Introduction:

1.1 Introduction Of the Problem Under Study and Preliminaries:

Research begins with some population which we wish to study. Sometimes the population is so small that we can simply study it all. For example, all living former president of the United states. In such cases we can just collect data from each member of the population without any need of sampling. And, of course, our results will be as accurate a description of the population as our data gathering method allows. However in most cases the population is so large that we are prevented by considerations of time or effort from examining ever individual. In such cases we often turn to sampling. Nowadays, Our knowledge, our attitude and our actions are based to a very large extent on sample-based estimates. This is equally true in every day life and scientific research [Cochran (1977)]. Therefore sample survey are widely used as a means of gathering information in government, industry and trade, the investigation of social, psychological, educational and health problems, physical and life sciences and technology, and nowadays even in humanistic studies such as history and archeology, study of languages, literature, religion and the arts. Deming (1960) and Slonim (1960) have given many interesting example of applications of sampling methods in business. Neter (1972) has described the techniques of saving money by airlines and railways by using samples of records and estimates based on them to apportion income from freight.
and passenger services. Moreover, statistical techniques and theory for their design and analysis has at least partial applicability to investigation which are not usually referred to as survey such as quality control and inspection, auditing, experimental and computer simulation.

To estimate the amount of lichen available as food for caribou, a biologist collects lichen from selected small plots within the small study area. Based on the dry weight of these specimens the variables biomes for the whole region a few (highly expensive) sample holes are drilled. The situation is similar in a national opinion survey, in which only a sample of the people in the population is contacted, and the opinions in the sample are used to estimate the proportion with the various opinions in the whole population. To estimate prevalence of patients treated. To estimate the abundance of a rare and endangered birds species, the abundance of birds in the population is estimated based on the pattern of detection from a sample of sites in the study region.

The obvious questions are: how best to obtain the sample and make the observations and, once the sample data are in hand, how best to use them to estimate the characteristic of the whole population after obtaining the observations involving questions of sample size, how to select the sample, what observational method to use, and what measurements to record.

Getting good estimates with observations means picking out the relevant aspects of the data deciding whether to use
auxiliary information in estimation, and choosing the form of the estimator [Thompson (1992)].

The estimates based on sample are always subject to some uncertainty because only a part of the population has been measured and because of error of measurement. This uncertainty can be reduced by taking large samples, superior instruments of measurement and better sampling techniques.

For getting a better sample, several sampling schemes have been proposed in literature [see Cochran (1977), Jessen (1978), Kish (1965), Konjin (1973), Mukhopadhyay (1998), Raj and Chandak (1998), Sampat (2001), Sarandal et al. (1992) Sukhatme et al.(1992)]. A collection of identifiable units is known as population, e.g. collection of households in an area where households where as a batch of electric light bulbs. Samples are generally collected from a population by random selection where each unit in the population, has an equal chance of being included in the sample. This is simplest sampling scheme, also known as simple random sampling (SRS). The method of randomization, known as “DYUTVIDYA”, was well used in ancient India in Mahabharat era where Pandavas and Kauravas used to play Gambling. In 72th PARVA (Chapter) of Mahabharat, a character claims to count, exactly, the number of fruits in the branches of a tree with the help of this technique. Whenever the unit for including in the sample, the technique is called simple random sampling with replacement (SRSWR)
otherwise is called simple random sampling without replacement (SRSWOR). Estimates based on SRSWOR are found to be more efficient than SRSWR if population is not large, otherwise both methods give estimate with equal precision.

In case list of elementary units in the population is not available, systematic sampling is adopted where instead of selecting n units at random; the sample units are decided by a single number chose at random. This procedure is well used in catch of fish and forest surveys for the estimation of timbers. Whenever the units have significant variation with respect to their values, probability proportional to size (PPS) sampling is adopted which utilizes additional information (variable) called size variable. In this regard cumulative total method and Lahiri (1951) method are well used to choose an elementary unit in the sample. This sampling scheme is highly used in area survey where area of fields is available and investigator is interested in the estimation of total yield of a crop. A review of PPS sampling is given by Brewer and Hanif (1983).

If the elementary units of the population are heterogeneous, the units are first divided into homogeneous groups and then the sampling units are selected from each group either by SRS, Systematic of PPS sampling scheme. Such a method is known as Stratified sampling and is due to Neyman (1934).

After collecting data by an appropriate sampling scheme, sample estimate $\hat{\mu}$, (a function of sampling units) is obtained
to estimate the population parameter $\mu$ (a function of population units). In practice parameters like mean, variance, coefficient of variation and ratio of two population means (or total) are estimated. In sample survey criterion of a good estimate are

(i) Unbiasedness i.e. if $E(\hat{\mu}) = \mu$

(ii) Consistency i.e. if $\hat{\mu} \rightarrow \mu$ as sample size $n \rightarrow$ population size $N$.

