Abstract

For a phenomenon that varies over a continuous (or even a large finite) spatial domain, it is seldom feasible to observe every potential datum of some study variable associated with that phenomenon. Thus, important parts of statistics are statistical sampling theory as well as design of experiment where inference about the study variable may be made from a subset or sample of the population of potential data. The theory of sampling and design of experiment has its origin way back in the history of mankind. Special sampling refers to the sampling of geo-referenced of spatially labeled phenomena. In the spatial context, interest is usually in the prediction of the study variable at un-sampled sites. For such purpose, it is necessary to collect information regarding the population with respect to some characteristics. For example, in agricultural surveys, to estimate the production of food, the data are collected on some portion of land under different crops. Most of the government and non-government bodies collect information regularly about the total population, its geographical distribution, sex, age, age-sex etc. for future planning. In business, information is also required for the role and character of wholesale, retail and service traders etc. Given some predictand together with its predictor, a best sampling plan or network refers to the choice of locations at which to sample the phenomenon in order to achieve optimally according to a given criterion. In practice, optimal sampling plans may be extremely difficult to achieve, but atleast good sampling plans may be obtained and designed by constructing best sampling frame which can be formed by using appropriate technique of design of experiments.
The main objective of the thesis is to construct best design of sample survey and design of experiments for estimating the parameters of interest and testing the hypothesis under study respectively. It has been done by developing an appropriate procedure for selecting best sample from the population and best partially balanced incomplete block designs under the given constraints. The best estimators and best test statistic have been constructed for estimating the parameters and testing the hypothesis based on the sample having minimum extraneous variation.

Design of experiments plays a vital role in statistical research. It includes planning of experiment, obtaining pertinent information from it regarding the population under study, making a statistical analysis of the data and drawing valid inferences from it. Proper consideration of the statistical analysis before conducting the experiment, forces (stimulates) the experimenter to plan more carefully the design of the experiment. The observations obtained from a carefully planned and well designed experiment in advance having homogeneous experimental materials give entirely valid inferences over the treatments under study without introducing the complexity of design of experiment,. But in most of the practical situations, experimental material (sample) is not homogeneous. Therefore block designs are extensively used (employed) to reduce the heterogeneity among the experimental material. Block designs are used in many fields of research as these are more practicable. When block size becomes very large (large number of treatments under study) then the purpose of blocking is defeated as it is sometimes impracticable to get one complete replicate of units which is relatively homogeneous and as a result the heterogeneity is introduced in the designed experiment and reduces the discriminating power of the tests of significance. Hence the precision of the block design is adversely affected if the treatments are large in number. In order to maintain the homogeneity
within the blocks, the experimenter must either cut down the number of treatments which means loss of information over sacrificed treatments or use an incomplete block design. Therefore, in such situations incomplete block designs are used because there is no loss of information for using them. That is why, incomplete block designs are very popular and widely used in agriculture, animal husbandry and some allied fields. Among the class of binary, equi-replicated, equi-block sized designs, Balanced Incomplete Block (BIB) designs are the most efficient class of designs. But, it is difficult to develop BIB designs for all parametric combinations needed for various experimental situations. So, in such situations, partially balanced incomplete block (PBIB) designs have been generally used. PBIB designs are based on the concept of association scheme defined on treatments. First of all, we define terminology used in the thesis.

The main problem in sample survey is to estimate the population parameters of interest of characteristics under study with the desired precision and accuracy within the available resources and to develop the procedures of selection of sampling units. Survey sampling techniques are being implemented in private and government bodies like in state, industry, business, scientific institutions, public organizations and international agencies. Market research, demography and epidemiological studies are heavily dependent on the sampling approach. Survey sampling is responsible for gathering reliable information on the rates of births, deaths, population growth, incidence of diseases, on present nutritional standard, level of education and living conditions of the people for planning for the improvement of social and economic life of the people. This indicates that survey sampling plays a vital role in the maintenance and development of the economic and social welfare of people in a country. Survey sampling also most widely used in research studies carried out in various fields of sciences as well as social sciences. A sample survey may also become a necessity in dealing with characteristics where serious biases or non-
sampling errors are expected when special precautionary measures cannot be taking during collection and tabulation of data. Sampling theory plays a very important role to develop the test statistic for testing the hypothesis.

