N_h} \tag{2.1}$$

The total cost $C$ of the survey may be express as

$$C = c_o + \sum_{h=1}^{L} c_h n_h$$

or $$C_o = \sum_{h=1}^{L} c_h n_h \tag{2.2}$$

where $C_o = C - c_o$, $c_o$ represents the overhead cost and $c_h$ represents the per unit measurement cost for the hth stratum.

## 2.3 PROPORTIONAL ALLOCATION

The allocation in which $n_h$ are proportional to $N_h$ is called the proportional allocation and was originally proposed by Bowley (1926). Under proportional allocation

$n_h \propto N_h$

or $n_h = KN_h$

where $K$ is the constant of proportionality.

Substituting this value of $n_h$ in (2.2)

$$C_o = \sum_{h=1}^{L} K c_h N_h$$

or $K = \dfrac{C_o}{\sum_{h=1}^{L} c_h N_h}$

Thus $n_h = \dfrac{C_o N_h}{\sum_{h=1}^{L} c_h N_h} = \dfrac{C_o W_h}{\sum_{h=1}^{L} c_h W_h}; h = 1,2,\ldots,L$ (2.3)

If $c_h = c$ for all h then (2.3) gives

$$n_h = nW_h ; h = 1,2,\ldots,L \tag{2.4}$$

where $n = \dfrac{C_o}{c}$,is the total sample size.

Expression (2.4) gives the proportional allocation for fixed total sample size.

Under proportional allocation fixed cost the sampling variance of $\bar{y}_{st}$ is given by

$$V(\bar{y}_{st})_{prop} = \frac{(\sum_{h=1}^{L} W_h S_h^2)(\sum_{h=1}^{L} c_h W_h)}{C_o} - \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{N_h} \tag{2.5}$$

and for fixed total sample size

$$V(\bar{y}_{st})_{prop} = \frac{\sum_{h=1}^{L} W_h S_h^2}{n} - \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{N_h} \tag{2.6}$$

## 2.4 OPTIMUM ALLOCATION:

Staurt (1954) used Cauchy-Schwarz inequality to show that the

expression: $$n_h = n.\frac{W_h S_h / \sqrt{c_h}}{\sum_{h=1}^{L} W_h S_h \sqrt{c_h}};h = 1,2,...,L \qquad (2.7)$$

gives the values of $n_h$ that minimize (i) $V(\bar{y}_{st})$ when the total cost $C$ is fixed and (ii) $C$ when the variance $V(\bar{y}_{st})$ is fixed.

Allocation given in (2.7) is known as optimum allocation.

If the total cost is fixed then the total sample size $n$ is given by

$$n = \frac{(C - c_o)\sum_{h=1}^{L}\left(W_h S_h / \sqrt{c_h}\right)}{\sum_{h=1}^{L}\left(W_h S_h \sqrt{c_h}\right)} \qquad (2.8)$$

The expression (2.8) is obtained by substituting the values of $n_h$ from (2.7) in (2.2). On the other hand if the variance is fixed then $n$ is given by

$$n = \frac{N\left(\sum_{h=1}^{L} W_h S_h \sqrt{c_h}\right)\sum_{h=1}^{L}\left(W_h S_h / \sqrt{c_h}\right)}{NV + \sum_{h=1}^{L} W_h S_h^2} \qquad (2.9)$$

where $V$ is the fixed value of the variance $V(\bar{y}_{st})$.

The expression (2.9) is obtained by substituting the values of $n_h$ from (2.7) in (2.1) (See Cochran (1977)).

If $c_h = c$ for all h, that is the per unit measurement cost is same in each stratum then total cost $C$ given by (2.2) becomes $C = c_o + cn$. In this case the optimum allocation for fixed cost reduces to the optimum allocation for fixed sample size $n$ and we have the allocation problem as

"Minimize $V(\bar{y}_{st})$

subject to $\sum_{h=1}^{L} n_h = n$"

Neyman (1934) showed that $V(\bar{y}_{st})$ is minimum for fixed $n$ if $n_h$ are given by

$$n_h = n.\frac{W_h S_h}{\sum_{h=1}^{L} W_h S_h}; h = 1,2,...,L. \tag{2.10}$$

Therefore, $n_h$ given by (2.10) is sometimes called the Neyman allocation.

The variance under Neyman allocation is given as:

$$V(\bar{y}_{st})_{\min} = \frac{(\sum_{h=1}^{L} W_h S_h)^2}{n} - \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{N_h} \tag{2.11}$$

Usually the total cost of the survey is fixed in advance. Hence here in after by an allocation we mean the allocation that minimizes $V(\bar{y}_{st})$ for fixed total cost of the survey.

## 2.5 PROBLEM OF ALLOCATION: THE MULTIVARIATE CASE

When several characteristics (say 'p') are to be measured on each selected unit of the sample the problem of optimum allocation becomes more complicated. In such cases $S_h^2$ and $c_h$ may vary from stratum to stratum as well as from characteristic to characteristic and the optimum allocation given by (2.5) becomes

$$n_{hj} = n\frac{W_h S_{hj} / \sqrt{c_{hj}}}{\sum_{h=1}^{L} W_h S_{hj} / \sqrt{c_{hj}}}; h = 1,2,...,L; j = 1,2,...,p \,. \qquad (2.12)$$

where

$n_{hj}$=sample size for measuring jth characteristic; $j=1,2, \;\; ,p$ in hth stratum; $h=1,2,..,L$

$S_{hj}^2$ = stratum variance of the jth characteristic in the hth stratum.

$c_{hj}$ = per unit cost of measuring the jth characteristic in the hth stratum.

For different characteristics there are different sets of optimum allocations. In such cases $n_{hj}$ given by (2.10) can be arranged as an (L x p) matrix whose jth column represents the optimum allocation with respect to the jth characteristic. Hence there is no unique set of values of $n_h$ that minimizes all the variances $V(\bar{y}_{jst}), j = 1,2,...,p$ simultaneously, where

$$V(\bar{y}_{jst}) = \sum_{h=1}^{L} \frac{W_h^2 S_{hj}^2}{n_{hj}} - \sum_{h=1}^{L} \frac{W_h^2 S_{hj}^2}{N_h} \qquad (2.13)$$

In such situations we need a single representative for each row of the matrix $((n_{hj}))$.

There are two ways to deal with this situation. One way is to select a single representative for each row according to some reasonable criterion. And the other way is to reformulate and solve the problem of allocation in which the objective is to find $n_h$ that minimize, some function of $V(\bar{y}_{jst})$ for fixed total cost.

In the remaining sections of this chapter the problem of optimum allocation in multivariate stratified random sampling is studied in detail and allocations proposed by various authors are discussed. Two new allocations are proposed and compared with the already existing allocations in the sampling literature through numerical examples.

## 2.6 COMPROMISE ALLOCATION BASED ON THE ROW REPRESENTATIVES

Since the optimum allocation with respect to different charactertics are different there is no unique set of values of $n_h; h = 1,2,...,L$ that minimize every $V(\bar{y}_{jst}), j = 1,2,...,p$, simultaneously. Therefore, for practical purposes some compromise must be reached in a multivariable survey regarding the sample sizes from various strata.

An allocation based on some compromise criterion may be called a compromise allocation (See Cochran (1977)). If the correlations between the characteristics are sufficiently high the individual optimum allocations may vary relatively little. In such situations Cochran (1977) proposed the compromise allocation based on the averages of the individual optimum sample sizes with respect to different characteristics. If $n^*_{hj}; h = 1,2,...,L \; ; j = 1,2,...,p$ denote the individual optimum allocation for jth characteristic in the hth stratum then by formula (2.10).

$$n^*_{hj} = n.\frac{W_h S_{hj}}{\sum_{h=1}^{L} W_h S_{hj}}; h = 1,2,.....L; j = 1,2,....,p \qquad (2.14)$$

where the optimum allocation is for a fixed total sample size $n$. As suggested in Section 2.5 $n^*_{hj}$ given by (2.14) can be arranged as an (L x p) matrix whose jth column represents the optimum allocation with respect to the jth characteristic.

Let $n_{h(a)}$ denote the compromise allocation based on averages, as suggested by Cochran (1977) then

$$n_{h(a)} = \frac{1}{p}\sum_{j=1}^{p} n^*_{hj}; h = 1,2,...,L \qquad (2.15)$$

where the symbol $(a)$ stands for the average.

For the jth characteristic using (2.1) and ignoring finite population correction (fpc) the variances $V(\bar{y}_{jst})$ under this compromise allocation are given by

$$V_{j(a)} = V(\bar{y}_{jst}) = \sum_{h=1}^{L} \frac{W_h^2 S_{hj}^2}{n_{h(a)}}; j = 1,2,...,p \qquad (2.16)$$

Using (2.11) the variances $V(\bar{y}_{jst})$ under individual optimum allocations ignoring fpc are given by

$$V_j^* = V(\bar{y}_{jst})_{opt} = \frac{(\sum_{h=1}^{L} W_h S_{hj})^2}{n}; j = 1,2,\ldots,p \tag{2.17}$$

Using (2.4) the variances under proportional allocation ignoring fpc are given as

$$V(\bar{y}_{jst})_{prop} = \frac{\sum_{h=1}^{L} W_h S_{hj}^2}{n}; j = 1,2,\ldots,p. \tag{2.18}$$

Using the data given by Jessen (1942), Cochran showed that the average allocation gives results almost as precise as if it were possible to use individual optimum allocations.

In working out the compromise allocation he assumed all characteristics equally important. The author suggests that a more precise compromise allocation may be obtained if weighted averages are used instead of simple averages of $n_{hj}^*$. As regards the selection of weights for various characteristics it would be reasonable to take them proportional to the respective individual optimum variances given by (2.17) that is:

$a_j \propto V_j^*$

or $a_j = KV_j^*; j = 1,2,\ldots,p$ (2.19)

where $a_j > 0; j = 1,2,\ldots,p$ denote the weights assigned to the individual optimum allocations $n_{hj}^*$ and $K$ is the constant of proportionality.

or $\sum_{j=1}^{p} a_j = K \sum_{j=1}^{p} V_j^*$

$$\text{or } K = \frac{\sum_{j=1}^{p} a_j}{\sum_{j=1}^{p} V_j^*} = \frac{1}{V_j^*} \qquad (2.20)$$

(by putting the sum of weights $\sum_{j=1}^{p} a_j$ equal to 1)

where $V_j^* = V(\bar{y}_{jst})_{opt}$; $j = 1,2,...,p$ are as given in (2.15).

Substituting the value of $K$ from (2.20) in (2.19) we get

$$a_j = \frac{V_j^*}{\sum_{j=1}^{p} V_j^*}, j = 1,2,....,p \qquad (2\ 21)$$

where $a_j > 0$ and $\sum_{j=1}^{p} a_j = 1$

The weighted averages of $n_{hj}^*$; $h = 1,2,...,L$ as the proposed compromise allocation are thus given as:

$$n_{h(w)} = \sum_{j=1}^{p} a_j n_{hj}^*$$

$$= \sum_{j=1}^{p} \frac{V_j^*}{\sum_{j=1}^{p} V_j^*} n_{hj}^*$$

$$= \frac{(\sum_{j=1}^{p} V_j^* n_{hj}^*)}{\sum_{j=1}^{p} V_j^*}; h = 1,2,\ldots,L \tag{2.22}$$

Now using (2.14) and (2.17) we get

$$V_j^* n_{hj}^* = \frac{1}{n}(\sum_{h=1}^{L} W_h S_{hj})^2 \left( n.\frac{W_h S_{hj}}{\sum_{h=1}^{L} W_h S_{hj}} \right)$$

$$= W_h S_{hj} (\sum_{h=1}^{L} W_h S_{hj}) \tag{2.23}$$

By (2.22) and (2.23)

$$n_{h(w)} = \frac{\sum_{j=1}^{p} W_h S_{hj} (\sum_{h=1}^{L} W_h S_{hj})}{\frac{1}{n}\sum_{j=1}^{p}(\sum_{h=1}^{L} W_h S_{hj})^2}$$

$$= n\frac{\left(\sum_{j=1}^{p} W_h S_{hj} (\sum_{h=1}^{L} W_h S_{hj})\right)}{\sum_{j=1}^{p}(\sum_{h=1}^{L} W_h S_{hj})^2}; h = 1,2,\ldots,L. \tag{2.24}$$

The variances ignoring (fpc) under this allocation may be obtained by substituting $n_{h(w)}$ given by (2.24) for $n_h$ in (2.1) as

$$V_{j(w)} = V(\bar{y}_{jst}) = \sum_{h=1}^{L} \frac{W_h^2 S_{hj}^2}{n\sum_{j=1}^{p}(W_h S_{hj} \sum_{h=1}^{L} W_h S_{hj}) / \sum_{j=1}^{p}(\sum_{h=1}^{L} W_h S_{hj})}$$

$$= \left( \frac{\sum_{j=1}^{p}(\sum_{h=1}^{L} W_h S_{hj})^2}{n} \right) \sum_{h=1}^{L} \left( \frac{W_h^2 S_{hj}^2}{\sum_{j=1}^{p} W_h S_{hj} (\sum_{h=1}^{L} W_h S_{hj})} \right); j = 1,2,\ldots,p \qquad (2.25)$$

In practice usually the values of $S_{hj}^2$ are not known. In such situations their usual unbiased sample estimates $s_{hj}^2$ may be used. All the above expressions will be exactly same in this situation except that $S_{hj}$ is replaced by $s_{hj}$.

## 2.7 THE MINIMUM DEVIATION COMPROMISE ALLOCATION

Chatterjee (1967) used, the compromise criterion of minimizing the total proportional increase in individual optimum variances due to the use of a non-optimal allocation for obtaining a compromise allocation. He worked out the expression for the sample size $n_h$ for the hth stratum for a fixed total sample size $n$ as

$$n_{h(c)} = n.\frac{\sqrt{\sum_{j=1}^{p} n_{hj}^{*2}}}{\sum_{h=1}^{L} \sqrt{\sum_{j=1}^{p} n_{hj}^{*2}}}; h = 1,2,..,L \qquad (2.26)$$

where the symbol $(c)$ stands for Chatterjee. (see Cochran (1977)).

The corresponding variances are

$$V_{j(c)} = \sum_{h=1}^{L} \frac{W_h^2 S_{hj}^2}{n_{h(c)}}; j = 1,2,\ldots,p \tag{2.27}$$

We may obtain a more precise compromise allocation if instead of minimizing the total proportional increase in the individual optimum variances due to the use of a non-optimal allocation, we minimize the total deviation '$D$' from the individual optimum variances.

Using (2.11) ignoring fpc and (2.17) the total deviation $D$ may be expressed as

$$D = \sum_{j=1}^{p} \left(V_j - V_j^*\right)$$

$$= \sum_{j=1}^{p} \left( \sum_{h=1}^{L} \frac{W_h^2 S_{hj}^2}{n_h} - \frac{(\sum_{h=1}^{L} W_h S_{hj})^2}{n} \right) \tag{2.28}$$

where $V_j = \sum_{h=1}^{L} \frac{W_h^2 S_{hj}^2}{n_h}$ denote the sampling variance of $\bar{y}_{jst}$ under any general allocation $n_h$ and $V_j^* = \frac{(\sum_{h=1}^{L} W_h S_{hj})^2}{n}$ denote the sampling variance of $\bar{y}_{jst}$ under optimum allocation for fixed total sample size $n$ ignoring fpc.

As $V_j \geq V_j^*; j = 1,2,\ldots,p$, the quantity inside ( ) in (2.26) is always positive and $D = \sum_{j=1}^{p} \left(V_j - V_j^*\right)$ will present the true magnitude of the total deviation of the sampling variances of $\bar{y}_{jst}$ from $V_j^*$ for not using the individual optimum allocations. Thus a reasonable compromise criterion for working out the values of the compromise allocation $n_h$ would be to minimize $D$ subject to $\sum_{h=1}^{L} n_h = n$, that is by solving the optimization problem:

"Minimize $D$ given by (2.28) subject to $\sum_{h=1}^{L} n_h = n$" (2.29)

The problem (2.29) can be solved easily by using Lagrange multipliers technique as follows. Define the Lagrangian function $\phi$ as

$$\phi(n_h, \lambda) = D + \lambda\left(\sum_{h=1}^{L} n_h - n\right)$$

$$= \sum_{j=1}^{p} \left( \sum_{h=1}^{L} \frac{W_h^2 S_{hj}^2}{n_h} - \frac{\left(\sum_{h=1}^{L} W_h S_{hj}\right)^2}{n} \right) + \lambda\left(\sum_{h=1}^{L} n_h - n\right)$$

Differentiating $\phi$ with respect to $n_h; h = 1,2,\ldots,L$ and $\lambda$ and equating the partial derivatives thus obtained to zero we get the following $L+1$ simultaneous equations.

$$= -\sum_{j=1}^{p} \frac{W_h^2 S_{hj}^2}{n_h^2} + \lambda = 0; h = 1,2,\ldots,L \qquad (2.30)$$

and $\frac{\partial \phi}{\partial \lambda} = \sum_{h=1}^{L} n_h - n = 0$ (2.32)

(2.32) gives

$$\lambda = \sum_{j=1}^{p} \frac{W_h^2 S_{hj}^2}{n_h^2}$$

or $n_h^2 = \frac{1}{\lambda} \sum_{j=1}^{p} W_h^2 S_{hj}^2 ; h = 1,2,...,L$

or $n_h = \frac{1}{\sqrt{\lambda}} \sqrt{\sum_{j=1}^{p} W_h^2 S_{hj}^2} ; h = 1,2,...,L$ (2.33)

Substitution of the value of $n_h$ from (2.33) in (2.32) gives

$$\frac{1}{\sqrt{\lambda}} \sum_{h=1}^{L} \sqrt{\sum_{j=1}^{p} W_h^2 S_{hj}^2} - n = 0$$

or $n = \frac{1}{\sqrt{\lambda}} \sum_{h=1}^{L} \sqrt{\sum_{j=1}^{p} W_h^2 S_{hj}^2}$

or $\frac{1}{\sqrt{\lambda}} = \frac{n}{\sum_{h=1}^{L} \sqrt{\sum_{j=1}^{p} W_h^2 S_{hj}^2}}$ (2.34)

Substituting the value of $\frac{1}{\sqrt{\lambda}}$ from (2.34) in (2.33) we get the compromise allocation $n_{h(d)}$ based on minimum total deviation as

$$n_{h(d)} = n.\frac{\sqrt{\sum_{j=1}^{p} W_h^2 S_{hj}^2}}{\sum_{h=1}^{L}\sqrt{\sum_{j=1}^{p} W_h^2 S_{hj}^2}}; h = 1,2,...,L \tag{2.35}$$

where the symbol $(d)$ stands for deviation.

Thc variances $V_{j(d)}$ (ignoring fpc) under this allocation can be obtained by substituting $n_h = n_{h(d)}$ in (2.1). Thus

$$V_{j(d)} = \sum_{h=1}^{L} \frac{W_h^2 S_{hj}^2}{n_{h(d)}}; j = 1,2,...,p \tag{2.36}$$

## 2.8 NUMERICAL COMPARISONS

**Example 1**: Data used in the example are from Jessen (1942). The state of Iowa was divided into five geographic regions, each denoted by its major agricultural enterprise. These regions are to be used as strata in survey on dairy farming. The three items of most interest are the number of cows milked per day, the number of gallons of milk per day, and the total annual cash receipts from daily products. From a survey made in 1938, the estimated standard deviations $s_{hj}$ within strata are shown in Table 2.1. It has been decided to fix the total sample size $n$ as 1000. .

The proposed compromise allocation given by (2.24) based on weighted averages along with the corresponding expected variances given by (2.25) for the values given in Table 2.1 are worked out as

$n_{1(w)} = 236$, $n_{2(w)} = 246$, $n_{3(w)} = 194$, $n_{4(w)} = 115$ and $n_{5(w)} = 209$

with $V_{1(w)} = 0.0130$, $V_{2(w)} = 0.0811$, $V_{3(w)} = 76.9$ respectively.

**Table 2.1**

Standard deviations within strata

| Stratum No. $h$ | $W_h$ | Cows Milked $s_{h1}$ | Gallons of Milk $s_{h2}$ | Receipts for Dairy Products $s_{h3}$ |
|---|---|---|---|---|
| 1. | 0.197 | 4.6 | 11.7 | 332 |
| 2. | 0.191 | 3.4 | 9.8 | 357 |
| 3. | 0.219 | 3.3 | 7.0 | 246 |
| 4. | 0.184 | 2.8 | 6.5 | 173 |
| 5. | 0.208 | 3.7 | 9.8 | 279 |

The proposed compromise allocation based on minimum total deviation given by (2.34) and the corresponding expected variances given by (2.33) are worked out as:

$n_{1(d)} = 236$, $n_{2(d)} = 246$, $n_{3(d)} = 194$, $n_{4(d)} = 115$ and $n_{5(d)} = 209$

with $V_{1(d)} = 0.0130$, $V_{2(d)} = 0.0811$, $V_{3(d)} = 76.9$ respectively.

The sample sizes for a fixed total of 1000 under different allocations discussed in Sections 2.3 to 2.7 are summarized in Table 2.2. Table 2.3 shows the expected variances of $\bar{y}_{jst}$ under the allocations given in Table 2.2.