(iii) Efficiency i.e. for two estimates $\hat{\mu}_1$ and $\hat{\mu}_2$ of $\mu$, $\hat{\mu}_1$ is considered more efficient than $\hat{\mu}_2$ if

$$E(\hat{\mu}_1 - \mu)^2 < E(\hat{\mu}_2 - \mu)^2 \quad (1.1.1)$$

If $E(\hat{\mu}) = \mu$, the expression of deviation of estimator $\hat{\mu}$ from $\mu$ for all sample sizes i.e. $E(\hat{\mu} - \mu)^2$ is called variance of $\hat{\mu}$ and is denoted by $V(\hat{\mu})$ otherwise it is called mean squared error (MSE) and may be denoted by $M(\hat{\mu})$. The bias of $\hat{\mu}$ is defined as

$$B(\hat{\mu}) = E(\hat{\mu}) - \mu \quad (1.1.2)$$

A relation between bias and MSE of $\hat{\mu}$ is given by

$$M(\hat{\mu}) = V(\hat{\mu}) + B^2(\hat{\mu}) \quad (1.1.3)$$

Since bias gives the weighted average of the difference between the estimator and parameter, and mean square error gives weighted squared difference of the estimator of the parameter, it is always better to choose an estimator which has smaller bias (if possible unbiased) and lesser mean square error.
However, in sample survey theory sometimes it becomes necessary to consider biased estimators for two reasons [Cochran (1977, pp 12)].

(i) In some of most common problems, particularly in the estimation of ratios, estimators that are otherwise convenient and suitable found to be biased.

(ii) Even with estimators that are unbiased in probability sampling, error of measurement and non-response may produce biases in numbers that we are able to compute from the data.

Hurwitz and Thompson (1952), Godambe (1955, 1960) have given the estimators of population mean based on order of selection of a particular unit. These estimators include all the estimators based on any sampling design. In a storming example Basu (1971) has shown Horwitz Thompson estimator less efficient. But in general all these estimators including the study by Basu (1958), Koop (1957, 1963), Cassel, Sarandal and Wretman (1977) have provided firm base to the philosophical and theoretical foundations of sampling theory from finite population and estimation of some population parameters.

1.2 Use of Auxiliary Information

In surveys, it is sometimes possible to measure certain characters, other than the character under study, which are highly correlated with the study character. The additional information, thus obtained is known as the ancillary information and the character on which this information is obtained are known as auxiliary or auxiliary character. In most
of the surveys, it has been observed that auxiliary information, if used intelligently; may provide sampling strategies (a combination of sampling scheme and estimator) better than those in which no such information in used. The emphasis lay and the use of auxiliary information for improving the precision of estimates is chief characteristic of sampling theory. It is some times, possible to measure certain characters, other than the character under study, which are highly correlated with the study character.

An auxiliary variable assists in the estimation of the study variable. The goal is to obtain an estimator with increased accuracy. Some sampling frame are equipped from the out set with one or more auxiliary variables through simple numerical manipulations i.e.; the frame provides not only the identification characteristics of the units, but attached to each unit is also the value of one or more auxiliary variables.

Basically, the use of auxiliary information in probability sampling can be made either at the designing stage or at the estimation stage. Further, the use of auxiliary information at the designing stage can be made either for constructing the strata, formation of strata or determining the probabilities for selecting the sample. At the estimation stage, the ancillary information can be used for obtaining some estimators, which under certain conditions are, known to be more efficient than the estimators based on simple random sampling. The use of ancillary information may also be used in mixed ways; for example, a register of rearms may contain information about
the area of each farm, a list of districts may contain information about the number of people living in each district at the time of the latest population is an often used sampling farm in surveys of individuals or households. The register contains some quantitative and some categorical variables, among the former are age (abstained from the date of birth) and taxable income. Categorical auxiliary variables available in the register are sex, marital status, and residential district. The information on a character z may be used in defining the set of inclusion probabilities while that on another checker x may be used in constructing some efficient estimators for certain population parameters.

The efficiency of the procedure with the use of auxiliary information heavily depends upon the way in which the estimator has been proposed, that is, the form in which the estimator has been taken into account. The use of weighted means and ratio of weighted means of function of auxiliary character are the commonly used devices in most of the proposed estimators. Optimality of weights determines the final form of the estimator to be used in order to provide least mean square error or variance.

