ON SOME OPTIMUM STRATIFIED SAMPLING DESIGNS

ABSTRACT OF THE THESIS
SUBMITTED FOR THE AWARD OF THE DEGREE OF
Doctor of Philosophy
IN STATISTICS

BY
SHAZIA GHUFRAN

UNDER THE SUPERVISION OF
PROF. MOHAMMAD JAMEEL AHSAN

DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH
ALIGARH MUSLIM UNIVERSITY
ALIGARH - 202002, INDIA

2013
MANUAL OF ANALYSIS

INTENDED FOR

INSTRUCTORS IN

INDUSTRIAL CHEMISTRY

[Stamp and seal]
This thesis entitled "On Some Optimum Stratified Sampling Designs" is submitted to the Aligarh Muslim University, Aligarh, India, to supplicate the degree of Doctor of Philosophy in Statistics. It embodies of the research work carried out by me in the Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh.

The theory of probabilistic sampling is one of the topics of statistical theory that is most commonly used in many fields of scientific investigation. An effective sampling technique within a population represents an appropriate extraction of useful data which provides meaningful knowledge of the important aspects of the population. A well established sampling plan plays an important role to make the results obtained from statistical studied useful and reflect the reality. Although sampling methods are various, most frequently used method is stratified random sampling in practice, especially in case of heterogeneous population structure. As stratified sampling is the most popular sampling design, this thesis is devoted to the study of some optimization problems related to stratified sampling designs. The most important problem in stratified random sampling is to allocate the sample
Abstract

sizes to different strata. The individual stratum sample sizes, can be computed using different criteria.

When the estimation of population mean is of interest the simplest and the most frequently used allocation is the proportional allocation in which the sample size from a stratum is proportional to the stratum size. However, the best allocation, termed as optimum allocation, either minimizes the sampling variance of the estimator for a fixed cost of the survey or minimizes the cost of the survey for a fixed value of the variance of the estimator. The problem of finding the optimum allocation can be considered as a non-linear programming problem in which the objective function is the variance subject to cost restriction, or vice versa.

When the population is multivariate to work out an optimum allocation that is optimum for all the characteristics is not possible. Since the time of Tschuprow (1923) till date a large number of researchers discussed various aspects of ‘multivariate optimum allocation’, popularly known as the ‘compromise allocation’. In this thesis the problems of obtaining compromise allocations under various situations in stratified sampling designs has been critically studied.

This thesis consists of five chapters. **Chapter 1** is the usual introductory chapter. It provides an introduction to sample surveys, census, various
Abstract

sampling designs and a detailed introduction to multivariate stratified sampling with various sample size allocations. Problem of nonresponse in sample surveys is also introduced. This chapter also gives a brief idea of the application of mathematical programming techniques in sample surveys.

Most of the numerical solutions to the illustrated examples in different Chapters of this thesis are obtained by using the optimization software LINGO. An introductory reference of this software is also presented in this chapter.

In Chapter 2 the problem of determining the integer compromise allocation in multivariate stratified sampling is considered. The problem is formulated as a Multiobjective Integer Nonlinear Programming Problem (MINLPP). Using the value function technique the problem is converted into a single objective problem. Relaxing some of the conditions a formula for continuous sample sizes is obtained using Lagrange Multipliers Technique. The integer solution is obtained by rounding off the continuous sample sizes to their nearest integer values. In the later part of this chapter the problem of finding a compromise allocation for a multivariate stratified sample survey with a significant travel cost within strata is also formulated as a MINLPP. A solution procedure is proposed using the Goal Programming Technique. The numerical examples are presented to illustrate the computational details.
Abstract

The discussions in this chapter are based on my research papers as Khowaja et al. (2011a) and Ghufran et al. (2011a) published in the Journals “Communication in Statistics-Simulation and Computation” and the “South Pacific Journal of Natural and Applied Sciences” respectively.

Chapter 3 deals with the problem of optimum allocation in multivariate stratified sampling to estimate the $p-$overall population means. The objective is to minimize, simultaneously, the coefficients of variation of the estimators of $p-$population means under a cost constraint that includes the measurement as well as the travel cost. Usually some function of the sampling variances of the estimators of the $p-$population means is used as an objective function that is to be minimized for a fixed cost given as a linear function of sample allocations. Since the variances are not unit free it is more logical to consider the minimization of some function of squared coefficient of variations as an objective function. Furthermore, the investigators have to approach the sampled units in order to obtain the observations. This involves some travel cost. Usually this cost is neglected while constructing a cost function. The travel cost may be significant in some surveys. For example if the strata consist of geographically difficult areas.
Abstract

The formulated problem of obtaining an optimum compromise allocation turns out to be a MINLPP. In this chapter three different approaches are considered to solve the formulated MINLPP viz Value function approach, $\epsilon$-Constraint method and Distance based method. Additional restrictions are placed on the sample sizes to avoid oversampling and to ensure the availability of the estimates of the strata variances. Numerical examples are also presented to illustrate the computational details of the proposed methods. This work is based on my research paper Ghufran et al. (2012a) published in the journal “Communications in Statistics-Simulation and Computation”.

In Chapter 4 the optimum compromise allocation in multivariate stratified sampling with non-linear objective function and probabilistic non-linear cost constraint is considered. The probabilistic non-linear cost constraint is converted into equivalent deterministic one by using Chance Constrained programming. A numerical example is presented to illustrate the computational procedure. This work is based on my research paper Ghufran et al. (2011b) published in the “International Journal of Engineering, Science and Technology”.

In Chapter 5 the use of randomized response technique is explored. In the first part of the chapter a two-stage stratified Warner’s randomized response
model is used to determine the optimum allocation in the presence of non-response. The problem is formulated as a Nonlinear Programming Problem. A complete method of solution of the formulated problem is proposed. Two numerical examples are worked out to illustrate the computational details of the proposed method. In the second part of this chapter a two-stage stratified Warner's randomized response model is used to determine the optimum allocation in the presence of non-response and with travel cost.

In the third section of the chapter the randomized response model is extended for the multiple sensitive questions. Since social surveys are often based on questionnaires containing more than one sensitive question. The analysis of multiple RR data is of considerable interest. In multivariate stratified surveys with multiple RR data the choice of optimum sample sizes from various strata may be viewed as a MINLPP. The allocation thus obtained may be called a "compromise allocation". This section deals with the two-stage stratified Warner's RR model applied to multiple sensitive questions. The problem of obtaining a compromise allocation is formulated as Multi-objective Nonlinear Programming Problems (MNLPs) with linear and quadratic cost functions. The solution to the formulated problem is achieved through Goal Programming Technique. This chapter is based on my research papers Ghufran et al. (2012b) published in the journal
Abstract


REFERENCES


Ghufran, S., Khowaja, S and Ahsan, M. J. (2012b): Compromise allocation in multivariate stratified sample surveys under two stage randomized