If $T(\underline{n})_K$ denote the trace of the variance-covariance matrix of $\bar{y}_{jst}; j = 1,2,...,p$ for a given allocation $(\underline{n})_K = (n_1, n_2, ...., n_L)_K$. It is to be noted that this variance-covariance matrix will be a diagonal matrix when the

**Table 2.2**

Sample sizes within strata ($n$=1000)

| Stratum No. $h$ | Allocation | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Proportional $n_{h(p)}$ | Optimum for | | | Average $n_{h(a)}$ | Chatterjee's $n_{h(c)}$ | Proposed $n_{h(w)}$ | Proposed $n_{h(d)}$ |
| | | Cows $n_{h1}^{*}$ | Gallons $n_{h2}^{*}$ | Receipts $n_{h3}^{*}$ | | | | |
| 1 | 197 | 254 | 258 | 236 | 205 | 249 | 236 | 236 |
| 2 | 191 | 182 | 209 | 246 | 212 | 213 | 246 | 246 |
| 3 | 219 | 203 | 171 | 194 | 189 | 189 | 194 | 194 |
| 4 | 184 | 145 | 134 | 115 | 131 | 132 | 115 | 115 |
| 5 | 208 | 216 | 228 | 209 | 218 | 217 | 209 | 209 |

characteristics are mutually uncorrelated. The relative efficiency of the allocation $(\underline{n})_K$ with respect to another allocation $(n)_{K'} = (n_1, n_2 ......, n_L)_{K'}$ may be defined as the ratio:

$T(\underline{n})_{K'} / T(\underline{n})_K$ . (see Sukhatme et.al.(1994)).

$(\underline{n})_K = (n_1, n_2, ...., n_L)_K$ denotes an L-component vector such that $n_h > 0; h = 1,2,..., L$ and $\sum_{h=1}^{L} n_h = n$ (the total sample size).

The last column of Table 2.3 gives the relative efficiencies of various allocations with respect to the proportional allocation.

**Table 2.3**

Expected Variances of the estimated mean

| Type of allocation | Cows | Gallons | Receipts | Trace | R.E. w.r.t. Proportional allocation |
|---|---|---|---|---|---|
| Optimum $n_{hj}^*$ | 0.0127 | 0.0800 | 76.9 | 76.9927 | 1.0520 |
| Average $n_{h(a)}$ | 0.0128 | 0.0802 | 77.9 | 77.6930 | 1.0425 |
| Chatterjee $n_{h(c)}$ | 0.0128 | 0.0800 | 77.5 | 77.5928 | 1.0438 |
| Proposed $n_{h(w)}$ | 0.0130 | 0.0811 | 76.9 | 76.9941 | 1.0520 |
| Proposed $n_{h(d)}$ | 0.0130 | 0.0811 | 76.9 | 76.9941 | 1.0520 |
| Proportional $n_{h(p)}$ | 0.0130 | 0.0837 | 80.9 | 80.9968 | - |

It is observed that all the compromise allocations are more efficient than the proportional allocation. However, the proposed compromise allocation based on weighted averages and the minimum deviation are equally good and most efficient. The percentage gain in efficiency in using the proposed allocations over the proportional allocation is 5.2% where as the corresponding value for average allocation is 4.2% and for Chatterjee's allocation is 4.4%. Thus the proposed allocations are more precise than other compromise allocations.

**Example 2:** The data are from a farm survey in Iowa reported by Jessen (1942) (see Sukhatme et al., (1984)). The relevant data with respect to three characteristics (i) number of hogs bought during the year (ii) number of cattle bought during the year and (iii) number of cows milked during the year, are shown in Table 2.4

**Table 2.4**

Estimated strata mean squares

| Stratum No. $h$ | $W_h$ | Hogs bought $s_{h1}^2$ | Cattle bought $s_{h2}^2$ | Cows Milked $s_{h3}^2$ |
|---|---|---|---|---|
| 1 | 0.197 | 12 | 56 | 41.3 |
| 2 | 0.191 | 80 | 2,132 | 23.1 |
| 3 | 0.219 | 1,113 | 565 | 10.9 |
| 4 | 0.184 | 84 | 355 | 11.5 |
| 5 | 0.208 | 247 | 68 | 38.8 |

It has been decided to fix the total sample size as $n = 1000$

**Table 2.5**

Sample sizes within strata ($n$=1000)

| Stratum No $h$ | Proportional $n_{h(p)}$ | Optimum for | | | Average $n_h$ | Chatterjee's $n_{h(c)}$ | Proposed $n_{h(w)}$ | Proposed $n_{h(d)}$ |
|---|---|---|---|---|---|---|---|---|
| | | Hogs $n^*_{h1}$ | Cattle $n^*_{h2}$ | Milk $n^*_{h3}$ | | | | |
| 1 | 197 | 46 | 71 | 262 | 126 | 142 | 70 | 73 |
| 2 | 191 | 117 | 426 | 190 | 244 | 248 | 318 | 323 |
| 3 | 219 | 499 | 252 | 150 | 300 | 299 | 328 | 323 |
| 4 | 184 | 115 | 168 | 129 | 138 | 125 | 149 | 141 |
| 5 | 208 | 223 | 83 | 269 | 192 | 186 | 135 | 140 |

The proposed compromise allocation based on weighted averages given by (2.24) and the corresponding expected variances given by (2.25) are:

$n_{1(w)} = 69$, $n_{2(w)} = 318$, $n_{3(w)} = 328$, $n_{4(w)} = 150$ and $n_{5(w)} = 135$

with $V_{1(w)} = 0.2766$, $V_{2(w)} = 0.4606$ and $V_{3(w)} = 0.0424$ respectively.

The proposed compromise allocation based on minimum deviation given by (2.34) and the corresponding expected variances given by (2. 35) are

$n_{1(d)} = 73$, $n_{2(d)} = 323$, $n_{3(d)} = 323$, $n_{4(d)} = 141$ and $n_{5(d)} = 140$

with $V_{1(d)} = 0.2772$, $V_{2(d)} = 0.4607$ and $V_{3(d)} = 0.0409$ respectively.

Table 2.5 gives the different allocations. The optimum expected variances under various allocations are shown in Table 2.6. The last column shows the relative efficiency of different allocations with respect to proportional allocation based on the ratio of traces of the variance-covariance matrices of $\bar{y}_{jst}$ under different allocations.

It is observed that, all the compromise allocations are more efficient than the proportional allocation. However the two proposed compromise allocations are more efficient than other compromise allocations. The compromise allocation based on minimum deviation is the most efficient. The percentage gain in efficiency is about 24.4% where as the same figure corresponding to average allocation is only 11.7% and for Chatterjee's allocation is merely 5.9%.

**Table 2.6**

Expected variances of the estimated mean

| Type of allocation | Hog | Cattle | Milk | Trace | R.E. w.r.t. Proportional allocation |
|---|---|---|---|---|---|
| Optimum $n^*_{hj}$ | 0.2147 | 0.4277 | 0.0233 | 0.6657 | 1.4556 |
| Average $n_{h(a)}$ | 0.2694 | 0.5291 | 0.0293 | 0.8278 | 1.1170 |
| Chatterjee $n_{h(c)}$ | 0.2739 | 0.5314 | 0.1097 | 0.9150 | 1.0590 |
| Proposed $n_{h(w)}$ | 0.2768 | 0.4607 | 0.0408 | 0.7783 | 1.2450 |
| Proposed $n_{h(d)}$ | 0.2772 | 0.4607 | 0.0409 | 0.7788 | 1.2442 |
| Proportional $n_{h(p)}$ | 0.3260 | 0.6180 | 0.0250 | 0.9690 | – |

## 2.9 ALLOCATION WITH VARIABLE COST OF MEASUREMENT

Let $c_{hj}; h=1,2,...,L; j=1,2,...,p$, denote the per unit cost of measuring the jth characteristic in hth stratum. Also let, out of the total budget '$C$ ', $n_{h(d)}$, denote the cost allocated for measuring the jth characteristic. The individual optimum allocations using (2.7) are given as

$$n^*_{hj} = n\frac{W_h S_{hj} / \sqrt{c_{hj}}}{\sum\limits_{h=1}^{L} W_h S_{hj} / \sqrt{c_{hj}}}; h=1,2,....,L; j=1,2,....,p \tag{2.36}$$

where $n$ given by (2.8) is

$$n = \frac{C_j \sum_{h=1}^{L}\left(W_h S_{hj} / \sqrt{c_{hj}}\right)}{\sum_{h=1}^{L}\left(W_h S_{hj} \sqrt{c_{hj}}\right)} \tag{2.37}$$

Substitution of the value of $n$ from (2.37) in (2.36) gives

$$n_{hj}^{*} = \frac{C_j W_h S_{hj} / \sqrt{c_{hj}}}{\sum_{h=1}^{L} W_h S_{hj} \sqrt{c_{hj}}}; h = 1,2,....,h; j = 1,2,..........p \tag{2.38}$$

where overhead cost $c_o$ is ignored, that is the cost functions for individual allocations are taken as

$$C_j = \sum_{h=1}^{L} c_{hj} n_{hj}; j = 1,2,...,p \tag{2.39}$$

The optimum value of the variance $V(\bar{y}_{jst})$ (ignoring fpc) of the estimate $\bar{y}_{jst}$ of the population mean $\bar{Y}_j$ of the jth characteristic under the optimum allocation is given by

$$V_j^{*} = V(\bar{y}_{jst})_{opt} = \frac{\left(\sum_{h=1}^{L} W_h S_{hj} \sqrt{c_{hj}}\right)^2}{C_j}; j = 1,2,..., p \tag{2.40}$$

(see Cochran (1977)).

For working out a compromise allocation we have to restructure the cost setup as below.

Let $c_h = \sum_{j=1}^{p} c_{hj}$; denote the per unit cost of measuring all the 'p' characteristics in the hth stratum. Then for any compromise allocation $\underline{n} = (n_1, n_2, ..., n_L)$, we have

$$C = \sum_{h=1}^{L} c_h n_h \tag{2.41}$$

as the cost constraint, where $C = \sum_{j=1}^{p} C_j$ is the total fixed budget. The variances $V_j; j = 1,2,...,p$ (ignoring fpc) under a compromise allocation $n_h; h = 1,2,...,L$ can be worked out directly by using

$$V_j = \sum_{h=1}^{L} \frac{W_h^2 S_{hj}^2}{n_h}; j = 1,2,...,p \tag{2.42}$$

Due to this restructured cost setup it is not advisable to use the Cochran's average allocation or the allocation based weighted averages proposed in Section 2.6 because these compromise allocations either do not utilize the cost fully or become infeasible by violating the cost constraint in (2.41).

Chatterjee's allocation discussed in Section 2.7 can be used to work out compromise allocation for fixed total cost. It gives the compromise allocation as:

$$n_{h(c)} = \frac{C\sqrt{\sum_{j=1}^{p} n_{hj}^{*2}}}{\sum_{h=1}^{L} c_h \sqrt{\sum_{j=1}^{p} n_{hj}^{*2}}}; h = 1,2,...,L \tag{2.43}$$

The corresponding variances can be obtained by putting $n_h = n_{h(c)}$ in (2.42)

Thus

$$V_{j(c)} = \sum_{h=1}^{L} \frac{W_h^2 S_{hj}^2}{n_{h(c)}}; j = 1,2,...,p \qquad (2.44)$$

where the symbol $(c)$ stands for Chatterjee. In fact Chatterjee's allocation given in Section 2.7 is a special case $(c_h = c)$ of his general allocation given in (2.43)

## 2.11 THE MINIMUM DEVIATION COMPROMISE ALLOCATION FOR FIXED COST

As discussed in Section 2.9 the total deviation $D$ for fixed cost is given as $D = \sum_{j=1}^{p} \left(V_j - V_j^*\right)$

$$= \sum_{j=1}^{p} \left( \sum_{h=1}^{L} \frac{W_h S_{hj}^2}{n_h} - \frac{(\sum_{h=1}^{L} W_h S_{hj} \sqrt{c_{hj}})^2}{C_j} \right) \qquad (2.45)$$

where $V_j^*$ is given by (2.40) .The problem of allocation thus become to find $n_h; h = 1,2,...,L$ that minimize $D$ given by (2.45 )subject to the cost constraint in (2.41)

Defining the Lagrangian function as

$$\phi(n_h, \lambda) = D + \lambda \left( \sum_{h=1}^{L} c_h n_h - C \right) \qquad (2.46)$$

and equating the partial derivatives $\frac{\partial \phi}{\partial n_h}$ and $\frac{\partial \phi}{\partial \lambda}$ equal to zero and solving the

$(L+1)$ simultaneous equations thus obtained we get the minimum deviation compromise allocation $n_{h(d)}$ as

$$n_{h(d)} = \frac{C\sqrt{\sum_{j=1}^{p} W_h S_{hj}^2}}{\sqrt{c_h}\sqrt{\sum_{j=1}^{p} W_h S_{hj}^2}};h = 1,2,....,L \tag{2.47}$$

where the symbol ($d$) indicates that the compromise allocation is based on minimum deviation.

The corresponding variances can be obtained by using

$$V_{j(d)} = \sum_{h=1}^{L} \frac{W_h^2 S_{hj}^2}{n_{h(d)}}; j = 1,2,...,p \tag{2.48}$$

## 2.11 A NUMERICAL ILLUSTRATION

The data of this example are from Jahan, N. et al (1994). In a two-variate survey ($p=2$) the population is stratified into two strata ($L=2$). The following information are available.

Also

$$((c_{hj})) = \begin{pmatrix} 4 & 20 \\ 5 & 22 \end{pmatrix} \Rightarrow ((c_h)) = (24,27)$$

$((C_j)) = (400,2000) \Rightarrow C = 2400$

**Table 2.7**

Data for two strata and two characteristics

| Stratum No. $h$ | $W_h$ | $S_{h1}$ | $S_{h2}$ |
|---|---|---|---|
| 1 | 0.4 | 4.6 | 332 |
| 2 | 0.6 | 2.8 | 173 |

The individual optimum allocations are worked out using formula (2.36) as $((n_{hj}^*)) = \begin{pmatrix} 49 & 55 \\ 40 & 41 \end{pmatrix}$

with $V_1^* = 0.1384$ and $V_2^* = 584.0261$

The average allocation is given as

$$((n_{h(a)})) = \begin{pmatrix} 52 \\ 41 \end{pmatrix}$$

The cost associated with this allocation is

52 x 24 + 41 x 27 = 2355

which is less than the available cost 2400 and hence it is not advisable to use this allocation. However, for the sake of comparison the variances under this allocation are worked out as

$V_{1(a)} = 0.1339$ and $V_{2(a)} = 601.9420$.

The weighted average allocation proposed by author in Section 2.6 is given as $\left(\left(n_{h(w)}\right)\right)=\begin{pmatrix}55\\41\end{pmatrix}$

The cost associated with this allocation is

55 x 24 + 41 x 27 = 2427

which is more than the available cost 2400. Hence this allocation is infeasible and cannot be used. However, for the sake of comparison the variances under this allocation are worked out as

$V_{1(w)} = 0.1315$ and $V_{2(w)} = 589.38308$ ,

after adjusting the $n_{h(w)}$ to maintain the feasibility by multiplying it by an adjustment factor of 2400/2427=0.9889 .

Thus the adjusted $\left(\left(n_{h(w)}\right)\right)=\begin{pmatrix}54\\41\end{pmatrix}$.

Compromise allocation given by Chatterjee's, using formula (2.43) is

$\left(\left(n_{h(c)}\right)\right)=\begin{pmatrix}53\\41\end{pmatrix}$ with corresponding variances

$V_{1(c)} = 0.1327$ and $V_{2(c)} = 595.5429$ .

The proposed minimum deviation allocation using formula (2.47) is

$\left(\left(n_{h(d)}\right)\right)=\begin{pmatrix}55\\40\end{pmatrix}$ with corresponding variances

$V_{1(d)} = 0.1322$ and $V_{2(d)} = 590.0126$.

The proportional allocation for a fixed cost C is given by

$$n_{h(p)} = \frac{CW_h}{\sum_{h=1}^{L} W_h c_h}$$

(see Sukhatme et al (1984)).

For $C = 2400$, $W_1 = 0.4$, $W_2 = 0.6$, $c_1 = 24$, $c_2 = 27$

The proportional allocation $n_{h(p)}; h = 1,2$ is worked out as

$$n_{1(p)} = \frac{2400 \times 0.4}{0.4 \times 24 + 0.6 \times 27} = 37.2093 \cong 37$$

and $n_{2(p)} = \dfrac{2400 \times 0.6}{0.4 \times 24 + 0.6 \times 27} = 55.8139 \cong 56$

Thus $\left(\left(n_{h(p)}\right)\right) = \begin{pmatrix} 37 \\ 56 \end{pmatrix}$. The corresponding variances are

$V_{1(p)} = 0.1419$ and $V_{2(p)} = 669.0450$ .

Where the symbol $(p)$ stands for proportional.

These results arranged in a tabular form are given in Tables 2.8 and 2.9

**Table 2.8**

Sample sizes within strata

| Stratum No. $h$ | Compromise allocations $(n_h)$ | | | | |
|---|---|---|---|---|---|
| | Average $n_{h(a)}$ | Weighted average $n_{h(w)}$ | Chatterjee's $n_{h(c)}$ | Proposed $n_{h(d)}$ | Proportional $n_{h(p)}$ |
| 1 | 52 | 55 | 53 | 55 | 37 |
| 2 | 41 | 40 | 41 | 40 | 56 |

**Table 2.9**

Variances under different compromise allocations

| Allocations | $V_1$ | $V_2$ | Trace | R.E. w.r.t. Proportional |
|---|---|---|---|---|
| Average $n_{h(a)}$ | 0.1339 | 601.9420 | 602.0359 | 1.1115 |
| Chatterjee's $n_{h(c)}$ | 0.1327 | 595.5429 | 595.6756 | 1.1234 |
| Proposed $n_{h(w)}$ | 0.1315 | 589.3808 | 589.5123 | 1.1351 |
| Proposed $n_{h(d)}$ | 0.1322 | 590.0126 | 590.1448 | 1.1339 |
| Proportional $n_{h(p)}$ | 0.1419 | 669.0450 | 669.1869 | - |

The last column of Table 2.9 shows the relative efficiencies with respect to the proportional allocation based on the ratio of traces of the variance-covariance matrices.

It is observed that all the compromise allocations are more efficient than the proportional allocation. However, both the proposed allocations are more efficient than other compromise allocations. The compromise allocation based on weighted averages is the most efficient allocation for this example with the percentage gain in efficiency over proportional allocation as 13.5%.

## 2.12 CONCLUSION

The three numerical examples worked out in Sections 2.9 and 2.12 indicate that the compromise allocations based on (i) weighted averages (the weights $a_j$; proportional to individual optimum variances $V_j^*; j = 1,2,...,p$)

and (ii) minimizing total deviation $D=\sum_{j=1}^{p}\left(V_j - V_j^*\right)$ are more efficient than other compromise allocations existing in the sampling literature. The criterion for working out the relative efficiency is the ratio of the trace of the variance-covariance matrix of $\bar{y}_{jst}; j=1,2,...,p$ under proportional allocation to the trace under a given compromise allocation. Thus the proposed compromise allocations are an improvement over the compromise allocations already existing in the sampling literature.

# CHAPTER-III

## MIXED ALLOCATION IN STRATIFIED SAMPLING

# CHAPTER – III

## MIXED ALLOCATION IN STRATIFIED SAMPLING

### 3.1 INTRODUCTION

It is stratified sampling where the population of size N is divided into L strata of sizes $N_1, N_2, \ ,N_2$ $(\sum_{h=1}^{L} N_h = N)$ the variance of the stratified sample mean $\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{Y}_h$ is given by

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} - \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{N_h} \tag{3.1}$$

The total cost 'C' of the survey may be given as

$$C = c_o + \sum_{h=1}^{L} c_h n_h$$

or $$C - c_o = c_o + \sum_{h=1}^{L} c_h n_h = C_o \ \text{(say)} \tag{3.2}$$

where all the symbols have the same meaning as defined in Section 2, Chapter 2 of this thesis. The cost structure of the survey may be more complicated than given in (3.2) (See Hansen et .al. (1953) and Groves (1989)). For example the cost function may be of the form

$$C = c_o + \sum_{h=1}^{L} t_h \sqrt{n_h} + \sum c_h n_h$$

where $t_h$ denote the traveling cost between the selected units of the hth stratum. Csenki (1977) used the cost function of the form

$$C = c_o + \sum_{h=1}^{L} c_h n_h^{\delta}$$

where $\delta > 0$ is a known constant.

In spite of all the above discussed cost functions the cost function given in (3.2) is often serves as an adequate approximation for practical purposes.