1.3 (a) Estimation of population mean

Let y be character under study and x be auxiliary character, \( \bar{y} \) and \( \bar{x} \) be the sample means of the auxiliary character x, then for estimating population mean \( \bar{Y} \) of character y, Cochran (1942) defined a ratio type estimator
the aim of this method is to use the ratio of sample means of two characters which would be almost stable under sampling fluctuations and thus, would provide a better estimate of the true value. It has been a well-known fact that $\bar{y}_r$ is more efficient than the sample mean estimator $\bar{y}$, where on auxiliary information is used, if (1.3a.2) the coefficient of correlation between $y$ and $x$, is greater than half of the ratio of coefficient of variation of $x$ to coefficient of variation of $y$, i.e., if

$$\rho > \frac{C_x}{2C_y} \quad (1.3a.2)$$

Thus, if the information on auxiliary variable is either readily available or can be obtained at no extra cost and it has a high positive correlation with the main character, one would certainly prefer ratio estimator to simple mean estimator. Using knowledge of $C_x$, Sisodia and Dwivedi (1981), Pandey and Dubey (1989a) proposed modified ratio estimator which in more efficient than usual ratio estimator.

Contrary to the situation of ratio estimator, if variables are negatively correlated, then the product estimator

$$\bar{y}_p = \bar{y}(\frac{x}{X}); \quad \bar{X} = 0(1.3a.3)$$

has been proposed by Robson (1957), Murthy (1964) and also considered by Goodman (1960), Srivastava (1966), Wu and
Chung (1981) among others. The estimator $\bar{y}_p$ is observed to give higher precision than the sample mean estimator $\bar{y}$ under the condition

$$\rho > -\frac{C_x}{2C_y}$$

The expressions for bias and mean square error of $\bar{y}_r$ and $\bar{y}_p$ have been derived by Cochran (1942) and Murthy (1964) respectively, which are also available in the books by Murthy (1977), Cochran (1977), Jessen (1978), Sukhatme et al (1992) and Mukhopadhyay (1998). Using the knowledge of $C_x$, Pandey and Dubey (1987, 1989b) proposed modified product estimator in simple random sampling.

In case relation between $y$ and $x$ is linear, the regression method of estimation is used for determining $\bar{Y}$. In this method, the estimator is defined by

$$\bar{y}_{1r} = \bar{y} + b(\bar{X} - \bar{x}),$$

(1.3a.5)

where $b$ is an estimate of change in $y$ if $x$ is increased by unity. Watson (1937), Yates (1960) and others have used this method of estimation in practical situations. By situations. By suitable choices of $b$, the regression estimate includes all the mean per unit estimate, the ratio estimate and the product estimate as particular cases. If $b$ is a pre-assigned constant $b_0$, $\bar{y}_{1r}$ is unbiased. (In this case no assumption is required about the relation between $y$ and $x$ in the finite population).
The best value of $b_0$ that minimizes the variances of $y_{1r}$ is given by

$$
\beta = \frac{S_{yx}}{S_x^2} \quad (1.3a.6)
$$

Which may be called the linear regression coefficient of the line

$$
y = \beta_0 + \beta x \quad (1.3a.7)
$$

in the finite population. Such estimator is known as difference estimators and was first proposed by Hansen, Hurwitz and Madow (1953). If $b_0$ is computed from the sample, an effective estimate is likely to be familiar least square estimate of $B$, i.e.

$$
b = \frac{s_{yx}}{s_x^2} \quad (1.3a.8)
$$

The theory of linear regression plays a prominent role in statistical methodology. The standard results of this theory are not entirely suitable for sample surveys because they require the assumption that population regression of $y$ and $x$ is linear, that the residual variance of $y$ about regression line is constant, and the population is infinite. If the first two assumptions are violently wrong, a linear regression estimator will probably not be used. However, in surveys in which regression line of $y$ on $x$ is thought to be approximately linear, it is helpful to use $\bar{y}_{1r}$ without assuming exact linearity or constant residual variance (Cochran 1977, pp 193). The
regression estimate is, in large samples, more precise than all the estimators.
Rao (1969) found, in a Monte Carlo study, that regression estimates becomes inefficient than the ratio estimate for small sample sizes. If the scatter diagram of the sample values of \( y \) on \( x \) exhibits a linear trend, then the regression estimate will have least bias. The study on regression estimate was extended by Raj (1965) and Srivastava (1967), Tripathi (1970), Das and Tripathi (1981), Chaubey et al. (1984), Dubey (1988) and others.