Just as in sampling theory where the interest lies in finding out the estimators (statistic) of interest based upon the sample selected, the main interest in design of experiment is to find out the best possible combination of treatments under the given conditions so that we may be able to get a group of treatments which is significantly effective as compared to any other set of treatments over a predetermined experimental area. In 1920, Fisher introduced the concept of design of experiment in which he was able to get a set of treatments which is significantly effective under the condition that the experimental units are relatively homogeneous, which gave rise to the concept of completely randomized design (CRD). The condition of homogeneity is not usually justified except in some type of laboratory experiments. To overcome this problem, first a randomized complete block design (RCBD) and later on, Latin square designs (LSD) were introduced. These are actually the designs in which blocking is complete in the sense that all treatments under investigation occur in each block of RBD and in each row and column of LSD. When the number of treatments under study is large then the size of the block/row/column increases, as a result, variation enters among the experimental units of block/row/column. Obviously, experimental units within a block will not remain relatively homogeneous. So, we have already explained that partially balanced incomplete block (PBIB) designs are generally used in such situations.
In most of the survey sampling situations, we may have information on more than one variable defined on each unit of the population, which may be highly correlated with each other. Suppose one of them is a variable under study \( y \) and information on other variables, termed as auxiliary variables. The information on auxiliary variables may be available in one form or the other or can be made available by diverting a part of the survey resources at a moderate cost. For example, in many repetitive surveys, the value of the same variable on a previous occasion say last census is usually taken as the values of auxiliary variable. In such situations, it is a widely accepted phenomenon that the efficiency of usual estimators of population parameter of interest can be increased by utilizing judiciously such information of auxiliary variables. In whatever form of the information, on auxiliary variable(s) is available, one may always utilize it to devise sampling strategies which are better than those in which no auxiliary information is used. In survey sampling, as per established facts, the information on auxiliary variables may be utilized at the following three stages:

i. At the stage of planning or designing of the survey i.e. in stratifying the population.

ii. At the stage of sample selection i.e. in selecting the units for the sample by the use of unequal probabilities sampling with probability proportional to some measure of size of the units based on auxiliary variables

iii. At the stage of estimation i.e. in the use of ordinary ratio, product, difference and regression estimators etc.

For estimating the parameters of interest, ratio/ratio-type and product/product-type estimators have been widely used when there is respective positive and negative
correlation among the variable under study and auxiliary variables. Among the three estimators, namely ratio estimator, product estimator and mean per unit estimator of population mean, one should use the

\[
\text{Ratio estimator} \quad \text{when} \quad \rho > \frac{C_x}{2C_y} \\
\text{Product estimator} \quad \text{when} \quad \rho < -\frac{C_x}{2C_y} \\
\text{Mean per unit} \quad \text{when} \quad -\frac{C_x}{2C_y} < \rho < \frac{C_x}{2C_y}
\]

Where \( C_y \) and \( C_x \) are coefficient of variation of variables \( y \) and \( x \) respectively and \( \rho \) is the correlation coefficient between them.

The present thesis have been divided into seven chapters

First Chapter is introductory in which importance of topic, definitions, terminology used and literature relating to the topic has been given in detail.

In the second chapter, for estimating the population mean \( \bar{Y} \) of the variable under study \( y \), two ratio type estimators have been proposed by making the linear transformation on auxiliary variable \( x \) which are unbiased upto first order of approximation. The expressions for the biases and variances of the proposed estimators have been obtained upto second order of approximation. It has been found that upto first order of approximation; proposed estimators are unbiased as well as equally efficient with the existing ones. The comparison of proposed estimators with the existing ones has also been made upto second order of approximation w.r.t. their mean square errors. The results have also been illustrated numerically which show that the
proposed estimators are almost unbiased and decrease in efficiency is negligible as compare to the existing ones.