The fixed cost allocation that minimizes $V(\bar{y}_{st})$ is well known in sampling literature as optimum allocation is given as

$$n_h = n \frac{W_h S_h / \sqrt{c_h}}{\sum_{h=1}^{L} W_h S_h / \sqrt{c_h}}; h = 1,2.........,L \tag{3.3}$$

where the total sample size n for fixed cost C is given as

$$n = C_o \frac{\sum_{h=1}^{L} \left(W_h S_h / \sqrt{c_h}\right)}{\sum_{h=1}^{L} \left(W_h S_h / \sqrt{c_h}\right)}; h = 1,2,...,L \tag{3.4}$$

where $C_o = C - c_o$

Using (3.3) and (3.4), we get

$$n_h = C_o \frac{\left(W_h S_h / \sqrt{c_h}\right)}{\sum_{h=1}^{L}\left(W_h S_h \sqrt{c_h}\right)}; h = 1,2,..,L \tag{3.5}$$

Substituting $n_h$ given by (3.5) in (3.1) the value of the variance $V(\bar{y}_{st})$ under optimum allocation comes out to be

$$V^* = V(\bar{y})_{opt} = \frac{\left(\sum_{h=1}^{L} W_h S_h \sqrt{c_h}\right)^2}{C_o} - \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{N_h} \tag{3.6}$$

(see Cochran (1977)).

The optimum allocation can also be worked out to minimize the cost for fixed variance. Using Cauchy -Schwarz inequality Stuart (1945) showed that in terms of total sample size n the expression of the sample sizes $n_h; h = 1,2,...,L$ that minimize the cost for fixed variance $V(\bar{y}_{st})$ can also be given by (3.3). The value of $n$, the total sample size, in this case is given by

$$n = \frac{\left(\sum_{h=1}^{L} W_h S_h \sqrt{c_h}\right)\left(\sum_{h=1}^{L} W_h S_h / \sqrt{c_h}\right)}{V + \frac{1}{N}\sum_{h=1}^{L} W_h S_h^2} \tag{3.7}$$

where $V = V(\bar{y}_{st}) + \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{N_h} = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h}$, is fixed .

Substituting the value of n form (3.7) in (3.3) we get

$$n_h = \frac{(W_h S_h / \sqrt{c_h})(\sum_{h=1}^{L} W_h S_h \sqrt{c_h})}{V + \frac{1}{N}\sum_{h=1}^{L} W_h S_h^2}; h = 1,2,\ldots,L \tag{3.8}$$

The resulting minimum cost is

$$C_{opt} = c_o + \frac{(\sum_{h=1}^{L} W_h S_h \sqrt{c_h})^2}{V + \frac{1}{N}\sum_{h=1}^{1} W_h S_h^2} \tag{3.9}$$

obtained by substituting $n_h$ given by (3.8) in (3.2) ( see also Kish (1967)).

The practical difficulty in using optimum allocation is that usually $S_h^2$ are not known, thus we can only approximate this allocation by using estimated values of $S_h^2$. They may be the values computed on some previous occasion or values obtained by a pilot survey. Other allocations that are less precise than optimum allocation are proportional and equal allocations. In proportional allocation the sample sizes $n_h$ from various strata are proportional to the corresponding stratum weights $W_h$. This gives

$n_h \, \alpha \, W_h$

or $n_h = KW_h$. (3.10)

where $K$ is the constant of proportionality.

The proportional allocation may be worked out for fixed cost or for fixed total sample size. For fixed cost, under proportional allocation

$$n_h = \frac{C_o W_h}{\sum_{h=1}^{L} c_h W_h}; h = 1,2,...,L \tag{3.11}$$

$$\text{with } V(\bar{y}_{st})_{prop} = \frac{(\sum_{h=1}^{L} W_h S_h^2)(\sum_{h=1}^{L} c_h W_h)}{C_o} - \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{N_h} \tag{3.12}$$

for fixed total sample size, under proportional allocation

$$n_h = nW_h; h = 1,2,...,L \tag{3.13}$$

$$\text{with } V(\bar{y}_{st})_{prop} = \frac{\sum_{h=1}^{L} W_h S_h^2}{n} - \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{N_h} \tag{3.14}$$

(see Section 2.2, Chapter 2).

Practical implementation of proportional allocation is easy because usually the strata sizes $N_h$ and thus strata weights $W_h$ are known. In case $W_h$ are unknown they can also be estimated from a pilot survey.

In the absence of the true value of $W_h$, if other situations permit one can use equal allocation. To implement equal allocation only knowledge of the total sample size n and the number strata L are required. The sample sizes $n_h$ are given by

$$n_h = \frac{n}{L}; h = 1,2,...,L \tag{3.15}$$

The variance $V(\bar{y}_{st})$ under equal allocation is given as:

$$V(\bar{y}_{st})_{equal} = \frac{L\sum_{h=1}^{L} W_h^{\,2} S_h^{\,2}}{n} - \frac{\sum_{h=1}^{L} W_h S_h^{\,2}}{N} \tag{3.16}$$

In this chapter the problem of allocation of a sample to strata is discussed in general conditions. There are sometimes valid reasons due to which only a particular type of allocation is advisable in a particular part of a stratified population. Under this situation it would be reasonable to divide the group of strata into subgroups and use a particular type of allocation in one group. Clark and Steel (2000) used a similar idea in two-stage stratified sampling.

Such an allocation, which uses different type of allocations for different subgroups of strata, may be called a "Mixed allocation".

## 3.2.THE MIXED ALLOCATION

Let the group of L strata is divided into $k$ subgroups $G_1, G_2, ..., G_k$, where the group $G_j$ consists of $L_j; j = 1,2,...,k$ strata such that $\sum_{j=1}^{k} L_j = L$.

Without loss of generality we can assume that the first $L_1$ strata constitute the first subgroup $G_1$, the next $L_2$ strata constitute the second subgroup $G_2$, and so on and the last $L_k$ strata constitute the last subgroup $G_k$. Under this scheme, the jth subgroup $G_j; j = 1,2,...,k$ will consists of

$(\sum_{i=1}^{j-1} L_i + 1)$th, $(\sum_{i+1}^{j-1} L_i + 2)$ th,..., and $(\sum_{i=1}^{j-1} L_i + L_j) = (\sum_{i=1}^{j} L_i) th$ strata. .

Further let due to the prevailing circumstances in a particular subgroup a particular type of allocation is to be used. This could be done by letting

$$n_h = \alpha_j \beta_h ; h \in I_j ; j = 1,2,...,k \tag{3.17}$$

where $I_j ; j = 1,2,...,k$ is the set of indices of the strata that constitute the jth subgroup $G_j$,

$\beta_h ; h \in I_j ; j = 1,2,...,k$ are known constants depending upon the type of allocation to be used in the jth subgroup,

and $\alpha_j ; j = 1,2,...,k$ are to be determined.

For example if in any particular subgroup, say $G_p$, equal allocation is to be used then

$\beta_h = 1; h \in I_p$

Proportional allocation in the qth subgroup $G_q$ is characterized by

$\beta_h = W_h ; h \in I_q$

To use optimum allocation in the rth subgroup $G_r$, $\beta_h$ is given as

$\beta_h = \frac{W_h S_h}{\sqrt{c_h}} ; h \in I_r$ and so on.

Two other allocations that are used sometimes are allocation proportional to $W_h \bar{Y}_h$ and allocation proportional to $W_h R_h$ , where

$R_h; h = 1,2,...,L$ denote the range of the hth stratum (see Murthi (1967)). If any of the above allocations is to be introduced in a subgroup we may take $\beta_h = W_h\bar{Y}_h$ or $\beta_h = W_h R_h$ accordingly.

It can be seen that

$$I_j = \{ \sum_{l=1}^{j-1} L_l + 1,\ \sum_{l=1}^{j-1} L_l + 2, \ldots, \sum_{l=1}^{j-1} L_l + L_j = \sum_{l=1}^{j} L_l \}; j = 1,2, \quad ,k \tag{3.18}$$

where $I_r \bigcap I_s = \phi; r \neq s$ and $\bigcup_{j=1}^{k} I_j = \{1,2...,L\}$

The mixed allocation defined in (3.17) may be computed for minimizing $V(\bar{y}_{st})$ given by (3.1) for a fixed cost or for minimizing the total cost given by (3.2) for a fixed value of $V(\bar{y}_{st})$.

These optimization problems can be formulated as the following two nonlinear programming problems (NLPP)

**NLPP1:** (Minimizing $V(\bar{y}_{st})$ for fixed cost)

$$\text{Minimize } V(n_h) = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} \tag{3.19}$$

$$\text{subject to } \sum_{h=1}^{L} c_h n_h = C_o \tag{3.20}$$

$$n_h = \alpha_j \beta_h; h \in I_j; j = 1,2,...,k \tag{3.21}$$

$$n_h \geq 0; h = 1,2,...,L \tag{3.22}$$

where from the expression (3.1) of $V(\bar{y}_{st})$ the terms independent of $n_h$ are dropped and $C_o = C - c_o$ (3.23)

**NLPP2**: (Minimizing the cost for fixed value of $V(\bar{y}_{st})$ )

$$\text{Minimize } C(n_h) = \sum_{h=1}^{L} c_h n_h \tag{3.24}$$

$$\text{subject to } \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} = v \tag{3.25}$$

$$n_h = \alpha_j \beta_h ; h = 1,2,\ldots,k \tag{3.26}$$

$$\text{and} \qquad n_h \geq 0; h = 1,2,\ldots,L \tag{3.27}$$

where from the expression (3.2) of $C$ the term $c_o$ which is independent of $n_h$ is dropped

and $v = V + \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{N_h}$ ;$V$ being the fixed value of $V(\bar{y}_{st})$.

Using constraints $n_h = \alpha_j \beta_h ; h \in I_j ; j = 1,2,\ldots,k$ the expression $\sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h}$

and $\sum_{h=1}^{L} c_h n_h$ in NLPP 1 and 2 may be expressed as :

$$\sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} = \sum_{j=1}^{k} \sum_{h \in I_j} \frac{W_h^2 S_h^2}{\alpha_j \beta_h} \tag{3.28}$$

$$\text{and } \sum_{h=1}^{L} c_h n_h = \sum_{j=1}^{k} \sum_{h \in I_j} \alpha_j c_h \beta_h \tag{3.29}$$

respectively, where $I_j ; j = 1,2,\ldots,k$ are given by (3.18)

Using (3.24) and (3.25) the two NLPP can thus be restated as:

**NLPP 1:** Minimize $F_1(\alpha_j) = \sum_{j=1}^{k} \sum_{h \in I_j}^{k} \frac{W_h^2 S_h^2}{\alpha_j \beta_h}$ (3.30)

subject to $\sum_{j=1}^{k} \sum_{h \in I_j} \alpha_j c_h \beta_h = C_o$ (3.31)

and $\alpha_j \geq 0; j = 1,2,\ldots,k$ (3.32)

**NLPP2:** Minimize $F_2(\alpha_j) = \sum_{j=1}^{k} \sum_{h \in I_j} \alpha_j c_h \beta_h$ (3.33)

subject to $\sum_{j=1}^{k} \sum_{h \in I_j} \frac{W_h^2 S_h^2}{\alpha_j \beta_h} = v$ (3.34)

and $\alpha_j \geq 0; j = 1,2,\ldots,k$ (3.35)

Ignoring restrictions $\alpha_j \geq 0; j = 1,2,\ldots,k$ both the NLPP1 and 2 can be solved by using Lagrange multipliers technique. If the solutions thus obtained satisfy the restrictions $\alpha_j \geq 0; j = 1,2,\ldots,k$ and thus $n_h \geq 0; h = 1,2,\ldots,L$ also then the NLPP1 and 2 are solved completely, otherwise some nonlinear programming technique may be used to solve them.

## 3.3 THE SOLUTION

The NLPP1 after ignoring restrictions in (3.32) can be described as "Find $\alpha_j; j = 1,2,\ldots,k$ that minimize $F_1(\alpha_j)$ given by (3.30) subject to the constraint (3.31)".

The Lagrangian function $\phi_1(\alpha_j, \lambda_1)$ for this problem is

$$\phi_1(\alpha_j, \lambda_1) = \sum_{j=1}^{k} \sum_{h \in I_j} \frac{W_h^2 S_h^2}{\alpha_j \beta_h} + \lambda_1 (\sum_{j=1}^{k} \sum_{h \in I_j} \alpha_j c_h \beta_h - C_o) \quad (3.36)$$

where $\lambda_1$ is the Lagrange multiplier.

Differentiating $\phi_1$ with respect to $\alpha_j$ and $\lambda_1$ and equating the partial derivatives equal to zero we get the following $(k+1)$ simultaneous equations

$$\frac{\delta \phi_1}{\partial \alpha_j} = -\sum_{h \in I_j} \frac{W_h^2 S_h^2}{\alpha_j^2 \beta_h} + \lambda_1 \sum_{h \in I_j} c_h \beta_h = 0; \; j = 1,2,\ldots,k \quad (3.37)$$

and $$\frac{\delta \phi_1}{\partial \lambda_1} = \sum_{j=1}^{k} \sum_{h \in I_j} \alpha_j c_h \beta_h - C_o = 0 \quad (3.38)$$

From (3.37) $\lambda_1 \sum_{h \in I_j} c_h \beta_h = \frac{1}{\alpha_j^2} \sum_{h \in I_j} \frac{W_h^2 S_h^2}{\beta_h}$

(because $\alpha_j; j = 1,2,\ldots,k$ is constant within a particular subgroup)

or $$\alpha_j^2 = \frac{1}{\lambda_1} \frac{\sum_{h \in I_j} W_h^2 S_h^2 / \beta_h}{\sum_{h \in I_j} c_h \beta_h}$$

or $$\alpha_j = \frac{1}{\sqrt{\lambda_1}} \sqrt{\frac{\sum_{h \in I_j} W_h^2 S_h^2 / \beta_h}{\sum_{h \in I_j} c_h \beta_h}}; \; j = 1,2,\ldots,k \quad (3.39)$$

Substituting the value of $\alpha_j$ from (3.39) in (3.38) we get

$$\sum_{j=1}^{k}\sum_{h\in I_j}\frac{1}{\sqrt{\lambda_1}}\sqrt{\frac{\sum_{h\in I_j} W_h^2 S_h^2 / \beta_h}{\sum_{h\in I_j} c_h\beta_h}}\,c_h\beta_h = C_o$$

or $$\frac{1}{\sqrt{\lambda_1}} = \frac{C_o}{\sum_{j=1}^{k}\left( (\sum_{h\in I_j} c_h\beta_h)\sqrt{\frac{\sum_{h\in I_j} W_h^2 S_h^2 / \beta_h}{\sum_{h\in I_j} c_h\beta_h}}\right)}$$

or $$\frac{1}{\sqrt{\lambda_1}} = \frac{C_o}{\sum_{j=1}^{k}\sqrt{(\sum_{h\in I_j} c_h\beta_h)(\sum_{h\in I_j} W_h^2 S_h^2 / \beta_h)}} \tag{3.40}$$

From (3.39) and (3.40)

$$\alpha_j = C_o\frac{\sqrt{(\sum_{h\in I_j} W_h^2 S_h^2 / \beta_h)/(\sum_{h\in I_j} c_h\beta_h)}}{\sum_{j=1}^{k}\sqrt{(\sum_{h\in I_j} W_h^2 S_h^2 / \beta_h)(\sum_{h\in I_j} c_h\beta_h)}};\ j = 1,2,\ldots,k \tag{3.41}$$

The values of the sample sizes $n_h$ for the strata belonging to a particular subgroup say $G_p$, that is for $h \in I_p$ can be obtained by substituting the value of $\alpha_j$ given by (3.41) for $j=p$, in (3.17), where $p \in \{1,2,\ldots,k\}$

The resulting variance (ignoring fpc) is

$$V_{(mixed)} = \sum_{j=1}^{k}\sum_{h\in I_j}\frac{W_h^2 S_h^2}{\alpha_j \beta_h}$$

$$= \sum_{j=1}^{k} \sum_{h\in I_j} \frac{\left(W_h^2 S_h^2 / \beta_h\right)\sum_{j=1}^{k}\sqrt{(\sum_{h\in I_j} W_h^2 S_h^2 / \beta_h)(\sum_{h\in I_j} c_h \beta_h)}}{C_o \sum_{j=1}^{k}\sqrt{(\sum_{h\in I_j} W_h^2 S_h^2 / \beta_h)/(\sum_{h\in I_j} c_h \beta_h)}}$$

$$= \frac{\left(\sum_{j=1}^{k}\sqrt{(\sum_{h\in I_j} W_h^2 S_h^2 / \beta_h)(\sum_{h\in I_j} c_h \beta_h)}\right)^2}{C_o} \tag{3.42}$$

where the symbol '(mixed)' corresponds to the mixed allocation.

The NLPP2 after ignoring restrictions in (3.31) may be described as: "Find $\alpha_j; j = 1,2,...,k$ that minimize $F_2(\alpha_j)$ given by (3.33) subject to be constraint (3.34)".

To solve this problem, define the Lagragian function $\phi_2(\alpha_j, \lambda_2)$ as

$$\phi_2(\alpha_j, \lambda_2) = \sum_{j=1}^{k}\sum_{h\in I_j} \alpha_j c_h \beta_h + \lambda_2\left(\sum_{j=1}^{k}\sum_{h\in I_j} \frac{W_h^2 S_h^2}{\alpha_j \beta_h} - v\right) \tag{3.43}$$

where $\lambda_2$ is the Lagrange multiplier.

As before we have the (k+1) equations as

$$\frac{\partial \phi_2}{\partial \alpha_j} = \sum_{h\in I_j} c_h \beta_h - \lambda_2 \sum_{h\in I_j} \frac{W_h^2 S_h^2}{\alpha_j^2 \beta_h} = 0; j = 1,2,...,k \tag{3.44}$$

and $$\frac{\partial \phi}{\partial \lambda_2} = \sum_{j=1}^{k}\sum_{h\in I_j} \frac{W_h^2 S_h^2}{\alpha_j \beta_h} - v = 0; j = 1,2,...,k \tag{3.45}$$

From (3.44)

$$\alpha_j = \sqrt{\lambda_2}\sqrt{\frac{\sum_{h\in I_j} W_h^2 S_h^2 / \beta_h}{\sum_{h\in I_j} c_h \beta_h}}; j = 1,2,\ldots,k \tag{3.46}$$

Substituting the value of $\alpha_j$ from (3.46) in (3.45) we get

$$\sqrt{\lambda_2} = \frac{\sum_{j=1}^{k}\sqrt{(\sum_{h\in I_j} W_h^2 S_h^2 / \beta_h)(\sum_{h\in I_j} c_h \beta_h)}}{v} \tag{3.47}$$

From (3.46) and (3.47)

$$\alpha_j = \frac{1}{v}\left(\sqrt{(\sum_{h\in I_j} W_h^2 S_h^2 / \beta_h)/(\sum_{h\in I_j} c_h \beta_h)}\right)\left(\sum_{j=1}^{k}\sqrt{(\sum_{h\in I_j} W_h^2 S_h^2 / \beta_h)(\sum_{h\in I_j} c_h \beta_h)}\right),$$

$$j = 1,2,\ldots,k \tag{3.48}$$

The values of the sample sizes $n_h$ for the strata belonging to a particular subgroup say $G_q$, that is for $h \in I_q$ can be obtained by substituting the value of $\alpha_j$ given by (3.48) for $j=p$ in (3.17), where $q \in \{1,2,\ldots,k\}$

The resulting cost is

$$C_{(mixed)} = \sum_{j=1}^{k}\sum_{h\in I_j} \alpha_j c_h \beta_h$$

$$= \frac{1}{v}\sum_{j=1}^{k}\left[\sum_{h\in I_j} c_h \beta_h \left(\sqrt{\sum_{h\in I_j} (W_h^2 S_h^2 / \beta_h)/(\sum_{h\in I_j} c_h \beta_h)}\right) \times \left(\sqrt{(\sum_{h\in I_j} W_h^2 S_h^2 / \beta_h)(\sum_{h\in I_j} c_h \beta_h)}\right)\right]$$

$$= \frac{\left[ \sum_{j=1}^{k} \sqrt{(\sum_{h \in I_j} W_h^2 S_h^2 / \beta_h)(\sum_{h \in I_j} c_h \beta_h)} \right]^2}{v} \qquad (3.49)$$

The results obtained in this section can also be obtained by using Cauchy- Schwarz inequality.

## 3.4 THE INEFFICIENCY OF THE MIXED ALLOCATION

It is well known that the optimum allocation given by (3.5) is the most efficient allocation for fixed cost. But there are certain limitations to the use of optimum allocation in practice. The most severe of all the limitations is the absence of the knowledge of strata variances $S_h^2$. In such situations in the formula (3.5) $S_h^2$ may be replaced by its sample estimate

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$$

The values of the sample allocations in this case will be

$$\hat{n}_h = \frac{C_o (W_h s_h / \sqrt{c_h})}{\sum_{h=1}^{L} W_h s_h \sqrt{c_h}} ; h = 1,2,\ldots, L$$

where $\hat{n}_h$ are called the modified optimum allocation. Unfortunately, in general there is no guarantee that this modified optimum allocation is really optimum. At times it proved to be less efficient than a proportional allocation. So that even if an estimate of $S_h^2$ is available it is not always advisable to use the modified optimum allocation (see Sukhatme et al (1984)). As an alternative the use of the proposed mixed allocation is advised.

The relative efficiency (R.E.) of the optimum allocation for fixed cost as compared to the corresponding mixed allocation is given by

$$(R.E.)_{opt} = \frac{V_{mixed} - V_{opt}}{V_{opt}} \tag{3.50}$$

where $(R.E.)_{opt}$, stands for the relative efficiency of optimum allocation as compared to the mixed allocation.