In some situations of practical importance, the information on more than one auxiliary character correlated with the study variable \( y \) is available. To cope with such situations, Olkin (1958), Sukhatme and Chand (1977, 1978) proposed weighted multivariate ratio estimators. In case all the auxiliary variables are negatively correlated, Shukla (1966) and Srivastava (1966a) proposed weighted multivariate product estimators. Srivastava (1966b), Singh (1967), Rao and Modhulkar (1967), Sahal et al. (1980), Kothwala and Gupta (1989), extended the idea in case some auxiliary variables are positively correlated and others negatively correlated. John (1969) considered an alternative multivariate generalization of ratio and product estimators, applicable to any sampling design. If \( y \) and \( x \) have curvilinear relationship, Singh et al (1980) proposed a non-linear estimator and expressed efficiency of the estimator in terms of intra-class correlation coefficient.

12
1.3a Bias adjustment in the estimators

Since the ratio, product and regression estimators were observed to be more precise than the usual sample mean estimator under different conditions, several researchers diverted their attention in the direction of modifying estimation procedure so that less biased or unbiased estimators can be obtained. The work of Hartley and Ross (1954) deserves special mention in this direction. They proposed a bias adjusted ratio estimator, which is unbiased under simple random sampling without replacement. Robson (1957) obtained the exact expression of variance of this estimator and later on, Goodman and Hartley (1958) furnished the condition under which the unbiased estimator will be preferred over $\bar{y}_r$. Several other unbiased ratio-type estimator have also been proposed. Murthy and Nanjamma (1959), Rao (1966, 1967, 81a, 81b) proposed almost unbiased ratio estimator by weighting separate and combined (usual) type ratio estimator. Beale (1962) and Tin (1965) proposed bias adjusted ratio estimator, which are equally efficient in finite population. Sahoo (1983, 1987, 1995), Pandey and Dubey (1989c) also suggested some bias adjusted ratio estimators.

A skillful device for bias adjusted ratio estimator was proposed by Quenouille (1956). The technique consists of splitting the sample into $g$ sub-samples of equal size and define a weighted estimator based on sub-samples and the entire sample. This technique is also known as Jackknife-technique (Tukey, 1958). The problem of optimum value of $g$
was discussed by Rao and Rao (1971) under a super-population model for the ratio estimator. In order to compare the different methods of the bias reduction in ratio estimator, a number of Monte Carlo studies under various super-population models have been performed by Hutchison (1971), Rao and Beegle (1967), Rao and Kuzik (1972), Tin (1965) and others.


Micky (1959) proposed an unbiased regression estimator, which is based on splitting of the sample into $g$ groups. Williams (1963) also proposed an unbiased regression estimator. The empirical study of Rao (1969) showed that Mickey's unbiased regression estimator is defined inferior to the classical regression estimator when the samples are of small sizes. Mickey (1959) also developed a general theory for the construction of unbiased ratio-type estimators in simple random sampling without replacement, using information in
the population mean of several auxiliary variables. Shastry (1964) made a critical study of Mickey's works and attempted some of the problems not covered by him.

1.3.a (ii) Modification through generalization

In recent past, a number of modified ratio and product estimators came into existence. The fact that ratio or product estimators have superiority over sample mean estimator only when the correlation between the study and auxiliary variable is positively or negatively high, let the researchers to think over modifying such estimators so that the modified estimators can work efficiently even if correlation is low. Such modifications (or generalizations) are usually made by introducing some unknown constants in the estimator or by mixing two or more estimators out of sample mean, ratio and product estimators with unknown weights. The optimum values of the unknown constants are then determined by minimizing the mean squared error of the estimator, which generally depend upon the population parameters. Various authors including Adhvaryu and Gupta (1983), Chakraborty (1967), Tripathi (1970), Srivastava (1967, 1971, 1981), Gupta (1971,1978), Ray and Sahai (1978, 80), Vos (1980), Shah and Patel (1984), Das and Tripathi (1981), Singh and Upadhyaya (1986), Dubey (1988), Shukla (1989), Biradar and Singh (1992-93), Dubey and Singh (2001), Dubey and Kant (2004b) and others have suggested modified ratio or product type
estimators and studied their properties theoretically and empirically.

In the sequence of suggesting modification over ratio and product type estimators, Bandopadhyaya (1980) and Srivenkataramana (1980) proposed dual to ratio estimators applicable for the case of positive correlation. Such estimators are more précised than $\bar{y}$ if sample size is smaller than half or the population size. Dubey and Kant (2004b) extensively and found optimum sample size for dual to ratio estimator, which may be greater than half of population size. For this choice of sample size, dual to ratio estimator attains maximum efficiency as compared to its efficiency for any other sample size. But for this optimum choice of sample size, dual to ratio estimator becomes less efficient as compared to regression estimator.