In the third chapter, following the same idea of chapter II, an almost unbiased estimator of population mean has been proposed when the relationship between the variable under study and auxiliary variable is completely linear. The expressions for the bias and mean square error of the proposed estimator has been obtained upto second order of approximation and it has been shown that proposed estimator is unbiased as well as equally efficient with the regression estimator upto first order of approximation. The comparison of proposed estimator with the regression estimator has also been made upto second order of approximation w.r.t. their biases and mean square errors. The results have also been illustrated numerically.

In the fourth chapter, starting from the association matrices of a group divisible (GD) partially balanced incomplete block (PBIB) design with two associate classes, we obtain some new partially balanced incomplete block designs with two, three and four associate classes following certain juxtaposition patterns by using some association matrices and some other types of well known matrices. For the purpose of completeness, association scheme of group divisible design and one new association scheme have been discussed. Methods of construction of regular group divisible (RGD), singular group divisible and designs with three associate classes following new association scheme have also been discussed which are supported by illustrations.

Table 4.5.1 lists parameters of the new regular group divisible designs constructed from Theorem 4.4.1 for the values of the parameter $r^*, k^* \leq 20$ & $v^*, b^* \leq 25$; Table 4.5.2 lists parameters of the new singular group divisible designs constructed from Theorem 4.4.2 for the values of the parameter $r^*, k^* < 11$ & $v^*, b^* \leq 25$, Table- 4.5.3 lists parameters of the three associate class
partially balanced incomplete block designs constructed from Theorem 4.4.3, Theorem 4.4.4, Theorem 4.4.5 and Remark 4.4.5 for the values of the parameter \( r^*, k^* \leq 10 \) & \( v^*, b^* \leq 20 \) and Table-4.5.4 lists parameters of the four associate class PBIB designs constructed from Theorem 4.4.6 for the values of the parameter \( r^*, k^* \leq 20 \) & \( v^*, b^* \leq 40 \). In Table-4.5.3, the column “Application” means the theorem number from which the design has been obtained. The symbol ‘E’ denotes the average efficiency factor of the partially balanced incomplete block designs with respect to randomized block design (RBD), ‘E;’ denotes the efficiency factor of the treatment contrasts which are \( i^{th} \) \( (i = 1, 2, \ldots, m) \) associates with respect to randomized block designs. The efficiencies of various treatment contrasts and average efficiency factor have been calculated for the purpose of comparison.

In the fifth chapter, following the work done by Ramakrishnan (1956) and R.N. Mohan (1983), some new series of triangular and four associate class PBIB designs with two replications have been constructed by using dualization technique. The efficiencies of the constructed designs have been calculated for the purpose of comparison. The effort has also been made to illustrate the results numerically.

In the sixth chapter, application of Design of Experiment in Sampling has been discussed. For estimating the population mean, efficient sampling strategy has been proposed under controlled sampling design. Special case of partially balanced incomplete block design is used for preparing the sampling frame which gives quite high probability of selection of preferred samples as compared to non-preferred sample. The expressions for bias and variance of proposed sampling strategy have been obtained. It has been shown that efficiency of the estimator under the given design remains the same as in the case of simple random sampling design but the
probability of selection of non-preferred samples becomes less in the proposed sampling design as compared to the conventional simple random sampling design.

In the seventh chapter, a study of the primary hypertension among the euglycemic patients of ischemic heart disease has been done. The work is related to the prediction of the hypertension. So, two separate multiple linear regression models have been used by taking Systolic Blood Pressure (SBP) & Diastolic Blood Pressure (DBP) as dependent variables and body mass index, tri-glyceride, low density lipoprotein-cholesterol and high density lipoprotein-cholesterol as independent variables (independent risk factors) to study the essential hypertension among the euglycemic patients of ischemic heart disease. Adequacy of the models has been verified using analysis of variance (F-test), histogram, Q-Q plot and box whiskers plot. In this chapter, an effort has been made to analyze the data and interpret it with the help of statistical software package SPSS 14.0.