The quantity on the R.H.S. of (3.50) can also be called the inefficiency of the mixed allocation as compared to the optimum allocation.

Thus $(R.I.E.)_{mixed} = \dfrac{V_{mixed} - V_{opt}}{V_{opt}}$ (3.51)

where $(R.I.E.)_{mixed}$, stands for the relative inefficiency of the mixed allocation as compared to the optimum allocation.

In the expression (3.50) and (3.51) $V_{mixed}$ is given by (3.42) and $V_{opt}$ (ignoring fpc) is given by

$$V_{opt} = \frac{(\sum_{h=1}^{L} W_h S_h \sqrt{c_h})^2}{C_o} \tag{3.52}$$

The contribution towards the total relative inefficiency of a particular allocation applied to the subgroup $G_p$ can be assessed by the term

$$\frac{\left(\sum_{h \in I_p} \frac{W_h^2 S_h^2}{\alpha_j \beta_h} - \sum_{h \in I_p} \frac{W_h^2 S_h^2}{n_{h(opt)}}\right)}{V_{opt}} \tag{3.53}$$

in the RHS of (3.51), where $n_{h(opt)}$ denote the sample sizes under optimum allocation for $h \in I_p$.

The expression (3.53) will help in deciding whether to use any particular allocation in a particular subgroup or not. If a particular allocation results in a large contribution towards the total relative inefficiency when applied to a particular, subgroup of strata, then the reasons for applying it may be reviewed.

## 3.5 A NUMERICAL ILLUSTRATION

In stratification with seven strata the values of $N_h, s_h$ and $c_h$ are given in Table 3.1. It assumed that the total available budget of the survey in C = 4500 units which includes an overhead cost $c_0$ = 500 units. The data are artificially constructed to illustrate the use mixed allocation.

We have $C_0 = C - c_o = 4500 - 500 = 4000$ units of cost available for measurements.

In Table 3.1 the strata are so arranged that

(i) Strata 1,2 & 3 constitute sub group $G_1$ in which equal allocation is to be used.

(ii) Strata 4 & 5 constitute subgroup $G_2$ in which proportional allocation is to used.

(iii) Strata 6 & 7 constitute subgroup $G_3$ in which optimum allocation is to be used.

**Table 3.1**

Values of $N_h, s_h$ and $c_h$ for seven strata

| Stratum No h | Stratum size $N_h$ | Stratum S.D. $s_h$ | Per units of measurement $c_h$ |
|---|---|---|---|
| 1 | 472 | 5.237 | 6 |
| 2 | 559 | 5.821 | 8 |
| 3 | 425 | 5.238 | 7 |
| 4 | 218 | 25.528 | 12 |
| 5 | 233 | 22.232 | 11 |
| 6 | 328 | 15.129 | 10 |
| 7 | 265 | 40.125 | 15 |

Thus, $I_1 = \{1,2,3\}$

$I_2 = \{4,5\}$

$I_3 = \{6,7\}$

It can be see that $I_j; j = 1,2,3$ are mutually exclusive and exhaustive as

$I_1 \cap I_2 = I_1 \cap I_3 = I_2 \cap I_3 = \phi$ and $\bigcup_{j=1}^{3} I_j = \{1,2,3,4,5,6,7\}$

The reason for using equal allocation to strata 1,2 and 3 is that these strata are relatively more homogeneous as compared to other strata since their corresponding estimated strata S.D.($s_h$) are small. Proportional allocation is used in strata 4 and 5 because they have relatively smaller size ($N_h$) among the remaining four strata. The above set up will help in reducing the variance $V(\bar{y}_{st})$ under mixed allocation.

**Table 3.2**

Sample sizes under over all optimum allocation

| $h$ | $W_h$ | $s_h$ | $c_h$ | $W_h s_h$ | $W_h s_h / \sqrt{c_h}$ | $W_h s_h \sqrt{c_h}$ | $n_{h(opt)}$ (rounded) |
|---|---|---|---|---|---|---|---|
| 1 | 0.189 | 5.237 | 6 | 0.990 | 0.404 | 2.425 | 35 |
| 2 | 0.224 | 5.821 | 8 | 1.304 | 0.461 | 3.688 | 40 |
| 3 | 0.170 | 5.238 | 7 | 0.890 | 0.336 | 2.355 | 29 |
| 4 | 0.087 | 25.528 | 12 | 2.221 | 0.641 | 7.694 | 56 |
| 5 | 0.093 | 22.232 | 11 | 2.067 | 0.623 | 6.855 | 55 |
| 6 | 0.131 | 15.129 | 10 | 1.982 | 0.627 | 6.268 | 55 |
| 7 | 0.106 | 40.125 | 15 | 4.253 | 1.098 | 16.472 | 96 |
| | | | | | $\sum$ | 45.757 | |

Table 3.2 gives the sample sizes when as gives overall optimum allocation is used.

The estimated variance $v(\bar{y}_{st})$ under optimum allocation ignoring fpc is

$$v_{opt} = \frac{\left(\sum_{h=1}^{L} W_h s_h \sqrt{c_h}\right)^2}{C_o}$$

$$= \frac{(45.757)^2}{4000}$$

$$= 0.5234 \tag{3.54}$$

The application of the mixed allocation to the various subgroups of the strata according to the given scheme may be characterized by letting

$$n_{h(m)} = \alpha_j \beta_h ; j = 1,2,3; h \in I_j \tag{3.55}$$

where $n_{h(m)}; h = 1,2,...,L$ denote the sample sizes under mixed allocation and $\beta_h$ are defined as below.

In subgroup $G_1$ for applying equal allocation

$\beta_h = 1$ for $h \in I_1 = \{1,2,3\}$

In subgroup $G_2$ for applying proportional allocation

$\beta_h = 1$ for $h \in I_2 = \{4,5\}$

In subgroup $G_3$ for applying optimum allocation

$\beta_h = \left(W_h s_h / \sqrt{c_h}\right)$ for $h \in I_3 = \{6,7\}$

In Table 3.3 the values of $W_h^2 s_h^2 / \beta_h$ and $c_h \beta_h$ are tabulated. These values are to be used in the calculation of $\alpha_j ; j = 1,2,3$.

**Table 3.3**

Values of $W_h^2 s_h^2 / \beta_h$ and $c_h \beta_h$

| $h$ | $W_h s_h$ | $W_h^2 s_h^2$ | $c_h$ | $\beta_h$ | $W_h^2 s_h^2 / \beta_h$ | $c_h \beta_h$ |
|---|---|---|---|---|---|---|
| 1 | 0.990 | 0.980 | 6 | 1 | 0.980 | 6 |
| 2 | 1.304 | 1.700 | 8 | 1 | 1.700 | 8 |
| 3 | 0.890 | 0.792 | 7 | 1 | 0.792 | 7 |
| Subtotal for $h \in I_1$ | | | | | 3.472 | 21.000 |
| 4 | 2.221 | 4.933 | 12 | 0.087 | 56.701 | 1.044 |
| 5 | 2.067 | 4.272 | 11 | 0.093 | 45.935 | 1.023 |
| Sub total for $h \in I_2$ | | | | | 102.636 | 2.067 |
| 6 | 1.982 | 3.928 | 10 | 0.627 | 6.270 | 6.270 |
| 7 | 4.253 | 18.088 | 15 | 1.098 | 16.470 | 16.470 |
| Subtotal for $h \in I_3$ | | | | | 22.740 | 22.740 |

Table 3.4 gives the values of $\alpha_j$; $j = 1,2,3$

**Table 3.4**

Calculation of $\alpha_J$

| Subgroup No. $J$ | $(A)$ $\sum_{h \in I_J} W_h^2 s_h^2 / \beta_h$ | $(B)$ $\sum_{h \in I_J} c_h \beta_h$ | $(C)$ $\sqrt{(A)/(B)}$ | $(D)$ $\sqrt{(A)(B)}$ | $\alpha_J = 4000 \times \frac{(C)}{\sum(D)}$ |
|---|---|---|---|---|---|
| 1 | 3.472 | 21.000 | 0.407 | 8.539 | 35.512 |
| 2 | 102.636 | 2.067 | 7.047 | 14.565 | 614.868 |
| 3 | 22.740 | 22.740 | 1 | 22.740 | 87.252 |
| | | | $\sum(D)$ | 45.844 | |

The values of $\beta_h$ from Table 3.3 and values of $\alpha_J$ from Table 3.4 when substituted in the formula (3.55) gives the mixed allocation $n_{h(m)}; h = 1,2,...,7$ as

For $J = 1$, that is $h \in I_1 = \{1,2,3\}$

$n_{1(m)} = \alpha_1 \beta_1 = 35.512 \times 1 = 35.512 \cong 35$

$n_{2(m)} = \alpha_1 \beta_2 = 35.512 \times 1 = 35.512 \cong 35$ 35

$n_{3(m)} = \alpha_1 \beta_3 = 35.512 \times 1 = 35.512 \cong 35$ 35

For $J = 2$, that is $h \in I_2 = \{4,5\}$

$n_{4(m)} = \alpha_2 \beta_4 = 614.868 \times 0.087 = 53.493 \cong 54$

$n_{5(m)} = \alpha_2 \beta_5 = 614.868 \times 0.093 = 57.183 \cong 57$

For $J = 3$, that in $h \in I_3 = \{6,7\}$

$n_{6(m)} = \alpha_3 \beta_6 = 87.252 \times 0.627 = 54.707 \cong 55$

$$n_{7(m)} = \alpha_3 \beta_7 = 87.252 \times 1.098 = 95.803 \cong 96$$

The estimated variance $v(\bar{y}_{st})$ (ignoring fpc) under mixed allocation is given as

$$v_{mixed} = \sum_{h=1}^{7} \frac{W_h^2 s_h^2}{n_{h(m)}} = 0.5356$$

## 3.6 CONCLUSION

The estimated relative inefficiency of the mixed allocation as compared to the overall optimum allocation is given by (3.51) is

$$(R.I.E)_{mixed} = \frac{v_{mixed} - v_{opt}}{v_{opt}} \times 100\%$$

$$= \frac{0.5253 - 0.5234}{0.5234} \times 100\%$$

$$= 0.363\,\%$$

The estimated relative inefficiency of equal and proportional allocations are given as

$$(R.I.E.)_{equal} = \frac{\sum_{h \in I_1} \frac{W_h^2 s_h^2}{n_{h(m)}} - \sum_{h \in I_1} \frac{W_h^2 s_h^2}{n_{h(m)}}}{v_{opt}} \times 100\%$$

$$= \frac{0.0992 - 0.0978}{0.5234} \times 100\%$$

$$= 0.267\,\%$$

$$(R.I.E.)_{prop} = \frac{\sum_{h \in I_1} \frac{W_h^2 s_h^2}{n_{h(m)}} - \sum_{h \in I_1} \frac{W_h^2 s_h^2}{n_{h(m)}}}{v_{opt}} \times 100\%$$

$$= \frac{0.1663 - 0.1658}{0.5234} \times 100\%$$

$$= 0.096\,\% \quad \text{respectively.}$$

It can be seen that

$(R.I.E)_{mixed} = (R.I.E)_{equal} + (R.I.E)_{prop}$.

Since the *R.I.E.* in using mixed allocation is only $0.363\%$ we conclude that we used mixed allocation safely instead of overall optimum allocation.

# CHAPTER-IV

# DOUBLE SAMPLING FOR STRATIFICATION WITH SUBSAMPLING THE NONRESPONDENTS: A DYNAMIC PROGRAMMING APPROACH

# CHAPTER IV

## DOUBLE SAMPLING FOR STRATIFICATION WITH SUBSAMPLING THE NONRESPONDENTS: A DYNAMIC PROGRAMMING APPROACH

### 4.1 INTRODUCTION

In stratified sampling, the population is divided into $L$ strata which are homogeneous within themselves and whose means are widely different The stratum weights are used in estimating unbiasedly the mean or the total of the character under study

If these weights are not known, the technique of double sampling can be used, which consists of selecting a preliminary sample of size $n'$ by simple random sampling, without replacement (SRSWOR), to estimate the stratum weights and then selecting the subsample of $n$ units with $n_h$ units from the h-th stratum, to collect information on the characteristic under study, such that $\sum_{h=1}^{L} n_h = n$

Rao (1973) proposed the method of double sampling for stratification (DSS) for the estimation of the population mean $\overline{Y}$, of the variate $y$, using the values of the auxiliary variate collected at the first phase for stratification only

Ige and Tripathi (1987) used the information collected at the first phase for stratification as well as in constructing ratio and difference estimators of the population mean $\overline{Y}$

One of the sources of error in surveys is non-contact or refusals In a household survey the selected family may not be available at home when the interviewer calls The selected person may refuse to cooperate, saying that he has not time to answer question or that he consider the purpose of the survey to be senseless Persuasion and further recalls are therefore necessary for achieving complete coverage of the sample But it is expensive to call and call again At the same time we cannot afford to neglect the non-response Results based on response alone will not apply to the entire population from which the sample was selected Experience from different surveys show that non-response generally differs from the response in several respects and neglecting them will introduce a bias in the results Under these circumstances, one solution is to take a small subsample of the non-respondents and use all the persuasion, ingenuity and other resources at our command to get a response from them The two samples can then be combined suitably to get a better estimate of the population parameter

Hansen and Hurwitz (1946) discussed a method of tackling total non-response in mail interviews Rao (1986) applied this method of subsampling the non-respondents for the ratio estimation of the mean when the population mean of the auxiliary character is known

Using an auxiliary variable Okafor (1994) derived the DSS estimator based on the subsampling of the non-respondents, when there is total response on the auxiliary variable and incomplete response on the main character

For practical application of any allocation integer values of the sample sizes are required This could be done by simply rounding off noninteger sample sizes to their nearest integral values When the sample sizes are large enough and (or) the measurement cost in various strata are not too high, the rounded off sample allocation may work well

However in situations other than described above the rounded off sample allocations may become infeasible and (or) non optimal This means that the rounded off values may violate some of the constraint of the problem and (or) there may exist other sets of integer sample allocations with a better value of the objective function of the formulated NLPP In such situations we have to use some integer programming technique to obtain an optimum integer solution In this chapter the problem of obtaining an optimum allocation in DSS, when there is incomplete response on the main character and total response on the auxiliary character, is considered as an all integer nonlinear programming problem (AINLPP) A solution procedure is developed using the dynamic programming technique. A numerical example is also presented to illustrate the computational details.

## 4.2 THE PROBLEM

From a population of $N$ units a large sample of size $n'$ is selected by simple random sampling without replacement (SRSWOR). Information on the auxiliary variable $x$ is collected with which an unbiased estimate $w_h = n'_h / n'$ of the true stratum weight $W_h = N_h / N$ is computed.

where $n'_h$ is the number of units in the initial sample that falls in stratum $h$, $(h = 1, 2, \ldots, L)$, with $\sum_{h=1}^{L} n'_h = n'$

In each stratum a subsample of size $n_h = v_h \, n'_h$, $(0 < v_h < 1)$, $v_h$ is prefixed, is selected from $n'_h$ by SRSWOR. The main character $y$ is then observed on these $n_h$ units, $h = 1, 2, \ldots, L$.

The DSS estimator of the population mean for the total response is

$$\bar{y}_{ds} = \sum_{h=1}^{L} w_h \, \bar{y}_h \tag{4.1}$$

where $\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$ , sample mean

The variance of $\bar{y}_{ds}$ is

$$V(\bar{y}_{ds}) = \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^2 + \frac{1}{n'} \sum_{h=1}^{L} W_h \left(\frac{1}{v_h} - 1\right) S_y^2 \tag{4.2}$$

where $S_y^2 = \frac{1}{N-1} \sum_{h=1}^{N} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y})^2$

and $S_{yh}^2 = \frac{1}{N-1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2$ , variance of $y$ in h-th stratum.

Let

$n_{1h}$: unit respond at the first call from the $n_h$ units selected in stratum $h$.

$n_{2h}$: units do not respond.

Thus the subsample of size $n_h$ is again subdivided into respondent and non-respondent subsamples of sizes $n_{1h}$ and $n_{2h}$ respectively, where $n_{1h} + n_{2h} = n_h$. A subsample of size $m_{2h}$ out of the $n_{2h}$ non-respondents of h-th stratum is selected and interviewed with improved methods, where $m_{2h} = k_h^* n_{2h}$ $(0 < k_h^* < 1)$, $k_h^*$ is prefixed.

An unbiased, estimator $\bar{y}_{ds}^*$ for $\bar{Y}$ based on the sample means from respondents and non-respondents (in second attempt) is given as

$$\bar{y}_{ds}^* = \sum_{h=1}^{L} w_h \bar{y}_h^*, \text{ where } \bar{y}_h^* = \frac{n_{1h}\bar{y}_{1h} + n_{2h}\bar{y}_{m_{2h}}}{n_h} \tag{4.3}$$

$\bar{y}_{1h}$ = sample mean for respondents based on $n_{1h}$ units

$\bar{y}_{m_{2h}}$ = sample mean for the non-respondents based on $m_{2h}$ units

The variance of $\bar{y}_{ds}^{*}$ is

$$V(\bar{y}_{ds}^{*}) = V(\bar{y}_{ds}) + \frac{1}{n'} \sum_{h=1}^{L} W_{2h} \frac{1-k_h^{*}}{k_h^{*} \nu_h} S_{2yh}^{2} \tag{4.4}$$

$W_{2h} = N_{2h} / N$, population proportion of the non-respondents in stratum $h$.

$S_{2yh}^{2}$, is the population variance among the non-respondents in stratum $h$.

(see Hansen and Hurwitz (1946) and Rao (1986)).

The problem now is to find the optimum sizes of the subsamples $m_{2h}$, $h = 1, 2, \ldots, L$ for which $V(\bar{y}_{ds}^{*})$ given by (4.4) is minimum for a fixed cost. This problem may be divided into two phases.

**Phase I**: In this phase the optimum values of $n_h$, $h = 1, 2, \ldots, L$ are obtained for which $V(\bar{y}_{ds}^{*})$ is minimum for a fixed sample size $n = \sum_{h=1}^{L} n_h$.

**Phase II**: In this phase the optimum values of $m_{2h}$; $h = 1, 2, \ldots, L$ are obtained for a fixed total cost of the survey.

## 4.3 FORMULATION OF THE PHASE-I PROBLEM

Using (4.2) and (4.4) the problem of first phase can be formulated as

Minimize $V(\bar{y}_{ds}^{*}) = \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^2 + \frac{1}{n'} \sum_{h=1}^{L} W_h \left(\frac{1}{\nu_h} - 1\right) S_{yh}^2$

$$+\frac{1}{n'}\sum_{h=1}^{L} W_{2h}\left(\frac{1-k_h^*}{k_h^* \nu_h}\right)S_{2yh}^2 \tag{4.5}$$

subject to $\sum_{h=1}^{L} n_h = n$ (4.6)

$$1 \le n_h \le N_h \tag{4.7}$$

and $n_h$ are integers; $h = 1,2,\ldots,L$ (4.8)

Ignoring the terms independent of $n_h$ the objective function in (4.5) can be expressed as

$$Z(n_1, n_2, \ldots, n_L) = \frac{1}{n'}\sum_{h=1}^{L}\left(\frac{W_h n'_h S_{yh}^2 + W_{2h}\{(1-k_h^*)/k_h^*\} n'_h S_{2yh}^2}{n_h}\right)$$

$$= \sum_{h=1}^{L}\frac{a_h}{n_h}$$

where $a_h = \left[\frac{W_h n'_h S_{yh}^2 + W_{2h}\{(1-k_h^*)/k_h^*\} n'_h S_{2yh}^2}{n'}\right]$ (4.9)

The problem (4.5)-(4.8) may be simplified as

Minimize $Z(n_1, n_2, \ldots, n_L) = \sum_{h=1}^{L}\frac{a_h}{n_h}$ (4.10)

subject to $\sum_{h=1}^{L} n_h = n$ (4.11)

$$1 \le n_h \le N_h \tag{4.12}$$

and $n_h$ are integers; $h = 1,2,\ldots,L$ (4.13)

The restriction (4.12) are imposed to avoid over sampling, that is, the situation where $n_h > N_h$ and to have the representation of every stratum in the sample.

## 4.4 SOLUTION OF THE PHASE-I PROBLEM

Ignoring restrictions in (4.12) and (4.13) and using Lagrangians multipliers technique, the optimum value of $n_h$ that minimize (4.10) subject to (4.11) may be obtained as given below.

$$\phi(n_h, \lambda) = \sum_{h=1}^{L}\left(\frac{a_h}{n_h}\right) + \lambda\left(\sum_{h=1}^{L} n_h - n\right)$$

differentiating $\phi$ partially w.r.t. $n_h$ and equating to zero we get

$$\frac{\partial \phi}{\partial n_h} = -\frac{a_h}{n_h^2} + \lambda = 0; \quad h = 1,2,\ldots,L$$

$$\text{or} \quad a_h = \lambda n_h^2; \quad h = 1,2,\ldots,L$$

$$n_h = \frac{\sqrt{a_h}}{\sqrt{\lambda}}; \quad h = 1,2,\ldots,L$$

Taking summation on both the sides we get

$$\sum_{h=1}^{L} n_h = \frac{1}{\sqrt{\lambda}}\sum_{h=1}^{L}\sqrt{a_h} \quad \text{or} \quad n = \frac{1}{\sqrt{\lambda}}\sum_{h=1}^{L}\sqrt{a_h}$$

or $\frac{1}{\sqrt{\lambda}} = \frac{n}{\sum_{h=1}^{L} \sqrt{a_h}}$

which gives

$$n_h = \frac{n\sqrt{a_h}}{\sum_{h=1}^{L} \sqrt{a_h}}; \quad h = 1,2,\ldots,L \tag{4.14}$$

If the above values of $n_h$ satisfies (4.12) also the non-linear programming problem (NLPP) (4.10)-(4.12) is solved.