Due to the dependence of the values of unknown constants upon population parameters used in the modified estimators researchers tried to locate some parameters, which could be used their guessed values in the development of the estimators. One such quantity is $pC_y / C_x = \beta \bar{X} / \bar{Y}$. Several authors have suggested that its value can be guessed accurately from a pilot surveys, past experience or repeated surveys. Reddy (1978) has shown that it does not fluctuate much in repeated surveys and, therefore, could be guessed from past data. Similarly, in many biological, agricultural and earth science problems, the values of the shape parameter such
as coefficient of variation, skewness and kurtosis are easily available.

**1.3a (iii) Modification through transformation on variables**

Walsh (1970) considered a ratio-type estimator with the use of weighted mean of $\bar{x}$ and $\bar{X}$ in place of $x$ in the usual estimator and shown that the estimator when some guessed value of $K$ is known. Mohanty and Das (1971) used transformed auxiliary variable $x'$ in place of $x$ on the fact that if regression line of $y$ on $x$ is

$$y = a_0 + a_1 x$$

then regression line of $y$ on $x'$

$$y = a_1 x$$

passes through origin where $X' = \frac{a_0}{a_1} + X$. Kulkarni (1978) found that bias of usual ratio estimator reduces by using such a transformation Reddy (1974) has considered similar estimator for population total by making transformation on the auxiliary character. Dubey (1988) studied transformed ratio and product type estimators and investigated several estimators better than ratio, regression and product estimators. Srivenkataramana and Tracy (1979, 1980, 1981, 1983) considered some useful transformations for defining ratio and product-type estimators.
1.3b Estimation of population variance

The use of an auxiliary variable $x$ in the estimation of the finite population total or mean of a characteristic $y$ is a common occurrence in practice. Ratio, difference, and regression estimators utilize an auxiliary variable for more efficient estimation of the parameter in question. Such estimators take advantage of the correlation between the $x$ variable and the characteristic $y$. In many situations, the problem of estimating population variance $\sigma_y^2$ may be of considerable importance. This problem has drawn the attention of some researchers. Wakimoto (1971) gave unbiased estimators of population variance based on a stratified random sample. Liu (1974) considered the problem of estimating population variance in general set-up. Das and Tripathi (1978) suggested various estimators of population’s variance $\sigma_y^2$ of variable $y$ using information. Later Srivastava and Jahjii (1980, 1995), Isaki (1983) Upadhyay and Singh (1983), Singh et al (1988) Biradar and Singh (1994, 1998), Garcia and Cerbrian (1997), Singh and Singh (2001) considered the problem of estimating population variance using auxiliary information Singh et al. (1973) and Searls and Intarapanich (1990) proposed Searls (1964) type estimator of population variance for $\sigma_y^2$ in case of normal populations. Dubey and Kant (2001) suggested an estimator, which is more
efficient than all the above estimators if population variance \( \sigma_x^2 \) of auxiliary variable \( x \) is closer to \( \sigma_y^2 \).

In a similar manner, than, it seems reasonable that under suitable conditions efficient estimation of the variance of finite population total or mean of the characteristic \( y \) is also possible using such estimation techniques. Fuller (1970) proposed a regression estimate of the variance of the Horvitz-Thompson estimator of the population total using as \( X \) variables the quantities \( (\pi_{ij} - \pi_{ij}) \) and \( (\pi_{ij} - \pi_{ij})(i - j)^2 \), where \( \pi_i \) and \( \pi_{ij} \) denote the individual and joint inclusion probabilities, respectively. A small numerical example illustrated a significant reduction in the variance of the estimator of variance over the use of the Yates-Grundy estimator. Ogus and Clark (1971) proposed the use of ratio or difference estimators of the variance under a Poisson sampling design (a design in which each sampling unit is given an independent chance of being selected into the sample without replacement) for the purpose of reducing the effect of the random sample size on the variance estimator. Under a sample design in which one sample unit is selected in each stratum with probability proportional to size (PPS), Hansen, Hurwitz, and Madow (1953) proposed the use of a correlated variable in conjunction with a collapsed stratum variance estimation technique and showed that the resulting estimator was approximately positively biased. Under the same
sampling design, Hartley, Rao, and Kiefer (1969) proposed a variance estimator based on an assumed good regression fit between the true stratum means and some auxiliary variables. Their examples, using a single auxiliary variable, indicated considerable improvement in terms of absolute bias over the estimator proposed by Hansen, Hurwitz, and Madow (1953). However, the Hansen et al. method has the advantage over the Hartley et al. method of ease of applicability as the latter requires matrix inversion. In addition the Hartley estimator has not been shown to be nonnegative under all conditions. Recently, Shapiro and Bateman (1978) considered reducing the bias of the estimator of the variance in a one-per-stratum design by using as a variance estimator the Yates-Grundy variance estimator for a two sample unit per stratum design with joint inclusion probabilities, \( \pi_{ij} \), calculated on the basis of Durbin’s (1967) sampling scheme.