In case either some or all of the $n_h$ given by (4.14) violates (4.12) or to get an integer solution restricted by (4.13) the Lagrange multipliers technique could not provide the solution and some other constrained optimization technique is to be used. In the next section a computational procedure to obtain integer values of $n_h$ is developed using dynamic programming technique.

## 4.5 SOLUTION OF THE PHASE-I PROBLEM USING THE DYNAMIC PROGRAMMING TECHNIQUE

The problem (4.10)-(4.13) can be restated as

$$\text{Minimize } Z(n_1, n_2, \ldots, n_L) = \frac{a_1}{n_1} + \frac{a_2}{n_2} + \ldots + \frac{a_L}{n_L} \tag{4.15}$$

$$\text{subject to } \quad n_1 + n_2 + \ldots + n_L = n \tag{4.16}$$

$$1 \le n_1 \le N_1, \ \ldots, 1 \le n_L \le N_L \tag{4.17}$$

and $n_h$ are integers; $h=1,2,\ldots,L$

(4.18) The objective function and the constraints of the AINLPP (4.15)-(4.18) are the sum of independent functions of $n_h$, $h=1,2,\ldots,L$

The AINLPP, which is an $L$-stage decision problem, can be decomposed into $L$-stage single variable decision problems.

In the following a solution procedure for solving the formulated AINLPP using dynamic programming technique is developed.

Consider the sub-problem called the k-th sub-problem, involving the first $(k<L)$ strata and let $f(k,r)$ be the minimum value of the objective function for the first $k$ strata with total sample size $r$, i.e.

$$f(k,r)=\min\sum_{h=1}^{k}\frac{a_h}{n_h} \tag{4.19}$$

$$\text{subject to } \sum_{h=1}^{k}n_h = r \tag{4.20}$$

$$1\le n_h \le N_h \tag{4.21}$$

$$\text{and} \quad n_h \text{ are integers,} \quad h=1,2,\ldots,k \tag{4.22}$$

Thus the problem (4.15)-(4.18) is equivalent to the problem of finding $f(L,n)$. $f(L,n)$ is found recursively by finding $f(k,r)$ for $k=1,2,\ldots,L$ and $r=0,1,\ldots,n$

$$\text{Now} \quad f(k,r)=\min\left(\frac{a_k}{n_k}+\sum_{h=1}^{k-1}\frac{a_h}{n_h}\right)$$

subject to $\sum_{h=1}^{k} n_h = r - n_k$

$$1 \le n_h \le N_h$$

and $n_h$ are integers, $h = 1,2,\ldots,k$

For fixed integer value of $n_k$, $1 \le n_k \le \min[r, N_k]$, $f(k,r)$ is given by

$$f(k,r) = \frac{a_k}{n_k} + \left\{ \min \sum_{h=1}^{k-1} \frac{a_k}{n_k} \middle| \sum_{h=1}^{k-1} n_h = r - n_k , \right.$$

$$\left. 1 \le n_h \le N_h ;\ n_h \text{ are integers, where } h = 1,2,\ldots,k-1 \right\}$$

But by definition the terms in {} above is equal to $f(k-1, r-n_k)$.

Suppose we assume that for a given $k$, $f(k-1,r)$ is known for all possible $r = 0,1,\ldots,n$. Then

$$f(k,r) = \min_{n_k = 1,2,\ldots,n} \left[ \frac{a_k}{n_k} + f(k-1, r-n_k) \right] \tag{4.23}$$

This is the required dynamic programming recursive formula. Using the relation (4.23) for each $k = 1,2,\ldots,L$ and $r = 0,1,\ldots,n$, $f(L,n)$ can be calculated.

Initially we set $f(k,r) = \infty$, if $r < k$ since we wish to have $n_h \ge 1$, for each $h = 1,2,\ldots,k$, $r$ must be at least equal to $k$.

Also $f(1,r) = \min[a_1 / n_1$, subject to $n_1 = r,\ 1 \le n_1 \le N_1]$

Thus $f(1,r)=\begin{cases}\infty & for \quad r>N_1 \quad or \quad r<1\\ a_1/r & for \quad 1\le r\le N_1\end{cases}$

We tabulate the value of $f(k,r)$ and the optimal $n_k$, for each $k$, systematically. Then from $f(L,n)$, optimal $n_L$ can be found; from $f(L-1,n-n_L)$ optimal $n_{L-1}$ can be found and so on until finally we find optimal $n_1$.(see Arthenari and Dodge (1981)).

## 4.6 FORMULATION OF THE PHASE-II PROBLEM

For the second phase of the solution consider the variance function given in (4.5)

$$V(\bar{y}_{ds}^{*})=\left(\frac{1}{n'}-\frac{1}{N}\right)S_y^2+\frac{1}{n'}\sum_{h=1}^{L}W_h\left(\frac{1}{v_h}-1\right)S_{yh}^2$$

$$+\frac{1}{n'}\sum_{h=1}^{L}W_{2h}\left(\frac{1-k_h^{*}}{k_h^{*}v_h}\right)S_{2yh}^2 \qquad (4.24)$$

Assuming the cost function [see Okafor (1994)]

$$C=C_1n'+\sum_h C_{2h}\,n_h+\sum_k C_{21h}\,n_{1h}+\sum_h C_{22h}\,n_{2h}\,k_h^{*}$$

where

$C_1$ : cost of getting information on the first phase sample.

$C_{2h}$ : cost of first attempt on the main character in stratum $h$.

$C_{21h}$: cost of processing the results on the main character from the respondent at the first attempt sample in the stratum $h$.

$C_{22h}$: cost of getting and processing results on the main character from the sub sample of the non-respondents at the second phase sample in stratum $h$.

We also must have $1 \le m_{2h} \le n_{2h}$

Ignoring the terms independent of $m_{2h}$ in the R.H.S. of (4.24) and putting $k_h^* = m_{2h} / n_{2h}$ and $v_h = n_h / n'_h$ .

The problem becomes

$$\text{Minimize}\, Z(m_{21}, m_{22}, \ldots, m_{2L}) = \frac{1}{n'} \sum_{h=1}^{L} W_{2h} \left( \frac{n_{2h}}{m_{2h}} \right) \frac{n'_h}{n_h} S_{2yh}^2 \tag{4.25}$$

$$\text{subject to} \sum_{h=1}^{L} C_{22h}\, m_{2h} \le C_0 \tag{4.26}$$

$$\text{and} \quad 1 \le m_{2h} \le n_{2h} \tag{4.27}$$

$m_{2h}$ are integers, $h = 1, 2, \ldots, L$.

And $C_0 = C_1 n' + \sum_h C_{2h} n_h + \sum_h C_{21h} n_{1h}$

Let

$$b_h = \frac{1}{n'} W_{2h} n_{2h} \frac{n'_h}{n_h} S_{2yh}^2 \tag{4.28}$$

The AINLPP (4.25)-(4.27) may be restated as

Minimize $Z(m_{21}, m_{22}, \ldots, m_{2L}) = \sum_{h=1}^{L} \frac{b_h}{m_{2h}}$

(4.29)

subject to $\sum_{h=1}^{L} C_{22h} m_{2h} \le C_0$ (4.30)

and $1 \le m_{2h} \le n_{2h}$ (4.31)

where $m_{2h}$ are integers, $h = 1, 2, \ldots, L$ (4.32)

## 4.7 SOLUTION OF THE PHASE-II PROBLEM

Like phase-I applying Lagrangian multipliers technique, with equality in (4.30) and ignoring (4.31) and (4.32) we get,

$$\phi(m_{2h}, \lambda) = \sum_{h=1}^{L} \frac{b_h}{m_{2h}} + \lambda\left(\sum_{h=1}^{L} C_{22h} m_{2h} - C_0\right)$$

Differentiating $\phi$ with respect to $m_{2h}$ and $\lambda$ and equating to zero we get

$$\frac{\partial \phi}{\partial m_{2h}} = -\frac{b_h}{m_{2h}^2} + \lambda C_{22h} = 0$$

$$\frac{\partial \phi}{\partial \lambda} = \sum_{h} C_{22h} m_{2h} - C_0 = 0$$

Solving the above equations we get the optimum value of $m_{2h}$ as

$$m_{2h} = C_0 \frac{b_h / C_{22h}}{\sum_{h=1}^{L} \sqrt{b_h} C_{22h}} \qquad (4.33)$$

## 4.8 SOLUTION OF THE PHASE-II PROBLEM USING THE DYNAMIC PROGRAMMING TECHNIQUE

Let $f(k,r)$ be the minimum value of the objective function of the problem (4.29)-(4.32), the first $k$ strata with $C_0 = r$ i.e.

$$f(k,r) = \left\{ \min \sum_{h=1}^{k} \frac{b_h}{m_{2h}} \,\middle|\, \sum_{h=1}^{k} C_{22h} m_{2h} \le r \right.$$

$$\left. 1 \le m_{2h} \le n_{2h},\ m_{2h} \text{ are integer, } h = 1,2,\ldots,k \right\} \qquad (4.34)$$

with the above definition of $f(k,r)$ the problem is equivalent to the problem of finding $f(L,C_0)$. $f(L,C_0)$ is found recursively by using (4.34) for $k = 1,2,\ldots,L$ and $r = 0,1,\ldots,C_0$.

Now $f(k,r) = \min \left( \frac{b_k}{m_{2k}} + \sum_{h=1}^{k-1} \frac{b_h}{m_{2h}} \right)$

subject to $\sum_{h=1}^{k-1} C_{22h} m_{2h} = r - C_{22k} m_{2k}$

and $1 \le m_{2h} \le n_{2h}$,

where $m_{2h}$ are integers, $h = 1,2,\ldots,k-1$.

$$f(k,r)=\left\{\begin{array}{l}\min\left(\frac{b_k}{m_{2k}}+\sum_{h=1}^{k-1}\frac{b_h}{m_{2h}}\right)\left|\ \sum_{h=1}^{k-1}C_{22}m_{2h}=r-C_{22k}m_{2k}\right. \\ and..\ 1\le m_{2h}\le n_{2h}, m_{2h}\ are\ integers\ h=1,2,...,k-1\end{array}\right\}$$

and $\quad 1\le m_{2h}\le n_{2h},\ m_{2h}$ are integers $h=1,2,\ldots,k-1\}$

For a fixed integer value of $m_{2k}$, $1\le m_{2k}\le[r,n_{2k}], f(k,r)$ , is given by

or $\{\ \min\left(\frac{b_k}{m_{2k}}+\sum_{h=1}^{k-1}\frac{b_h}{m_{2h}}\right)|\sum_{h=1}^{k-1}C_{22}m_{2h}=r-C_{22k}m_{2k}$

$1\le m_{2h}\le n_{2h}$, and $m_{2h}$are integers $h=1,2,\ldots,k-1\}$ (4.35)

By definition the terms in the braces is equivalent to $f(k-1,r)$ is known for all possible $r=0,1,\ldots,C_0$. Then

$$f(k,r)=\min_{m_{2k}=1,2,\ldots,C_0}\left[\frac{b_k}{m_{2k}}+f(k-1,r-C_{22h}m_{2k})\right] \quad (4.36)$$

Using the relation (4.36) for each $k=1,2,\ldots,L$ and $r=0,1,\ldots,C_0$, $f(L,C_0)$ can be calculated. Initially we set $f(k,r)=\infty$, if $r\le k$. Since we wish to have $m_{2h}\ge 1$ for each $h=1,2,\ldots,k; r$ must be at least equal to $k$.

Also $\quad f(1,r)=\min[b_1/m_{21}\quad$ Subject to $m_{21}=r,\ 1\le m_{21}\le n_{21}]$

Thus $\quad f(1,r)=\begin{cases}\infty & for\quad r>n_{21}\quad or\quad r<1\\ b_1/r & for\quad 1\le r\le n_{21}\end{cases}$

We tabulate the value of $f(k,r)$ and the optimal $m_{2k}$, for each $k$, systematically. Then from $f(L,C_0)$ optimal $m_{2L}$ can be found from

$f(L-1, C_0 - m_{2L})$. Optimal $m_{2L-1}$ can be found and so on, until finally we find optimal $m_{21}$.

## 4.9 NUMERICAL EXAMPLE

The following numerical example demonstrates the use of the solution procedure. The data used in this example is from Murthy (1967). Here DSS is used to estimate the mean area under cultivation. The area of each village and the area cultivated in the village are converted to hectares and grouped into three strata. Within each stratum, the population was again subdivided into respondent and non-respondent groups. Villages with larger area considered in non-respondent group.

Table 4.1 and 4.2 gives the population parameters obtained from the data as given in Okafor (1994).

**Table 4.1**

Overall stratum population parameters

| Stratum | $W_h$ | $S_{yh}^2$ | $v_h$ | $k_h^*$ |
|---|---|---|---|---|
| 0-930 | 0.336 | 39974.81 | 0.4 | 0.5 |
| 931-1700 | 0.352 | 61455.48 | 0.5 | 0.6 |
| 1701-4300 | 0.313 | 172425.05 | 0.6 | 0.7 |

It is assumed that $N = 200$, $n' = 100$, $n = 50$

Using proportional allocation $n'_h$ may be obtained as

$n'_1 = 33.6 \cong 34$, $n'_2 = 35.2 \cong 35$, and $n'_3 = 31.3 \cong 31$

**Table 4.2**

Class stratum population parameters

| Stratum | Class | $S^2_{yh}$ | $W_h$ |
|---|---|---|---|
| 0-930 | Respondent | 7162.51 | 0.188 |
| | Non Respondent | 14549.99 | 0.148 |
| 931-1700 | Respondent | 19564.45 | 0.219 |
| | Non Respondent | 17386.54 | 0.133 |
| 1701-4300 | Respondent | 5042.50 | 0.188 |
| | Non Respondent | 71175.11 | 0.125 |

For L=3 the Phase-I problem (4.15)-(4.18) can be expressed as

$$\text{Minimize } Z = \frac{a_1}{n_1} + \frac{a_2}{n_2} + \frac{a_3}{n_3} \tag{4.37}$$

$$\text{subject to } n_1 + n_2 + n_3 = 50 \tag{4.38}$$

$$\left.\begin{aligned} 1 \le n_1 \le 34 \\ 1 \le n_2 \le 35 \\ 1 \le n_3 \le 31 \end{aligned}\right\} \tag{4.39}$$

where $n_h$ are integers; $h = 1, 2, 3$ (4.40)

Table (4 3) gives the optimum values of $n_h$ using formula (4 14) These values of $n_h$ satisfy (4 39) also, hence they will solve NLPP (4 37)-(4 40) completely

**Table 4.3**

Calculation of $n_h$ using formula (4 14)

| $h$ | $a_h$ | $\sqrt{a_h}$ | $n\sqrt{a_h}$ | $n_h$ |
|---|---|---|---|---|
| 1 | 5236 5381 | 72 363928 | 3618 1964 | 12 176312 $\cong$ 12 |
| 2 | 8157 2253 | 90 317359 | 4515 868 | 15 197245 $\cong$ 15 |
| 3 | 18085 764 | 134 48332 | 6724 1661 | 22 627748 $\cong$ 23 |
| | $\sum\sqrt{a_h}$ | = 297 16461 | | |

The optimal value of the objective function is $Z^* = 1766\,09$

For the sake of illustration, the dynamic programming approach to find the integer optimum allocation in Phase-I is also applied to the same problem Execution of the computer program (in C language, given in Appendix-I) for the procedure given in Section 4 5 for solving the AINLPP (4 19)-(4 22) gives the following solution to the Phase I problem

$n_1 = 12, \quad n_2 = 15, \quad n_3 = 23$

The corresponding value of the objective function is $Z^* = 1766\,5308$

It can be seen that this solution is same as given in Table 4 3 except for a negligible change in the value of the objective function

For formulating the Phase-II problem, let $C_{22h} = 10, 12, 8$ for $h = 1, 2, 3$ respectively and $C_0 = 100$

Since $W_{1h}$ and $W_{2h}$ are known for $h = 1, 2, 3$ they are used to work out the expected values of $n_{2h}$, $h = 1, 2, 3$ as $n_{2h} = n_h W_{2h} / (W_{1h} + W_{2h})$ These values are tabulated in Table 4 4

**Table 4.4**

Calculation of $n_{2h}$

| $h$ | $W_{1h}$ | $W_{2h}$ | $n'_h$ | $n_h$ | $S^2_{2yh}$ | $C_{22h}$ | $n_{2h}$ |
|---|---|---|---|---|---|---|---|
| 1 | 0 188 | 0 148 | 33 60 | 12 | 14549 99 | 10 | 5 2857 $\cong$ 5 |
| 2 | 0 219 | 0 133 | 35 20 | 15 | 17386 54 | 12 | 5 6676 $\cong$ 6 |
| 3 | 0 188 | 0 125 | 31 20 | 23 | 71175 11 | 8 | 9 1853 $\cong$ 9 |

For $L = 3$, the Phase-II problem as given in (4 29) to (4 32) is

$$\text{Minimize } Z = \frac{b_1}{m_{21}} + \frac{b_2}{m_{22}} + \frac{b_3}{m_{23}} \tag{4 41}$$

$$\text{subject to } C_{221} m_{21} + C_{222} m_{22} + C_{223} m_{23} \le C_0 \tag{4 42}$$

$$\left.\begin{array}{l} 1 \le m_{21} \le n_{21} \\ 1 \le m_{22} \le n_{22} \\ 1 \le m_{23} \le n_{23} \end{array}\right\} \tag{4 43}$$

where $m_{2h}$ are integers, $h = 1, 2, 3$ (4 44)

Table 4 5 gives the optimum values of $m_{2h}$ using formula (4 35) These values of $m_{2h}$ are infeasible, since they violate the restriction $\sum_{h=1}^{3} C_{22h} m_{2h} \le C_0$ in (4 37), hence as an alternative, the dynamic programming approach given in Section 4 8 may be used

**Table 4.5**

Calculation of $m_{2h}$ using formula (4 35)

| $h$ | $b_h$ | $\sqrt{b_h}$ | $\sqrt{b_h / C_{22h}}$ | $\sqrt{b_h C_{22h}}$ | $m_{2h}$ |
|---|---|---|---|---|---|
| 1 | 318 70212 | 17 85223 | 5 6453708 | 56 453708 | 2 6707656 $\cong$ 3 |
| 2 | 307 54977 | 17 537097 | 5 0625238 | 60 750286 | 2 3950268 $\cong$ 2 |
| 3 | 1108 5576 | 33 295009 | 11 771563 | 94 172506 | 5 5690027 $\cong$ 6 |
| | | | $\sum \sqrt{b_h C_{22h}}$ | = 211 3765 | |

Execution of the computer program (in C language, given in Appendix–II of this chapter) for the procedure developed in Section 4 8 for solving the AINLPP (4 29)-(4 32) gives the following results

$m_{21} = 3, \quad m_{22} = 2, \quad m_{23} = 5$

The optimum value of the objective function (4 41) is $Z^* = 481\ 72045$

# APPENDIX-I

```
#include<stdio.h>
#define K_MAX 3
#define R_MAX 50
#define INF 9999999.0

main( )
    {
    int l,n[4][51],k,r,i,m,nk;
    float f[4][51],min;
    float a[4]={1,5236.5381,8157.2253,18085.764};
    float Nk[4]={1,34,35,31};
   FILE *op;
   op=fopen("result1.dat","w+");
   f[0][0]=0.0;
   f[1][0]=INF;
   f[2][0]=INF;
   f[3][0]=INF;
   l=0;
/*Initialization of zero point functions */
   for(i=1;i<=50;i++)
   f[1][i]=INF;
/*Starting with k */
   for(k=1;k<=K_MAX;k++)
   {
/*Starting with r */
     for(r=1;r<=R_MAX;r++)
     {
       if(r<k)
       f[k][r]=INF;
       min=INF;
       for(nk=1;nk<=r;nk++)
       {
         if(nk>=1 && nk<=Nk[k])
/* Implementing the recursion function */
         f[k][r]=a[k]/nk+f[k-1][r-nk];
         if(f[k][r]<min)
         {
         min=f[k][r];
         n[k][r]=nk;
         } /* End of if */
       } /* End of nk loop */
       f[k][r]=min;
     } /* End of r loop */
```

```
   }  /* End of K loop */
/* Saving Output in a file */
fprintf(op,"|------------------------------------
-----------------------------------------|\n");
fprintf(op,"  r f[1,r]      n1      f[2,r]    n2
  f[3,r]    n3\n");
fprintf(op,"|------------------------------------
-----------------------------------------
|\n");
for(r=1;r<=R_MAX;r++)
  for(k=1;k<=K_MAX;k++)
  {
    if(f[k][r]==INF)
    {
    f[k][r]=0;
    n[k][r]=0;
    }
    if(k==1)
fprintf(op," %d    %10.4f
%d\t",r,f[k][r],n[k][r]);
    if(k>1)
fprintf(op,"    %10.4f     %d\t",f[k][r],n[k][r]);
    if(k==3)
  fprintf(op,"\n");
  }
/* Appending the result to the output file */
  m=R_MAX;
  for(k=K_MAX;k>=1;k--)
  {
fprintf(op,"\nThe value of n[%d]=%d",k,n[k][m]);
    m=m-n[k][m];
  }
fprintf(op,"\n|--------------END------------------
|\n");
getch( );
return;
}
/* End of Program */
```

## APPENDIX-II

```
#include<stdio.h>
#define K_MAX 3
#define R_MAX 100
#define INF 9999999.0

     main( )
    {
         int l,m2[4][101],k,j,r,i,m,m2k;
     double f[4][101],min;
     double
b[4]={1,318.70212,307.54977,1108.5576};
     double n2k[4]={1,5,6,9};
     double c22k[4]={1,10,12,8};
         FILE *op;
           op=fopen("result4.dat","w+");
         f[0][0]=0.0;
                         f[1][0]=INF;
                         f[2][0]=INF;
                         f[3][0]=INF;
                         l=0;
           for(i=1;i<=100;i++)
                f[l][i]=0.0;
          for(k=1;k<=K_MAX;k++)
                         {
                                   for(r=1;r<=R_MAX;r++)
                                    {
                                            if(r<k)

f[k][r]=INF;
                                                min=INF;
                     for(m2k=1;m2k<=r;m2k++)
          {
        if(r<c22k[k]*m2k)
          f[k-1][r-c22k[k]*m2k]=INF;
                                  if(m2k>=1
&&m2k<=n2k[k])
      f[k][r]=b[k]/m2k+f[k-1][r-c22k[k]*m2k];

if(f[k][r]<min)


min=f[k][r];
          m2[k][r]=m2k;
```

```
                                    } /* End
of if */
                            } /* End of m2k loop
*/
                              } /* End of r loop
*/
                         }  /* End of K loop */
                   /* Saving Output in a file */
fprintf(op,"|---------------------------------------
-----------------------------|\n");
        fprintf(op,"    r  f[1,r]   m21   f[2,r]
m22          f[3,r]         m23\n");
        fprintf(op,"|------------------------------
-----------------------------|\n");
                   for(r=1;r<=R_MAX;r++)
                         for(k=1;k<=K_MAX;k++)
                       {
                              if(f[k][r]==INF)
                              {
                              f[k][r]=0;
                              m2[k][r]=0;
                              }
                              if(k==1)
 fprintf(op,"   %d    %10.4f
%d\t",r,f[k][r],m2[k][r]);
                              if(k>1)
 fprintf(op,"%10.4f      %d\t",f[k][r],m2[k][r]);
                              if(k==3)
 fprintf(op,"\n");
                         }
                 /* Appending the result to the output
file */
                         m=R_MAX;
                         for(k=K_MAX;k>=1;k--)
                         {
 fprintf(op,"\n The value of
m2[%d]=%d",k,m2[k][m]);
                              m=m-m2[k][m];
                          }
 fprintf(op,"----------------END-----------------
|\n");
   getch();
   return;
                                              }
                   /* END OF PROGRAM */
```

# CHAPTER-V

## THE PROBLEM OF OPTIMUM STRATIFICATION UNDER NEYMAN ALLOCATION: A MATHEMATICAL PROGRAMMING APPROACH

# CHAPTER-V

# THE POBLEM OF OPTIMUM STRATIFICATION UNDER NEYMAN ALLOCATION: A MATHEMATICAL PROGRAMMING APPROACH

## 5.1 INTRODUCTION

As given in chapter I, Section IA. 5, the use of stratified sampling involves the solution of four carefully formulated optimization problems according to the objective and available resources to the sample survey. These four optimization problems are related to the optimum choice of the

(i) Stratification variable

(ii) Number of strata

(iii) Stratum boundaries

(iv) Sample size allocations

In this chapter the problem of selecting the optimum strata boundaries is discussed as anMPP and a solution procedure is proposed that uses dynamic programming technique. This chapter is based on my research paper entitled "Optimum Stratification for exponential study variable under Neyman Allocation" accepted for presentation in the 5th International Symposium on Optimization and Statistics, to be held in this department during December 28-30, 2002.

The basic consideration involving in the formation of strata is that the strata should be internally as homogenous as possible, that is stratum variances $S_h^2$ are as small as possible. If the distribution of the study variable is available the strata would be created by cutting this distribution at suitable points.

Given the number of strata, Dalenius and Gruney (1951) suggested that the strata boundaries be so determined that $W_h S_h$ remain constant.

Maholanobis (1952) and Hansen, Hurwitz and Madow (1953) have suggested that strata boundaries be so determined that $W_h \overline{Y}_h$ remain constraint. Dalenius and Hodges (1959) have supported the work of Dalenius and Gruney (1951).

Dalenius (1957) has worked out the best stratum boundaries under proportional and Neyman allocation. Ekman (1959) has suggested approximation to complicated theoretical solutions. Cochran (1961) has examined the applications of these approximations through the empirical studies. Sethi (1963) has showed that the above suggestions fail to provide optimum strata boundaries for certain types of populations. He derived the solutions for optimum stratification points for certain populations. Hess, Sethi and Balakrishnan (1966) have applied these solutions to some empirical studies and made a comparison of various approximations. Singh and Sukhatme (1969) have suggested several approximate methods to obtain optimal points of stratification. Singh & Sukhatme (1973) have suggested certain rules for obtaining optimal stratification points based on auxiliary information. Some others who worked on this problem are Singh (1977), Unnithan (1978), Yadav and Singh (1984) etc.

Khan et al (2002b) have formulated the problem of optimum stratification as a mathematical programming problem and developed a solution procedure using dynamic programming technique. They have applied their procedure to work out optimum strata boundaries to populations having uniform and right triangular distributions.

Most of the authors who worked on this problem obtained minimal equations for optimum strata boundaries. Unfortunately these equations are difficult to solve for exact solutions. So that only approximate solution can be obtained. Some authors suggested iterative procedures that are very slow even to obtain a local minimum of the objective function. Moreover, the iterative procedures may oscillate and there is no guarantee that they will provide us with the approximate global minimum.

In this chapter the approach of Khan et al (2002b) is extended to work out optimum strata boundaries for an exponential population under Neyman allocation.

## 5.2 THE PROBLEM

Let the population under study is to be stratified into L strata and the estimation of the population mean is of interest. Let $x_0$ and $x_L$ be the smallest and largest values of the study variable $x$ in the population. The problem of optimum stratification can be described as to find the intermediate stratum boundaries $x_1, x_2, \ldots, x_{L-1}$ such that the variance of the stratified sample mean $\bar{x}_{st}$ under Neyman (1934) allocation is minimum.

The variance of the stratified sample mean

$$\bar{x}_{st} = \sum_{h=1}^{L} W_h \bar{x}_h \tag{5.1}$$

under Neyman allocation is given as

$$V(\bar{x}_{st}) = \frac{\left(\sum_{h=1}^{L} W_h S_h\right)^2}{n} - \frac{\sum_{h=1}^{L} W_h S_h^2}{N} \tag{5.2}$$

where the symbols have the same meaning as described in Section 1A.4 of Chapter I except that the study variable is denoted by $x$.

If the finite population correction is ignored, minimizing expression on the right hand side of (5.2) is equivalent to minimizing

$$\sum_{h=1}^{L} W_h S_h \tag{5.3}$$

Let $f(x)$ denotes frequency function of the study variable $x$, $x_0 \le x \le x_L$. The problem of determining the strata boundaries is equivalent to cut up the range

$$x_L - x_0 = d \ (say) \tag{5.4}$$

at points $x_1 \le x_2 \le ..., \le x_{L-1}$ such that (5.3) is minimum. Where the values of $W_h$ and $S_h$ are obtained by

$$W_h = \int_{x_{h-1}}^{x_h} f(x)dx \qquad (5\ 5)$$

$$S_h^2 = \frac{1}{W} \int_{x_{h-1}}^{x_h} f(x)dx - \mu_h^2 \qquad (5\ 6)$$

where $\mu_h = \frac{1}{W_h} \int_{x_{h-1}}^{x_h} xf(x)dx \qquad (5\ 7)$

and $(x_{h-1}, x_h)$ are the boundaries of hth the stratum.

When the frequency function $f(x)$ is known, using (5.5), (5.6) and (5.7), $W_h S_h$ could be expressed as a function of $x_h$ and $x_{h-1}$ only.

Let $f_h(x_h, x_{h-1}) = W_h S_h$

Then the problem of determining the optimum strata boundaries (OSB) can be expressed as:

"Find $x_1, x_2, ..., x_{L-1}$ that minimize $\sum_{h=1}^{L} f_h(x_h, x_{h-1})$, subject to the constraints

$x_o \le x_1 \le x_2 \le, ..., \le x_{L-1} \le x_L$"

Define

$y_h = x_h - x_{h-1}$ ; $h = 1,2,...,L$

where $y_h \ge 0$ denotes the width of the hth stratum.

With the above definition of $y_h$ (5.4) can be expressed as

$$\sum_{h=1}^{L} y_h = \sum_{h=1}^{L} (x_h - x_{h-1}) = x_L - x_o = d$$

The kth stratification point $x_k$ ; $k=1,2,...,L-1$ can then be given as

$$x_k = x_0 + y_1 + y_2 + ..... + y_k$$

Then the problem of determining optimum strata boundaries can be considered as the problem of determining optimum strata widths and can be expressed as the following Mathematical Programming Problem (MPP):

$$\left.\begin{aligned} &\text{Minimize} \quad \sum_{h=1}^{L} f_h(y_h, x_{h-1}) \\ &\text{subject to} \quad \sum_{h=1}^{L} y_h = d \\ &\text{and} \qquad y_h \geq 0; \; h = 1,2,...,L \end{aligned}\right\} \tag{5.8}$$

For h=1 the term $f_1(y_1, x_0)$ in the objective function of (5.8) is a function of $y_1$ alone, as $x_0$ is known for h=2 the term $f_2(y_2, x_1) = f_2(y_2, x_0 + y_1)$ will become a function of $y_2$ alone once $y_1$ is known. Thus, we may rewrite the MPP (5.8) expressing the objective function a function of $y_h$ alone as:

$$\left.\begin{aligned} &\text{Minimize} \quad \sum_{h=1}^{L} f_h(y_h) \\ &\text{subject to} \quad \sum_{h=1}^{L} y_h = d \\ &\text{and} \qquad y_h \geq 0; \; h = 1,2,...,L \end{aligned}\right\} \tag{5.9}$$

Let $x$ follows an exponential frequency function:

$$f(x) = e^{-x}, x \geq 0 \qquad (5.10)$$
$$= 0 \text{ ; otherwise}$$

In practice the actual populations are often finite, assuming the largest value of $x$ in the population as $D$, (5.10) can be rewritten as

$$f(x) = e^{-x} \;;\; 0 \leq x \leq D$$

$= 0$ ;otherwise

$$\Rightarrow x_o = 0 \text{ and } x_L = D$$

From (5.5)

$$W_h = \int_{x_{h-1}}^{x_h} f(x)dx$$

$$= \int_{x_{h-1}}^{y_h + x_{h-1}} f(x)dx \qquad \text{(because } x_h = y_h + x_{h-1}\text{)}$$

$$= \int_{x_{h-1}}^{y_h + x_{h-1}} e^{-x} dx$$

$$= \left[-e^{-x}\right]_{x_{h-1}}^{y_h + x_{h-1}}$$

$$= e^{-x_{h-1}} - e^{-(y_h + x_{h-1})}$$

$$\text{or } W_h = e^{-x_{h-1}}\left(1 - e^{-y_h}\right) \qquad (5.11)$$

From (5.7)

$$\mu_h = \frac{1}{W_h}\int_{x_h}^{y_h + x_{h-1}} x f(x_h) dx$$

$$= \frac{1}{W_h}\int_{x_{h-1}}^{y_h + x_{h-1}} x.e^{-x}\; dx$$

$$= \frac{1}{W_h}\left[-xe^{-x} + \int e^{-x}dx\right]_{x_{h-1}}^{y_h+x_{h-1}}$$

$$= \frac{1}{W_h}\left[-xe^{-x} - e^{-x}\right]_{x_{h-1}}^{y_h+x_{h-1}}$$

$$= \frac{1}{W_h}\left[-e^{-x}(1+x)\right]_{x_{h-1}}^{y_h+x_{h-1}}$$

$$= \frac{1}{W_h}\left[e^{-x_{h-1}}\left(1+x_{h-1}\right) - e^{-\left(y_h+x_{h-1}\right)}\left(1+y_h+x_{h-1}\right)\right]$$

$$= \frac{1}{W_h}e^{-x_{h-1}}\left[\left(1+x_{h-1}\right) - e^{-yh}\left(1+y_h+x_{h-1}\right)\right]$$

$$= \frac{e^{-x_{h-1}}\left[\left(1+x_{h-1}\right)\left(1-e^{-y_h}\right) - y_h e^{-yh}\right]}{W_h}$$

Therefore $\mu_h = \dfrac{e^{-x_{h-1}}\left[\left(1+x_{h-1}\right)\left(1-e^{-yh}\right) - y_n e^{-yh}\right]}{e^{-x_{h-1}}\left(1-e^{-yh}\right)}$

or $\mu_h = \dfrac{\left[\left(1+x_{h-1}\right)\left(1-e^{-yh}\right) - y_n e^{-yh}\right]}{1-e^{-yh}}$ (5.12)

From (5.6) $S_h^2 = \dfrac{1}{W_h}\int_{x_{h-1}}^{y_h+x_{h-1}} x^2 f(x)dx - \mu_h^2$

$$= \frac{1}{W_h}\int_{x_{h-1}}^{y_{h-1}+x_{h-1}} x^2 e^{-x}dx - \mu_h^2$$

$$= \frac{1}{W_h}\left[-x^2e^{-x} + \int 2x.e^{-x}dx\right]_{x_{h-1}}^{y_h+x_{h-1}} - \mu_h^2$$

$$= \frac{1}{W_h}\left[-x^2e^{-x} - 2xe^{-x} - 2e^{-x} - 2e^{-x})\right]_{x_{h-1}}^{y_h+x_{h-1}} - \mu_h^2$$

$$= \frac{1}{W_h}\left[-e^{-x}(x^2+2x+2)\right]_{x_{h-1}}^{y_h+x_{h-1}} - \mu_h^2$$

$$= \frac{1}{W_h}\left[e^{-x_{h-1}}\left(x_{h-1}^2+2_{x_{h-1}}+2\right) - e^{-(y_h+x_{h-1})}\left((y_h+x_{h-1})^2\right.\right.$$

$$\left.\left. + 2(y_h+x_{h-1})+2\right) - \mu_h^2\right]$$

$$= \frac{e^{-x_{h-1}}\left[\left(x_{h-1}^2+2x_{h-1}+2\right) - e^{-y_h}\left(y_h^2+2y_hx_{h-1}+x_{h-1}^2+2y_h+2x_{h-1}+2\right)\right]}{W_h}$$

$$- \mu_h^2$$

$$= \frac{\left[\left(x_{h-1}^2+2x_{h-1}+2\right) - e^{-y_h}\left(x_{h-1}^2+2x_{h-1}+2\right) - y_he^{-y_h}\left(y_h+2x_{h-1}+2\right)\right]}{\left(1-e^{-y_h}\right)}$$

$$- \mu_h^2$$

$$= \frac{\left(x_{h-1}^2+2x_{h-1}+2\right)\left(1-e^{-y_h}\right) - y_he^{-y_h}\left(y_h+2x_{h-1}+2\right)}{\left(1-e^{-y_h}\right)}$$

$$- \left[\frac{(1-x_{h-1})\left(1-e^{-y_h}\right) - y_ne^{-y_h}}{1-e^{-y_h}}\right]^2 \qquad \text{(using (5.12))}$$

Putting $a_h = \left(1-e^{-y_h}\right)$ we get

$$S_h^2 = \frac{\left(x_{h-1}^2+2x_{h-1}+2\right)a_h - y_he^{-y_h}\left(y_h+2x_{h-1}+2\right)}{a_h}$$

$$- \frac{(1+x_{h-1})^2a_h^2 - y_h^2e^{-2y_h}}{a_h^2}$$

$$=\frac{a_h^2-y_h^2e^{-y_h}a_h-y_h^2e^{-2y_h}}{a_h^2}$$

$$=\frac{a_h^2-y_h^2e^{-y_h}\left(a_h+e^{-y_h}\right)}{a_h^2}$$

$$=\frac{a_h^2-y_h^2e^{-y_h}}{a_h^2}$$

or $$S_h^2=\frac{\left(1-e^{-y_h}\right)^2-y_h^2e^{-y_h}}{\left(1-e^{-y_h}\right)^2}$$

Which gives

$$S_h=\frac{\left[\left(1-e^{-y_h}\right)^2-y_h^2e^{-y_h}\right]^{1/2}}{\left(1-e^{-y_h}\right)}$$

Using (5.11) we get

$$W_hS_h=e^{-x_{h-1}}\left[\left(1-e^{-y_h}\right)^2-y_h^2e^{-y_h}\right]^{1/2} \tag{5.13}$$

Using (5.11), (5.12) and (5.13) the MPP (5.9), can be restated as:

$$\left.\begin{aligned}&Minimize\ \sum_{h=1}^{L}e^{-x_{h-1}}\left[\left(1-e^{-y_h}\right)^2-y_h^2e^{-y_h}\right]^{1/2}\\ &subject\ to\ \sum_{h=1}^{L}y_h=d\\ &and\qquad y_h\geq 0;\ h=1,2,\ldots,L\end{aligned}\right\} \tag{5.14}$$

where $x$ follows the exponential distribution as defined in (5.10).

## 5.3 THE SOLUTION

Consider the following subproblems of (5.9) for first k (<L) strata

$$\left.\begin{aligned} &Minimize \ \sum_{h=1}^{L} f_h(y_h) \\ &subject\ to \ \sum_{h=1}^{L} y_h = d_k \\ &and \qquad y_h \geq 0;\ h = 1,2,\ldots,k \end{aligned}\right\} \tag{5.15}$$

where $d_k < d$ is the total width available for division into k strata.

Note that $d_k = d$ for $k = L$.

We then have

$$d_k = y_1 + y_2 + \ldots + y_k$$

$$d_{k-1} = y_1 + y_2 + \ldots + y_{k-1} = d_k - y_k$$

$$d_{k-2} = y_1 + y_2 + \ldots + y_{k-2} = d_{k-1} - y_{k-1}$$

$$\vdots$$

$$d_2 = y_1 + y_2 = d_3 - y_3$$

$and\ d_1 = d_2 - y_2$

Let $f(k, d_k)$ denotes the minimum value of objective function of (5.15), that

is, $f(k,d_k) = \min\left[\sum_{h=1}^{k} f_h(y_h) \mid \sum_{h=1}^{k} y_h = d_k \ and\ y_h \geq 0; h = 1,2,\ldots,k\right]$

The recurrence relation of the Dynamic Programming thus be given as

$$f(k, d_k) = \min_{0 \le y_k \le d_k} [f_k(y_k) + f(k-1, d_k - y_k)], k \ge 2 \tag{5.16}$$

For the first stage (i.e. k=1):

$$f(1, d_1) = f_1(d_1) \Rightarrow y^* = d_1 \tag{5.17}$$

From $f(L, d)$ the optimum width of Lth stratum, $y_L^*$, is obtained ;

from $f(L-1, d - y_L^*)$ the optimum width of (L-1) th stratum $y_{L-1}^*$, is obtained

and so on until $y_1^*$ is obtained.

Using (5.16) and (5.17) we get for first stage (k=1)

$$f(1, d_1) = \left[ \left(1 - e^{-d_1}\right)^2 - d_1^2 e^{-d_1} \right]^{1/2} \text{ at } y_1^* = d_1 \tag{5.18}$$

because $x_{k-1} = x_0 = 0$ when $K = 1$

For the stage $k \ge 2$

$$f(k, d_k) = \min_{0 \le y_k \le d_k} \left[ e^{-(d_k - y_k)} \left[ \left(1 - e^{-y_k}\right)^2 - y_k^2 e^{-y_k} \right]^{1/2} + f(k-1, d_k - y_k) \right] \tag{5.19}$$

where $x_{k-1} = x_0 + y_1 + y_2 + \dots\dots\dots\dots + y_{k-1} = d_k - y_k$

## 5.4 A NUMERICAL EXAMPLE

Relation (5.18) and (5.19) are the required relations of the dynamic programming. Execution of the computer program in 'Java SDK 2', given in

Appendix, of this chapter, gives the optimum stratum boundaries for the exponential study variable with density function

$$f(x) = \left.\begin{array}{ll} e^{-x}, & x > 0 \\ 0\ , & elsewhere \end{array}\right\} \tag{5.20}$$

for 2, 3, 4 and 5 strata. The results are presented in a tabular form in Table 5.1 along with the values of $\sum_{h=1}^{L} W_h S_h$ .