1.3c Estimation Of Population Coefficient Of Variation, Correlation And Regression:

The problem of estimating the population coefficient of variation \( C_Y = \sigma_Y / \bar{Y} \) is also of considerable importance in practice in many situations, where the variability and stability among \( y \)-values, e.g. dispersion per unit mean in the population is of interest to study. The problem of estimation of
$C_y$ is also related to the estimation of sampling error as it appears in the expression of means square error. The knowledge on a parameter of and auxiliary character $x$ is also utilized for simultaneous estimation of the various parameters of the principal character $y$, including $C_y$. Das and Tripathi (1981a, 1992) proposed several ratio and product type estimators of $C_y$. Tripathi et al (2001) proposed a generalized class of estimators.

This problem, using auxiliary information was considered by Srivastava et al. (1986), Dubey (1988) and generalized by Singh and Singh (1988), and Zaidi et al (1999).

1.4 Estimation In Double Sampling:

The usual ratio, regression and product estimators defined in (1.3a.1), (1.3a.3) and (1.3a.5) are based on the knowledge of population mean $\bar{X}$ of auxiliary variable $x$. In the practical situations, where $\bar{X}$ is not known, the method of double sampling or two phase sampling is adopted. This technique. The technique suggests to select in expensive information on one or more auxiliary variable of large size by simple random sampling. With the aid of the auxiliary information collected in the first phase, select a second phase sample by simple random sampling. The study variable $y$ is than observed for the elements in the second phase sample. The technique is called two-phase sampling (or some times
double sampling). Cochran (1963) has mention that in some application, if it is convenient, the second sample may be drawn independently of the first. Other earlier discussants of double sampling technique where Robson (1952), and Robson and King (1953). The way of selecting the second-phase sample varies from situation to situation. The following two methods are generally adopted:

(I): The second-phase sample is a sub-sample of the first-phase sample,

(II): The second-phase sample is drawn independent of the first-phase sample. [Bose (1943)].

A key to successful two-phase sampling is the creation of highly informative frame, not for the whole population (this may be too expensive) but for the part of the population from which the sub-sample is drawn. An addition reason for studying two-phase sampling is that the theory is useful for estimation in the presence of non-responses. In a survey with non-response, the selection of a probability sample can be seen as the first phase, and the respondents are viewed as a sub-selection.

Naturally, the double sampling results in decreasing the size of the main sample. This may not be a good proposition in many situations particularly when the collection of information on auxiliary characteristics is not relatively cheaper. In fact, double sampling sometimes may not be feasible at all, for instance, in biological experiments where observing X and Y may call for the destruction of sampling
unit. Still when information on auxiliary variable is lacking, double sampling is used frequently. Srivastava (1970, 71) generalized the ratio estimator considering both the above cases in exhaustive manner. Rao (1975) took into account the idea of sufficiency proposed by Pathak (1964) in order to discuss ratio estimator. He also made a vital contribution Rao (1972) by discussing regression estimator in two-phase sampling and taking into account of the distinct units present in the second sample when it has not been drawn independently from the first. Gupta (1978) studied the quadratic order of the ratio estimator under simple random sampling and double sampling. Srivastava (1981) proposed a general class in the form of function of sample means. Bedi (1985) extended the existing multi-variate regression estimator into Rao (1975). Some other fruitful contributions in two-phase sampling are due to Khan and Tripathi (1967), Sengupta (1981), Kawathekar and Ajgaonkar (1984), Singh (1986), Pandey and Dubey (1989d) etc.

Assuming two phase, Sukhatme (1962) considered a Hartley-Ross type two-phase estimator. De graft-Johnson and Sedransk (1974), Rao (1981) considered modified versions of the estimators of Beale (1962), Pascual (1961) and Tin (1965), by replacing \( \bar{x} \) by \( \bar{x}' \), the sample mean based on first phase sample, in the estimator. Rao (1981) also considered the Jackknife two-phase version of the classical single-phase ratio estimator.
where final sample of size $n$ is divided into of groups. [See Mukhopadhyay (1998).] Tripathi (1969) discussed regression estimator in double sampling where first phase sample is simple random sampling and second phase by probability proportional to size sampling. In contrary method of selection, Mukhopadhyay (1998) has discussed difference type estimator.