**Table 5.1**

Optimum strata boundaries for 2, 3, 4 and 5 strata

| No. of strata<br>$L$ | Strata widths<br>$y_h^*$ | Strata boundary points<br>$x_h^* = x_{h-1}^* + y_h^*$ | Optimum value of the objective function<br>$\sum_{h=1}^{L} f_h(y_h) = \sum_{h=1}^{L} W_h S_h$ |
|---|---|---|---|
| 2 | $y_1^*$= 1.2610<br>$y_2^*$= 18.7390 | $x_1^* = x_0 + y_1^*$= 1.2610 | 0.5341 |
| 3 | $y_1^*$= 0.7678<br>$y_2^*$= 1.2501<br>$y_3^*$= 17.9821 | $x_1^* = x_0 + y_1^*$= 0.7678<br>$x_2^* = x_1^* + y_2^*$= 2.0179 | 0.3648 |
| 4 | $y_1^*$= 0.5509<br>$y_2^*$= 0.7638<br>$y_3^*$= 1.2513<br>$y_4^*$= 17.4340 | $x_1^* = x_0 + y_1^*$= 0.5509<br>$x_2^* = x_1^* + y_2^*$= 1.3147<br>$x_3^* = x_2^* + y_3^*$= 2.5650 | 0.2770 |
| 5 | $y_1^*$= 0.4393<br>$y_2^*$= 0.5610<br>$y_3^*$= 0.7569<br>$y_4^*$= 1.2688<br>$y_5^*$= 16.9740 | $x_1^* = x_0 + y_1^*$= 0.4393<br>$x_2^* = x_1^* + y_2^*$= 1.0003<br>$x_3^* = x_2^* + y_3^*$= 1.7572<br>$x_4^* = x_3^* + y_4^*$= 2.0260 | 0.2233 |

The total width available for cutting stratum boundaries is taken as 20 units, that is $x_L = D = 20$, because the area above $x = 20$ for exponential distribution given in (5.20) is almost zero.

## 5.5 CONCLUSION

Unnithan (1978) showed that the iterative procedure by Dalenius and Hodge (1959) is slow even to obtain a local minimum; also it does not suggests any stopping rule and may oscillate. He also suggested an iterative solution procedure using modified Newton's method. Both these procedures require initial approximate solutions. Also there is no guarantee that these procedures will provide a global optimum. The advantage of the proposed solution procedure is that it provides a global minimum.

## APPENDIX

```
import java.io.*;
import java.util.*;

public class OptimumNew
{

        private RandomAccessFile randReader[] = null;
        private double e=2.718281828;
        private double increment = 0.10;
        private int intPreci = 1;
        private int intStage = 1;
        private int Dk = 999;
        DataOutputStream outputStream[];
        double storedFk[];

        public static void main(String args[])
        {
                new OptimumNew();
        }

        public OptimumNew()
        {
                System.out.println("enter the Stage value (1 to 9 only):");
                String str = Readline.readLine();
                intStage = Integer.parseInt(str);


                System.out.println("enter the summation Yk ( Dk ) value (integer
only):");
                str = Readline.readLine();
                Dk = Integer.parseInt(str);
                System.out.println("enter the desired precesion 1- 9 (integer
only):");
                str = Readline.readLine();
                intPreci = Integer.parseInt(str);

                try
                {
                        randReader = new RandomAccessFile[intStage];
```

```
                        for(int i =0; i < intStage; i++)
                        {
                                File file = new File("./Stage"+(i+1)+".txt");
                                randReader[i] = new RandomAccessFile(file, "r");
                        }
                        FileOutputStream fos[] = new
FileOutputStream[intStage];
                         outputStream = new DataOutputStream[intStage];
                        for(int i =0; i < intStage; i++)
                        {
                                File file = new File("./Stage"+(i+1)+".txt");
                                fos[i] = new FileOutputStream(file);
                                outputStream[i] = new DataOutputStream(fos[i]);
                        }
                        funF1D1()     ;
                        for(int i = 1; i < intStage; i++)
                                funFkDk(i)    ;
                        backWardCalculation();

                }
                catch(Exception ex)
                {
                        ex.printStackTrace();

                }
        }


                void funF1D1()
                {
                        storedFk = new double[(int)(Dk*Math.pow(10,
intPreci)+1)];
                        double Y1=0;
                        double dblTmp1 = 0;
                        double fx= 0;
                        double d1 = 0;
                        long d1Count=0;
                        int count = 0;
                        String strD1 = "", strFx="", strY1="";
                        increment = Math.pow(10, -intPreci);
                        //System.out.println(increment);
                        while(d1 <= Dk)
                        {
```

```
                    Y1 = d1;
                    fx = (1 - Math.pow(e, -Y1))*(1 - Math.pow(e, -
Y1)) - Y1*Y1*Math.pow(e, -Y1);
                    if(fx<0.0)
                    {
                         System.out.println("SQRT OF THE -VE
QUANTITY in funFkDk_");
                         System.out.println("d1="+d1+",
Y1="+Y1+" , fx= "+fx+"\n");
                         System.exit(0);
                    }
                    else
                         dblTmp1=Math.sqrt(fx);

                    fx=Math.pow(e, -(d1-Y1))*dblTmp1;
                    storedFk[count] = fx;
                    count++;
                    strFx = Double.toString(fx);
                    while(strFx.length() < 25)
                    {
                         strFx = "0" + strFx;
                    }
                    strY1 = Double.toString(Y1);
                    while(strY1.length() < 25)
                    {
                         strY1 = "0" + strY1;
                    }
                    strD1 = Double.toString(d1);
                    while(strD1.length() < 25)
                    {
                         strD1 = "0" + strD1;
                    }
                    try
                    {
                         outputStream[0].writeBytes(strD1+" " +
strY1+" " + strFx+"\n");
                    }
                    catch(Exception ex)
                    {
                         ex.printStackTrace();
                    }
                    //d1 += increment;
                    d1Count++;
```

```
                        d1 = d1Count*Math.pow(10, -
intPreci);//increment;
                }

        }

        double readFkDk1(int K, double Dk)
        {
                double tmpDk = Dk*Math.pow(10, intPreci);
                long lDk = (long)Dk;
                long n1 = (long)tmpDk;
    //Math.round(Dk*Math.pow(10, intPreci));
                String str= "";
                double ret=0, data1 =0, data2 =0;
                n1 = n1*78;
                try
                {
                        if(n1 < 0 || n1 > randReader[K].length()) return 0;
                        randReader[K].seek(n1);
                        str = randReader[K].readLine();
                        data1 = Double.parseDouble(str.substring(51));
                        if(str != null && str.length() >= 75)
                        {
                                data2 =
Double.parseDouble(str.substring(26, 51));
                        }
                        else
                         data2 = data1;
                         ret = data1+ (data2-data1)*(Dk*100 -
lDk*100)/100;
                        //System.out.println( "fkdk- Dk passed =" + Dk + ",
line = "+n1/78 +", Fx=" + ret );
                }
                catch(Exception ex)
                {
                        System.out.println( "fkdk- Dk passed =" + Dk + ",
line = "+n1/78 +", str=" + str );
                        ex.printStackTrace();
                        System.exit(0);

                }
                return ret;
        }
```

```
double readFkDk(int K, double Dk)
{
	double tmpDk = Dk*Math.pow(10, intPreci);
	long lDk = (long)Dk;
	int n1 = (int)tmpDk; //Math.round(Dk*Math.pow(10,
intPreci));
	String str= "";
	double ret=0, data1 =0, data2 =0;
	try
	{
		data1 = storedFk[n1];
		if(n1 < storedFk.length-1)
		{
			data2 = storedFk[n1+1];
		}
		else
		 data2 = data1;
		 ret = data1+ (data2-data1)*(Dk*100 -
lDk*100)/100;
		//System.out.println( "fkdk- Dk passed =" + Dk + ",
line = "+n1/78 +", Fx=" + ret );
	}
	catch(Exception ex)
	{
		System.out.println( "readFkDk- Dk passed =" + Dk
+ ", line = "+n1/78 +", str=" + str );
		ex.printStackTrace();
		System.exit(0);

	}
	return ret;
}

void funFkDk(int K)
{
	if(K > 1)
		readStoredFk(K);
	double Yk=0;
	double dblTmp1 = 0;
	long multi = (long)Math.pow(10, intPreci+1);
	double fx= 0;
```

```
double dk = increment;
long dkCount =1;
String strD1 = "", strFx="", strY1="";
//for 0<= Yk<= Dk
double increL = increment/10;
double minFx =999999;
double minYk = 45667;
try
{

outputStream[K].writeBytes("000000000000000000000000
000000000000000000000000 000000000000000000000000\n");
}
catch(Exception ex)
{
	ex.printStackTrace();
}
double lowLimit = 0, upperLimit = 0, increTmp =0;
while(dk <= Dk)
{
	//find min
	minFx = 999999;
	minYk = 999999;
	lowLimit = 0;
	upperLimit = dk;
	if(increL <= .01) increTmp = .1;
	if(upperLimit >= 20) increTmp = 1;
	if(upperLimit > 200) increTmp = 10;
	if(upperLimit <= 20*increL) increTmp = increL;
	while(increL <= increTmp)
	{
		minFx = 999999;
		Yk = lowLimit;
		//while(Yk<=dk)
		int count =0;
		while(Yk<=upperLimit)
		{
			count++;
			//calculateFkDk();
			dblTmp1 = (1 - Math.pow(e, -
Yk))*(1 - Math.pow(e, -Yk)) - Yk*Yk*Math.pow(e, -Yk);
			if(dblTmp1 < 0.0)
			{
```

```
                                        System.out.println("Sqrt of the
-ve quantity in fun2_");

        System.out.println("Yk="+Yk+" ,increTmp="+increTmp+" ,dk="+dk+"
,dblTmp1="+dblTmp1+" \n"+Math.pow(e, -Yk));
                                        System.exit(0);
                                }
                                else
                                        dblTmp1 =
Math.sqrt(dblTmp1);

                                fx = Math.pow(e,-(dk-Yk)) *
dblTmp1 + readFkDk(K-1, dk-Yk);
                                if(minFx > fx)
                                {
                                        minFx = fx;
                                        minYk = Yk;
                                }
                                Yk += increTmp;
                        }
                        lowLimit = minYk-increTmp;
                        upperLimit = minYk + increTmp;
                        if(upperLimit > dk ) upperLimit = dk;
                        if(lowLimit < 0 ) lowLimit = 0;
                        increTmp = increTmp/10;
                }

                strFx = Double.toString(minFx);
                while(strFx.length() < 25)
                {
                        strFx = "0" + strFx;
                }
                strY1 = Double.toString(minYk);
                while(strY1.length() < 25)
                {
                        strY1 = "0" + strY1;
                }
                strD1 = Double.toString(dk);
                while(strD1.length() < 25)
                {
                        strD1 = "0" + strD1;
                }
                try
```

```
                        {
                                outputStream[K].writeBytes(strD1+" " +
strY1+" " + strFx+"\n");
                        }
                        catch(Exception ex)
                        {
                                ex.printStackTrace();
                        }
                        Yk = dk;
                        dkCount++;
                        dk = dkCount*Math.pow(10, -
intPreci);//increment;
                }
                try
                {
                        System.out.println(K+" file-
"+randReader[K].length());
                }
                catch(Exception ex)
                {
                        ex.printStackTrace();
                }
        }

        void backWardCalculation()
        {
                try
                {
                File tmpFile = new File("./resultNew.txt");
                RandomAccessFile rand = new
RandomAccessFile(tmpFile, "rw");
                rand.seek(rand.length());
                double fxx[] = new double[intStage];
                double fyy[] = new double[intStage];
                double fdd[] = new double[intStage];
                int kk = intStage-1;
                fxx[kk] = readFkDk1(kk, Dk);
                fyy[kk] = readYk(kk, Dk);
                fdd[kk]= Dk;
                rand.writeBytes("\n Date: " + new Date() + "\nNumber of
stage = "+ intStage +", Dk = " + Dk + ", Precision = "+ intPreci
                );
                //System.out.println( "Yk- Dk =" + Dk );
```

```
                    for( int i =kk-1; i >= 0 ; i--)
                    {
                            fdd[i]= fdd[i+1] - fyy[i+1];
                            fxx[i] = readFkDk1(i, fdd[i+1]-fyy[i+1]);
                            fyy[i] = readYk(i, fdd[i+1]-fyy[i+1]);
                            //System.out.println(fdd[i+1] + ", fdd=" + fdd[i] );
                    }

                    for( int i =0; i <= kk ; i++)
                    {
                            rand.writeBytes("\nY" + (i+1) + " = " + fyy[i] +",
D" + (i+1) + " = " + fdd[i]);
                    }
                    rand.writeBytes("\nfx" + (kk+1) + " = " + fxx[kk]+
"\n\n\n\n");
            }
            catch(Exception ex)
            {
                    ex.printStackTrace();
            }
            }

                            double readYk(int K, double Dk)
                            {
                                    double tmpDk = Dk*Math.pow(10,
intPreci);
                                    long lDk = (long)Dk;
                                    long n1 = (long)tmpDk;
    //Math.round(Dk*Math.pow(10, intPreci));
                                    double ret=0, data1 =0, data2 =0;
                                    String str= "";
                                    n1 = n1*78;
                                    try
                                    {
                                            if(n1 < 0 || n1 >
randReader[K].length()) return 0;
                                            randReader[K].seek(n1);
                                            str= randReader[K].readLine();
                                            data1 =
Double.parseDouble(str.substring(26,51));
                                            str= randReader[K].readLine();
                                            if(str != null && str.length() >= 75)
                                            {
```

```
                                        data2 =
Double.parseDouble(str.substring(26, 51));
                                    }
                                    else
                                     data2 = data1;
                                     ret = data1+ (data2-data1)*(Dk*100
-lDk*100)/100;
                                    //System.out.println( "Dk passed =" +
Dk + ", line = "+n1/78 +", Fx=" + ret );
                              }
                              catch(Exception ex)
                              {
                                    System.out.println( K+",Dk passed
=" + Dk + ", line = "+n1/78 +", str=" + str );
                                    ex.printStackTrace();

                              }
                              return ret;
                        }


                  void readStoredFk(int k)
                  {
                        k--;
                        try
                        {
                              File file = new File("./Stage"+(k+1)+".txt");
                              RandomAccessFile randTmp = new
RandomAccessFile(file, "r");
                        //    randReader[k].seek(0);
                              System.out.println( "filelength read= "
+randReader[k].length() );
                              String str = null;
                              int line = 0;
                              System.out.println( "filelength= "
+randTmp.length() + ", array=" + storedFk.length);

                              while((str = randTmp.readLine()) != null
&& line < storedFk.length)
                              {
                                    storedFk[line] =
Double.parseDouble(str.substring(51));
                                    line++;
```

```
                    }
                    System.out.println( "k= " +k + ", line=" +
line);
                }
                catch(Exception ex)
                {
                    ex.printStackTrace();
                }

            }


//*************************************************************
*****
    static class Readline
    {

        public static void main(String args[])
        {
            try{
                // 1. Create an InputStreamReader using the
standard input stream
                InputStreamReader isr = new InputStreamReader(
System.in );

                // 2. Create a BufferedReader using the
InputStreamReader created.
                BufferedReader stdin = new BufferedReader( isr );

                // 3. Don't forget to prompt the user
                System.out.print( "Type some data for the program:
" );

                // 4. Use the BufferedReader to read a line of text
from the user.
                String input = stdin.readLine();

                // 5. Now, you can do anything with the input string
that you need to.
                // Like, output it to the user.
                System.out.println( "input = " + input );
            }catch(Exception ex){ex.printStackTrace();}
        }
```

```
public static String readLine()
{
        String input = "0";
        try{
                // 1. Create an InputStreamReader using the
standard input stream
                InputStreamReader isr = new InputStreamReader(
System.in );

                // 2. Create a BufferedReader using the
InputStreamReader created.
                BufferedReader stdin = new BufferedReader( isr );

                // 4. Use the BufferedReader to read a line of text
from the user.
                input = stdin.readLine();


        }catch(Exception
ex){ex.printStackTrace();System.exit(0);}
        finally
        {

        }
                return input;
}
    }


}
```

# REFERENCES

# REFERENCES


Aggrawal, O.P.(1974a): "On mixed integer quadratic problems", *Naval Research Logistics Quaterly*, **21.**

Aggrawal, O.P.(1974b): "On integer solution to quadratic problems by branch and bound technique",*Trabajas de Estatistica Y De investigation Opration*, **25**, 65-70.

Ahsan, M. J. (1975): "A procedure for the problem of optimum allocation in multivariate stratified random sampling", *Aligarh Bulletin of Mathematics*, **5 & 6**, 37-42.

Ahsan, M. J. (1978): "Allocation problem in multivariate stratified random sampling", *Journal of the Indian Statistical Association*, **16**, 1-5.

Ahsan, M. J. and Khan, S.U. (1977): "Optimum allocation in multivariate stratified sampling using prior information", *Journal of Indian Statistical Association*, **15**, 57-67.

Ahsan, M. J. and Khan, S.U. (1982): "Optimum allocation in multivariate stratified sampling with overhead cost", *Metrika*, **29**, 71-78.

Ahsan, M.J.; Khan, S.U. and Arshad, M. (1983): "Minimizing a non-linear function arising in stratification through approximation by a quadratic function", *Journal of Ind. Society Statist. Operation Research.*, **4**, **1-4**, 9-16.

Alldredge, J. R. and Armstrong, D. W. (1974): "Maximum likelihood estimation for the multinomial distribution using geometric programming", *Technometrics*, **16**, 585.

Anstreicher, K. M., Den Hertog , D. and Terlaky, T.(1994): "A long step barrier method for convex quadratic programming", *Algorithmica*, **19**, No.5, 365-382.

Aoyama, H.(1963): " Stratified random sampling with optimum allocation for multivariate populations", *Ann. Inst. Stat. Math.*, **14**, 251-258.

Aragon, J. and Pathak, P.K. (1990): "An algorithm for optimal integration of two surveys", *Sankhya, Ser. B*, **52**, No.2, 198-203.

Arbel, A. (1993): " An interior multiobjecitve linear programming algorithm", *Comput. Operations Res.*, **20**, 7, 723-735.

Arbel, A. (1994): "An interior multiobjective primal-dual linear programming algorithm using approximated gradients and sequential generation of anchor points", *Optimization*, **30**, No.2, 137-150.

Arshad, M., Khan , S. U. and Ahsan , M. J. (1981): "A procedure for solving concave quadratic programs", *Aligarh Journal of Statistics*, **1**, 106-112.

Arthanari, T.S. and Dodge, Y (1978), "Mathematical Programming and Construction of BIB Designs", *Technical Report 4*, School of Planning and Computer Applications, Tehran.

Arthenari, T. S. and Dodge, Yodolah (1981): *Mathematical Programming in Statistics*. John Wiley, New York.

Bahn, O. Du Merle, O, Goffin J. L. and Vial, J. P. (1995): "A cutting plane method from analytic centers for stochastic programming", *Mathematical Programming*, **69**, ser B, 45-73.

Bandinelli, R.D. (2000): "An optimal dynamic policy for hotel yield management", *European Journal of Operation Research*, **121**, 476-503.

Barankin, E. W. (1951): "On the system of linear equations with applications to linear programming and the theory of statistical hypothesis", *Publ. Stat.*, **1**, Univ. Calif. Press, Berkely, 161-214.

Beale, E. M. L.(1959): "On quadratic programming", *Naval Research Logistics Quaterly*, **6**, 227-243.

Bellman, R. E. and Dreyfus, S. E. (1962): *"Applied Dynamic Programming"*, Priceton University Press , Princeton.

Bellman, R.E. (1957): *"Dynamic Programming"*, Princeton University Press, Princeton.

Benzi, M. (1993): "Solution of equality- Constrained quadratic programming problems by a projective iterative method", *Rend. Math. Appli.*(7), **13**, No.2, 275-296.

Bethel, J. (1985): "An optimum allocation algorithm for multivariate surveys", *Proceedings of the Survey Research Section, American Statistical Association*, 204-212

Bethel, J. (1989): "Sample allocation in multivariate surveys", *Survey Methodology*, **15**, No.1, 47-57.

Bit, A. K., Biswal, M.P. and Alam, S. S. (1993): "An iterative fuzzy programming algorithm for multiobjective transportation problem", *Journal of Fuzzy Math.*, **1**, No. 4, 835-842.

Bowley, A. L (1926): "Measurement of the precision attained in sampling", *Bull. Inter. Stat. Inst.*, **22**(1), 1-62.

Bretthauer, K. M., Ross, A. and Shetty, B. (1999): "Nonlinear integer programming for optimal allocation in stratified sampling", *European Journal of Operational Research*, **116**, 667-680.

Buhler W., Aachen and Deutler T., Mannheim (1975): "Optimum stratification and grouping by Dynamic Programming", *Metrika*, **22**, 161-175.

Chaddha, R. L. Hardgrave, W. W. , Hudson , D. J. , Segal , M. and Suurballe, J. W. (1971): " Allocation of total sample size when only the stratum means are of interest", *Technometrics*, **13**, 4, 817-831.

Chakarborty, T. K. (1986): " A premptive single sampling attribute plan of given strength", *Opsearch* , **23**, 164-174.

Chakarborty, T. K. (1988): "A single sampling attribute plan of given strength based on fuzzy goal programming", *Opsearch*, **25**, 259-271.

Chakarborty, T. K. (1990): "The determination of indifference quality level single sampling attribute plan of given strength based on fuzzy goal programming", *Sankhya*, *Ser. B*, 238-245.

Chakarborty, T. K. (1991): "Fuzzy goal programming approach for designing single sampling attribute plans when sample size is fixed", *IAPQR Transactions*, **16**, 2, 1-8.

Charnes, A. and Cooper, W.W. (1959): "Chance constrained programming", *Management Science*, **6**, 73-79.

Charnes, A. and Cooper, W.W. (1960): "*Management models and industrial applications of the linear programming*", **1** & **2**, John Wiley and Sons, Inc. New York.

Charnes, A. and Cooper, W.W. (1962): "Programming with linear fractional functionals", *Naval Research Logistics Quaterly*, **9**, 181-186.

Charnes,A. and Cooper, W. W. (1977): "Goal programming and multiobjective optimizations- Part I", *European Journal of Operational Research* , **1**, 39-54.

Chatterjee, S. (1966): "A programming algorithm and its statistical application", *O. N. R. Technical report 1*, Department of Statistics, Harvard University, Cambridge.

Chatterjee, S. (1967): "A note on optimum allocation", *Skand. Akt.*, **50**, 40-44.

Chatterjee, S. (1968): "Multivariate stratified surveys", *Journal of American Statistical Association*, **63**, 530-534.

Chatterjee, S. (1972): "A study of optimum allocation in multivariate stratified surveys", *Skand. Akt.*, **55**, 73-80.

Chernayak, O. I. and Chorous , G. (2000): "Optimum allocation in stratified sampling with a nonlinear cost function", *Theory of Stochas*tic *Processes*, **6**(22), 3-4, 6-17.

Chernayak, O. I. and Starytskyy, G. (1998): " Optimum allocation in stratified sampling with convex cost function", *Visnyk of kviv.* University, Economics, N **39** , 42-46 (In Ukrainian).

Clark, R.G. and Steel, D. G. (2002): "Optimum allocation of sample to strata and stages with simple additional constraints", *The Statistician*, **49**, No.2, 197-207

Cochran, W. G. (1961): "Comparison of methods for determining stratum boundaries," *Bull. Int. Statist. Inst.*, **38**(2), Tokyo, 345-358.

Cochran, W. G. (1963): "*Sampling techniques*", (2nd ed.), New york, Wiley.

Cochran, W. G. (1977): "*Sampling techniques*", (3rd ed.), New york, Wiley.

Crowder, S. V. (1992): "An SPC model for short product on runs minimizing expected cost", *Technometrics*, **34**, 2, 64-73.

Csenki, A. (1997): " Optimum allocation in stratified random sampling via Holder's inequality", *The Statistician* , **46**, 439-441.

Dalenius, T. (1957): "*Sampling in Sweden* : Contributions to the methods and theories of sample survey practice", Almqvist and Wiksell, Stockholm.

Dalenius, T. and Gurney, M. (1951): "The problem of optimum stratification-II", *Skand. Akt.*, **34**, 133-148.

Dalenius, T. and Hodges, J. L. (1959): "Minimum variance stratification", *Journal of American Statistical Association*, **54**, 88-101.

Dantzig, G. B. (1959): " Notes on solving a linear program in integers", *Naval research Logistics Quaterly* , **6**, 75-76.

Dantzig, G.B. and Wald, A. (1951): "On the fundamental lemma of Neyman and Pearson", *Annals of Mathematical Statistics*, **22**, 87-93.

Dembo, R. S. (1982): "Sensitivity analysis in geometric programming", *Journal Optimization Theory Application*, **37**, 1-22.

Deming , W. E. (1950): "*Some theory of sampling*", John Wiley and Sons, New York.

Dinkelbach, W. (1967), "On nonlinear fractional programming", *Management Science*, **13**, 492-498.

Du, D. Z., Wu, F. and Zang, X. S. (1990): "On Rosen's gradient projection methods", *Anals Oper. Res.* **24**, 1-4, 11-28.

Duffin, R. J. Peterson, E. L. and Zener , C. (1967): "Geometric programming : Theory and Applications", Wiley , New York.

Ekman, G. (1959): ' Approximate expressions for the conditional mean and variance over small intervals of a continuous distribution", *Ann. Inst. Stat. Math.*, **30**, 1131-1134.

Ericson, W. A. (1967): "Optimal sample design with nonresponse", *Journal of American Statistical Association*, **62**, 63-78.

Ermer, D.S. (1971) : "Optimization of the constrained machining economics problem by geometric programming", *Journal of Engineering for Industry.* Transactions of A. S. M. E., **93**, 1067-1072.

Fahimi, M. and Pathak, P. K. (1992): "Optimal integration of surveys", *Current Issues in Statistical Inference: essays in Honour of D. Basu*, 208-224.

Finkbeiner, B. and Kall, P. (1978): "Direct algorithms in quadratic programming", *Zeitschrift fur Operations Research*, **17**, 45-54.

Flam, S. D. and Schult, R. (1993): "A new approach to stochastic linear programming" , *Numer. Funct. Anal. Optim.*, **14**, No. 5-6, 545-554.

Fletcher, R. (1971): "A general quadratic programming algorithm", *Journal of the Inst. of Mathematics and its Applications*, **7**, 76-91.

Folks, J. L. and Antle, C. E. (1965): "Optimum allocation of sampling units to the strata when there are R responses of interest", *Journal of American*

*Statistical Association*, **60**, 225-233.

Foody, W. and Hedayat, A. (1977): "On theory and application of BIB designs with repeated blocks", *Annals of Mathematical Statistics*, **5**, 932

Francis, R. L. and Wright , G. (1969): " Some duality relationships for the generalized Neyman-Pearson problems", *Journal of Optimization Theory and Applications*, **4**, 394.

Fukushima, M. (1986): "A successive quadratic programming algorithm with global and superlinear convergence properties", *Mathematical Programming*, **35**, No. 3, 253-264.

Ghosh, S. P. (1958): "A note on stratified random sampling with multiple characters", *Calcutta Statistical Association Bulletin*, **8**, 81-89.

Ghosh S.P. (1963): "Optimum stratification with two characters", *Ann. Math. Statistics*, **34**, 866-872

Goldfarb, D. (1969): " Extension of Davidon's variable metric method to maximization under linear inequality and equality constraints", *SIAM Journal of Applied Mathematics*, **17**, 739-764.

Gosh, D. T. (1989): "On the near optimum continuous sampling plan CSP-2 (with k=1) to minimize the amount of inspection and its performance as compared to optimum CSP-I plan", *Sankhya, Ser. B*, **51**, No. 3, 390-415.

Graves, R. L. (1967): "A principal pivoting simplex algorithm for linear and quadratic programming", *Operations Research*, **15**, 482-494.

Groves, R. (1989): "*Survey Errors and Survey Cost*", New York. Wiley.

Hadley, G. (1964): "*Nonlinear and Dynamic Programming*", Addison-Wesley.

Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1953): "*Sample Survey Methods and Theory*", New York. Wiley.

Hansen, M. H. and Hurwitz, W.N. (1946): "The problem of Non response in sample surveys", *American Statistical Association* **41**, 517-529.

Hansen, M. L., W. N. Hurwitz and W. G. Madow (1953): "*Sample Survey Methods and Theory*", **1** & **2**, New York, John Wiley and Sons, Inc.

Herleker R. K. (1967): "The problem of stratification Part-I , Investigation into some two-stage stratified sampling procedures for populations represented by probability density functions", *Skand. Akt.*, **50, 1-18**.

Hertley, H. O. (1965): "Multipurpose optimum allocation in stratified sampling", *Proceedings of American Statistical Association, Social Statistical Section*, **60**, 258-261.

Hess, I., Sethi, V. K., and Balakrishnan, T. R. (1966): "Stratification: A practical investigation", *Journal of American Statistical Association*, **61**, 74-90.

Hilderth, C. (1957): "A quadratic programming procedure", *Naval Research Logistics Quarterly*, **14**, 62-87.

Houthaker, (1960): "The capacity method of quadratic programming", *Econometrica*, **28**, 62-87.

Ige, A.F. and Tripathi, T.P. (1987): "On double sampling for stratification and use of auxiliary information", *Journal of Indian Society of Agricultural Statists*, **39**, 191-201.

Jahan, N. and Ahsan, M. J. (1995): " Optimum allocation using separable programming", *Dhaka University, Journal of Science*, **43**(i), 157-164.

Jahan, N., Khan, M.G.M. and Ahsan, M.J. (1994): "A generalized compromise allocation", *Journal of Indian Statistical Association*, **32**, No.2, 95-101.

Jahan, N., Khan, M.G.M. and Ahsan, M.J. (2001): "Optimum compromise allocation using dynamic programming", *Dhaka University Journal of Science*, **49** (2), 197-202.

Jarque C.M. (1981): "A solution to the problem of optimum stratification in multivariate sampling", *Appl. Statistics*, **30**, 2, 163-169.

Jessen, R.E. (1942): " Statistical investigation of a sample survey for obtaining farm facts", *Iowa Agr. Exp. Stat. Res. Bull.* , 304.

Jessen, R.E. (1969): "A dynamic programming algorithm for cluster analysis", *Operations Research*, 17, 1034.

Kabe, D.G. (1989): "On solving linear zero-one integer programming problems of hypothesis testing theory", *Indust. Math.*, **39**, No.1, 73-85.

Kefitz, N. (1951): "Sampling with probability proportional to size adjustment for changes in size", *Journal of American Statistical Association* **46**, 105.

Kelley, J. E. (1960): " The cutting plane method for solving convex programs", *J. Soc. Indust. Appl. Math.*, **8**, 703-712.

Khan, M. G. M., Ahsan, M. J. and Jahan, N. (1997): "Compromise allocation in multivariate stratified sampling: An integer solution", *Naval Research logistics*, **44**, 69-79.

Khan, E. A., Khan, M. G. M. and Ahsan, M. J. (2002a): "An optimum allocation in multivariate stratified sampling design using dynamic programming", *Australian and New Zealand Journal of Statistics*.(To appear).

Khan, E. A., Khan, M. G. M. and Ahsan, M. J. (2002b): "Optimum stratification: A mathematical programming approach", *Calcutta Statistical Association Bulletin*, **52**, 323-333.

Khare, B. B. (1987): "Allocation in stratified sampling in presence of non – response", *Metron*, **45** (I/II), 213-221.

Kish, L. (1967): "*Survey Samplıng*", 2nd edition. Wiley, New York.

Kokan, A.R. (1963): "Optimum allocation in multivariate surveys", *Journal of the Royal Statistical Society, Ser A*, **126**, 557-565.

Kokan A. R. and Khan, S. (1967): "Optimum allocation in multivariate surveys: An analytical solution", *Journal of Royal Statistical Society, Ser. B*, **29**, 115-125.

Kortanek, K, O. and Hoon , N. (1992): " A second order affine scaling algorithm for the geometric programming dual with logarithmic barrier", *Optimization*, **23**, No. 4. 303-322.

Kraft, O. (1970): *"Programming methods in Statistics and probability theory in nonlinear programming"*, J.B. Rosen, O.L. Mangasarian and K. Ritter , Eds., Academic Press, New York, 425-446.

Kreienbrock, L (1993): "Generalized measures of dispersion to solve the allocation problem in multivariate stratified random sampling", *Communications in Statistics*. **22**, 219-239.

Kuhn, H. W. and Tucker, A. W. (1951): "Nonlinear programming". *Proceedings of the second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 481-492.

Lahiri, D. H. (1954): " Technical paper on some aspects of the development of the sample design", *Indian National Sample Survey*, *Report 5*, repr. *Sankhya* , 14.

Lemke, C.E. (1962): "A method of solution for quadratic programs", *Management Science*, **8**, 442-453.

Li, D. (1990): " Multiple objective and nonseparability in stochastic dynamic programming", *Internat. J. Systems Science*, **21**, No.5, 933-950.

Li, D. and Hamimes , Y. Y. (1990): " New approach for nonseparable dynamic programming problems", *Journal of Optimization Theory and Applications.*, **64**, No.2, 311-330.

Liittschawager, J.M. and Wang, C. (1978): "Integer programming solution of a classification problem", *Management Science*, **24**, 1515.

Lin, C. C. (1994): " A systolic algorithm for dynamic programming", *Comput. Math. Appl.* , **27**,.No. 1, 1-10.

Malanobis, P. C. (1952): "Some aspects of the design of sample survey", *Sankhya*, **12**, 1-17.

Mandowara, P. C. (1994): " Optimum point for two or more stage designs", *Communications in Statistics Theory and Methods*, **23**, 947-958.

Mangasarian , O. L. (1969): "Nonlinear fractional programming", *Journal of Operations Research Society of Japan*, **12**, 1-10.

Markowitz, H. (1956): "The optimization of a quadratic function subject to linear constraints", *Naval Research Logistics Quaterly*, **3**, 111-133.

McCallion T. (1992): "Optimum allocation in stratified random sampling with ratio estimation as applied to the Northern Ireland December agricultural sample", *Applied Statistics*, **41**, 39-45.

Meeks, H. D. and Francis, R. L. (1973): "Duality relationships for nonlinear version of the generalized Neyman-Person Problem", *Journal of Optimization Theory Applications*, **11,** 360.

Miles, Davis and Robert H. Finch (1989): "Optimal allocation of stratified samples with several variance constraints and equal work loads over time by geometric programming", *Communications in Statistics, Theory and Methods*, **18**(4), 1507-1520.

Miles, Davis and Rudolph, E. Schwartz (1987): "Geometric programming for optimal allocation of integrated sample in quality control", *Communications in Statistics, Theory and Methods*, **16**(11), 3235-3254.

Mitra, S. K. and Pathak , P. K. (1984): "Algorithm for optimal integration of two or three surveys", *Scan. J. Statist.*, **11**, 257-263.

Murthy, M.N. (1967): "*Sampling Theory and Methods*", Statistical Publishing Society, Calcutta.

Neauhardt, J.B., Bradely, H.E. and Henning, R.W. (1973): "Computational results in selecting multi-factor experimental arrangements", *Journal of American Statistical Association*, **68**, 608.

Neyman, J. (1934): "On the two different aspects of the representative methods: the method of stratified sampling and the method of purposive selection", *Journal of Royal Statistical Soc*iety, **97**, 558-606.

Okada, S. (1993): "A method for solving multiobjective linear programming problems with trapezoidal fuzzy coefficient", *Japanease Journal of*

*Fuzzy Theory Systems*, **5**, No. 1, 15-15.

Okafor, F.C. (1994): "On double sampling for stratification with sub-sampling the non-respondents", *Aligarh Journal of Statistics*, **14**, 13-23.

Omule S.A.Y. (1985): "Optimum Design in multivariate stratified sampling", *Biometrical Journal,* **27**, 8, 907-912

Ozturk, A. (1991): "A general algorithm for univariate and multivariate goodness of fit tests based on graphical representation", *Communications in Statistics, Theory and Methods,* **20**, No. 10, 3111-3137.

Pukelsheim, F. (1978): "A quick introduction to mathematical programming with application to most powerful tests, non-negative variance estimation and optimal design theory", *Technical Report* **128**, Stanford University Press, Stanford, California.

Rahim , M. A. and Jocelyn, W. (1994): "Sample allocation in multivariate stratified design : An alternative to convex programming", *ASA Proceedings of Survey Research Methods Section,* 689-692.

Rahim, M. A. and Currie, S.(1993): "Optimizing sample allocation for multiple response variables", *ASA Proceedings of Survey Research Methods Section,* 346-351.

Raj, D. (1969): "*Sampling theory*", Mc Graw-Hill, New York.

Rao, J.N.K.(1973): "On double sampling for stratification and analytical surveys", *Biometrika*, **60**, 125-133.

Rao, M.R. (1971): "Cluster analysis and mathematical programming", *Journal of American Statistical Association*, **66**, 622.

Rao, P.S.R.S.(1986): "Ratio Estimation with subsampling the non respondents", *Surveys Methodology*, **12**, 2, 217-230.

Rosen, J.B. (1960): "The gradient projection method for nonlinear programming, Part-I: linear constraints", *SIAM Journal*, **3** , 181-217.

Rosen, J. B.(1961): "The gradient projection method for nonlinear programming Part-II: non linear constraints", *SIAM Journal*, **9**, 414-432.

Roy. A. and Wallenius , J. (1992): " Nonlinear multiple objective optimization : An algorithm and some theory", *Mathematical Programming* , *Ser. A*, **55**, No.2, 235-249.

Schoen, F. (1994): "An open optimization problem in statistical quality control", *Journal of global Optimization*, **1**, No.3, 295-303.

Seidel, W. (1991): "An open optimization problem in statistical quality control", *Journal of global Optimization*, **1**, No. 3, 295-303.

Serfling, R. J. (1967): "Approximately optimum stratification", *Journal of American Statistical Association*, **63**, 1298-1309.

Sethi, V. K. (1963): "A note on optimum stratification of population for estimating the population mean", *Australian Journal of Statistics*, **5**, 20

-23.

Shapiro, A. (1990): "On differential stability in stochastic programming", *Mathematical Programming*, **47**, 107-111.

Sheela, M. A. and Unnithan, V. K. G. (1992): " Optimum size of plots in multivariate case", *Journal of Indian Society of Agricultural Stats.*, **44**, No.3, 236-240.

Sherali, H. D. (1982): " Equivalent weights for lexicographic computations", *European Journal of Operational Research*, **11**, No. 4, 367-379.

Singh, R. and B. V. Sukhatme (1973): "Optimum stratification with ratio and regression methods of estimation", *Ann. Inst. Stat. Math.* **25**, 627-633.

Singh, R. and B. V. Sukhatme(1969): " Optimum stratification ", *Ann. Inst. Stat. Math.* **21**, 515-528.

Singh, R (1977): "A note on optimum stratification for equal allocation with ratio and regression methods of estimation", *Austratian Journal of Statistics* **19**(2), 96-104.

Staurt, A. (1954): " A simple presentation of optimum sampling results", *Journal of Royal Statistical Society, B,* **16**, 239-241.

Stock, J. S. and Frankel, L. R. (1939): "The allocations of sample among several strata", *Ann . Math. Stat.* **10**, 288.

Sukhatme, P. V., Sukhatme, B. V. Sukhatme, S. and Ashok, C. (1984):

"*Sampling Theory of Survey with Applications*", Iowa State University Press, Ames, Iowa (U. S. A.) and Ind. Soc of Agr. Stats, New delhi (India).

Tillman, F. A.: Hwang, C.L. and Kuo, W. (1977), "Optimization techniques for system reliability with redundancy-a review", *IEEE Tran. Reliability*, R-**26**, 148-155.

Todd, M. J. (1985), "Linear and quadratic programming in oriented matroids", Journal *of Combinatorial Theory, B*, **39**, 105-133.

Unnithan, V. K. G. (1978): "The minimum variance boundary points of stratification", *Sankhya*, **40**, C, 60-72.

Van de Panne, C. and Whinston, A., (1964a): "Simplical methods for quadratic programming", *Naval Research Logistics Quaterly*, **11**, 273-302.

Van de Panne C. and Whinston, A. (1964b): " The simplex and dual methods for quadratic programming", *Operations Research Quaterly*, **15**, 355-388.

Van de Panne, C. and Whinston A. (1966): " The symmetric function of the simplex method for quadratic programming", *Paper* **112**, Western Management Science Institute, University of California, Los Angeles.

Wachs, M. L. (1989): "On an efficient dynamic programming technique of F.F. Yao", *Journal of Algorithms*, **10**, No.4, 518-530.

Wang, C. L. (1990a): "Dynamic programming and inequality", *Journal of*

*Mathematical Analysis and Applications,* **150**, No.2, 528-555.

Wang, C. L. (1990b): "Dynamic programming and the Lagrange multipliers", *Journal of Mathematical Analysis and Applications*, **150**, No.2, 551-561.

Wang, C. L. and Xing A. Q. (1990): "Dynamic programming and penalty functions", *Journal of Mathematical Analysis and Applications*, **150**, No.2, 562-573.

Wei, Z. X. (1992): "A new successive quadratic programming algorithm", *Acta Math. Appl. Sinica.* (English ser.) **8**, No.3, 281-288.

Weintraub, A. and Vera, J. (1991), "A cutting plane approach for chance constrained linear programs", *Operations Research*, **39**, No.5, 776-785.

Whitaker, D., Thriggs, C.M. and John, J.A. (1990): "Construction of block designs using mathematical programming", *Journal of Royal Statistical Society, Ser B*, **52**, No. 3, 497-503.

Wolfe, P. (1959): "The simplex method for quadratic programming", *Econometrica* , **27**, 382-398.

Yadav, S. S. and Singh R. (1984): "Optimum stratification for allocation proportional to strata totals for simple random sampling scheme", *Communications in Statistics, Theory and Methods*, **13** (22), 2793-2806.

Yeh, A. (1993): " Two problems with GGP generalized geometric programming algorithm", *Comm. Numer. Methods Engg*, **9**, No.9, 767-772.

Yuan, Y. X. (1991), "A dual algorithm for minimizing a quadratic function with two quadratic constraints", *J. Comput. Math.*, **9**, No.4, 348-359.