In case population mean $\bar{X}$ and population variance $\sigma_X^2$ of auxiliary variable $x$ are unknown, Das and Tripathi (1989) used double sampling technique for estimating population mean and variance of study variable $y$. Some times even if $\bar{X}$ is unknown, information on a cheaply ascertainable variable $z$, closely related to $x$, is available on all units of the population. By analogy, if $z$ has a high positive correlation with $x$, the ratio estimator

$$\bar{x}_r = \frac{\bar{x}'}{\overline{z'}} \overline{z}$$

will estimate $\bar{X}$ more precisely than $\bar{x}'$. Thus, using $\bar{x}_r$ for $\bar{x}'$ in double sampling ratio estimator,

$$\bar{y}_dr = \bar{y} \frac{\bar{x}_r}{\bar{x}}$$

Chand (1975) developed a chain type ratio estimator,

$$\bar{y}_dr = \bar{y} \frac{\bar{x}_r}{\overline{z'}} \overline{z}$$

Kiregyera (1980) proposed generalized ratio estimator, Further improvement on this type of estimator has been made

1.5 Estimation In Presence Of Non-response:

It is a common experience in surveys that data cannot always be collected for all the units selected in the sample. Thus, the selected farmers or families may not be found at home at the first attempt and some may refuse to cooperate with the interviewer even if contacted. This experience is particularly true in mail surveys, where in questionnaires are sent to a sample of respondents and all of them are requested to send back their returns by some deadline. Many respondents do not reply and the available sample of returns is incomplete. This incompleteness, called non-response, is sometimes so large as to completely vitiate the result.

The problem of non-response is particularly pronounced in a survey with a very low response rate, in which the probability of responding is related to the characteristics to be measured magazine, readership surveys of sexual practices exemplify the problem. The effect the non-response problem may be reduced through additional sampling effort to estimate the characteristics of the non-response stratum of the population by judicious use of auxiliary information available on both responding and non-responding units, or by modeling of the non-response situation. But perhaps the best advice is to keep non-response rates as low as possible.
Because of its importance in practice, the problem of incomplete samples has received considerable attention, and several methods for recovering information from the non-respondents are available. As a result of very few low response rates in many mail surveys, estimates obtained from such surveys are unreliable. The main advantage of mail surveys is their low cost. Hansen and Hurwitz (1946) were first to deal with the problem of incomplete sample in mail survey. They proposed the following technique which is useful in obtaining unbiased estimators: (i) select a sample of respondents and mail a questionnaire to all of them (ii) after the deadline is over, identify the non-respondents and select a sub-sample of non-respondents (iii) collect data from the non-respondents in the sub-sample by interview and (iv) combine data from the two parts of the survey to estimate population values concerned. El-Badry (1956) has extended Hansen and Hurwitz’s technique. Repeated waves of questionnaires are sent to the non-responding units. As soon as a point is reached when further waves will not be effective, a sub sample from the remaining non-responding units is selected and interviewed. The final estimates are based on the pooled data from all the attempts put together. Foradori (1961) has generalized El-badry’s approach and has also studied the uses of Hansen and Hurwitz’s technique under different models.

Dalenius (1955) has considered the possibility of selecting a sub-sample of non-responding units during the course of the fieldwork for building the frame itself. The
technique helps to reduce the time lag between the initial mail survey and the subsequent interview survey.

Kish and Hess (1956) assumed that a record of non-responding units is kept from the past surveys. A sample of these is added to the units actually selected in the sample. In the course of fieldwork, data are collected for both original sample and the units added from the previous non-responding stratum. The latter part of the information is then used to estimates the contribution of non-responding units to the original samples.

Bartholomew (1961) has developed a two-call technique of arriving at unbiased estimates in interview surveys. The assumptions is that enumerators are able, in the course of first call, to arrange for the second call in such away that the units missed in the first call have the same probability of being contacted in the second call. The units actually contacted in second call, form a sub-sample of the units missed in first call. Data collected in the two visits make it possible to get unbiased estimates.

Hendricks (1956) has given an excellent example of reducing the bias through successive call in mail survey of farms. Three waves of questionnaires were sent out in secessionist regular intervals. The average number of trees per farm was calculated after each call. Even after three calls the non-response was 60 present and adjustment for bias was obviously necessary. As the characteristics under study and rate of non-response were correlated he used regression
method to arrive at an estimate of 344 trees per farm as against a true value of 329, which was known. This estimate was much closer to the value than any one directly obtained from the survey.

Srinath (1971) has proposed a rule for selecting a sub-sample of non-respondents under which the sub-sampling rate is not kept constants, but varies according to the sample non-response rates. Using this sampling rule the variance of the estimator of the population, mean is independent of the unknown rate of non-response in the population. He also modified El-badry’s method in an analogous manner.

**1.6 Study Under Superpopulation Model:**

Two main types of inferences are considered in the survey of literature, design-based and model-based inference. In design based inference, the finite population is considered fixed and the variable value are fixed. In model-based inference, the values of the variable in the finite population are assumed realizations from superpopulation model.

Assume that the value of \( y \) on \( i \) is a particular realization of a random vector \( Y = (Y_1, \ldots, Y_N) \) having a super population distribution \( \xi_\theta \), indexed by a parameter vector \( \theta \), \( \theta \in \Theta \) (the parameter space). The class of priors \( \{\xi_\theta, \theta \in \Theta\} \) is called a superpopulation model. The model \( \xi \) for \( Y \) is obtained through one's prior belief about \( Y \). As an example, in agricultural survey of yield of crops in \( N \) plots, if acreages \( x_1, \ldots, x_N \) under the crop on these plots are assumed fixed over
years, one may assume that the yield $y$ in a particular year is a random sample from a prior distribution $\xi$ of $Y$, which may depend among others, on $x_1, \ldots, x_N$.

The use of an appropriate superpopulation model distribution in survey sampling is justified by the fact that in surveys of usual interest (agricultural surveys, industrial surveys, Cost of living surveys, employment and unemployment surveys, traffic surveys, etc.), a $y$ can not take any value in $\mathbb{R}^N$, but takes values in particular domain in $\mathbb{R}^N$, some with higher probability one may therefore, postulate some reasonable superpopulation model $\xi$ for $Y$ and exploit $\xi$ to produce suitable sampling strategies.

Brewer (1963), Royall (1970 a, 1976), Royall and Herson (1973) and their co-workers considered the survey population as a random sample from a superpopulation model and attempted to draw inference about the population parameters from a prediction theorist's point. Scott and Smith (1974) , Ho (1980) have studied Horvitz-Thompson type estimator under superpopulation model and compare with other design based estimators. Chaudhuri (1977) considered the ratio type estimator of mean and compare with other sampling strategies under superpopulation model. Mukhopadhyay (1978 & 82) studied variance estimator of finite population under superpopulation model approach and obtained some optimal sampling strategies.
1.6 Problems Discussed In Subsequent Chapters

This dissertation illustrates some estimators of population mean (or total) and variance of a finite population using auxiliary information in sample surveys. A brief review of related literature including preliminaries has been presented above.

Chapter-2, deals with an estimator of population total after making a transformation on study variable where the units are selected with PPS sampling. Optimal properties of proposed estimator have been discussed. It is found that the proposed estimator is unbiased and is more efficient than other usual estimators under certain conditions.

If two auxiliary variables are available and PPS sampling as well as method of estimation (ratio, regression or product) are adopted for estimating population mean/total, then for such situations a rule has been discussed, in chapter-3, for choosing an appropriate auxiliary variable for using it in selection of units and other in building up the estimator. The efficiency of all the possible estimators have been compared for various live data. A simulation study has also been carried out. Further, efficiency of these estimators has been studied under a super-population model.

Chapter-4 presents a generalized estimator of population mean under any sampling design if first two raw moments of auxiliary variable are available. Properties of proposed estimator have been discussed and an optimum class of estimators has been obtained. Again, properties of
proposed class of estimators have been discussed in SRS, PPS and Stratified sampling schemes. It is seen that the proposed estimator is considerably more efficient than generalized regression type estimator discussed by Sarndal et al (1992), Srivastava and Jhajj (1980), and others. The idea has been extended to non-response in chapters 5. Allocation problems for proposed estimator are studied in the chapter if cost considerations are made.

In chapter-6, an attempt has been made to improve estimator of population variance under general sampling scheme if past data is used as auxiliary variable. It is seen that the proposed estimator includes the usual difference type estimator proposed by Das and Tripathi (1978), searls and Intarapanich (1990), Srivastava and Jhajj (1980) and Singh et al (1988) as its particular case. Properties of proposed estimator are discussed under simple random sampling. The case of bivariate symmetrical populations has been also considered for further investigation of its properties. The constants involved in the estimator have been estimated by sample values. In this case the proposed estimator has been found more efficient than regression type estimator of population variance by Isaki (1983).

All the results are supported by numerical illustrations.