1. C.W. Snedecor, and W.G.Cochran - Statistical Methods, Oxford and IBH.
2. P.G. Hoel – Introduction to Mathematical Statistics, Wiley International, 2012.

| Year | Subject Title | Sem | Sub Code |
|---|---|---|---|
| **2018–19 Onwards** | **Core VI: BASIC SAMPLING THEORY** | **V** | **18BST52C** |

**Objective:** To create an overview about sampling and its various methods.

**UNIT I**

Concept of Population and sample – Census method and survey method – Merits and limitations of sample survey-Need for Sampling – Design, Organization and Execution of Sample Survey – Principal Steps in Sample Surveys – Preparation of Questionnaire and Schedules – Pilot Survey – Sampling and Non-Sampling Errors

**UNIT II**

Probability and non-probability sampling-Sampling from Finite Population – Simple Random Sampling With and Without Replacement – Unbiased Estimate of Mean and Variance – Finite Population Correction – Estimation of Standard Error– Simple Random Sampling of Attributes – Determination of sample size. Estimation of mean and variance

**UNIT III**

Stratified Random Sampling: Concept of Stratifying Factor advantages of strartification-Unbiased Estimate of the Mean and Variance– Proportional and Optimum Allocation – Neyman's Allocation - Comparison of Stratified and Simple Random Sampling.

**UNIT IV**

Systematic Sampling: Linear – circular systematic sampling Estimation of the Mean and Variance – Comparison of Simple, Stratified and Systematic Sampling – Population with Linear Trend - advantages and disadvantages

**UNIT V**

Cluster Sampling – estimation of mean and variance under equal and unequal cluster size. Two Stage Sampling with respect to Simple Random Sampling –Estimation of the Mean and Population Variance.

**Text Books:**

1. Daroga singh, F.S.Chauwdhary – Theory And Analysis Of Sample Survey Designs, New Age International (P) Ltd, publishers, New Delhi, 2015.
2. S.C. Gupta and V.K.Kapoor- Fundamentals of Applied Statistics, Sultan Chand & Sons, New Delhi, 4$^{th}$ Edition 2015.

**Reference Books:**

1. P.V. Sukathme and B.V. Sukathme - Sampling Theory Of Survey With Applications, Asia Publishing House.

## 7·1. INTRODUCTION

Before giving the notion of sampling we will first define *population.* In a statistical investigation the interest usually lies in the assessment of the general magnitude and the study of variation with respect to one or more characteristics relating to individuals belonging to a group. This group of individuals under study is called *population or universe.* Thus in statistics, population is an aggregate of objects, animate or inanimate under study. The population may be finite or infinite.

It is obvious that for any statistical investigation complete enumeration of the population is rather impracticable. For example, if we want to have an idea of the average per capita (monthly) income of the people in India, we will have to enumerate all the earning individuals in the country, which is rather a very difficult task.

If the population is infinite, complete enumeration is not possible. Also if the units are destroyed in the course of inspection (*e.g.,* inspection of crackers, explosive materials, etc.), 100% inspection, though possible, is not at all desirable. But even if the population is finite or the inspection is not destructive, 100% inspection is not taken recourse to because of multiplicity of causes, *viz.,* administrative and financial implications, time factor, etc., and we take the help of *sampling.*

A finite sub-set of statistical individuals in a population is called a *sample* and the number of individuals in a sample is called the *sample size.*

For the purpose of determining population characteristics, instead of enumerating entire population, the individuals in the sample only are observed. Then the sample characteristics are utilised to approximately determine or estimate the population. For example on examining the sample of a particular stuff we arrive at a decision of purchasing or rejecting that stuff. The error involved in such approximation is known as *sampling error* and is inherent and unavoidable in any and every sampling scheme. But sampling results in considerable gains, especially in time and cost not only in respect of making observation of characteristics but also in subsequent handling of the data.

Sampling is quite often used in our day-to-day practical life. For example, in a shop we assess the quality of sugar, wheat or any other commodity by taking a handful of it from the bag and then decide to purchase it or not. A housewife normally tests the cooked products to find if they are properly cooked and contain the proper quantity of salt.

## 7·2. PARAMETERS AND STATISTICS

In order to avoid verbal confusion with the statistical constants of the population, *viz.,* mean, variance, etc., of the population which are usually referred to as parameters, statistical measures computed from the sample observations alone, *e.g.,* mean, variance, etc., of the sample have been termed by Professor R.A. Fisher as *statistic.*

In practice parameter values are not known and their estimates based on the sample values are generally used. Thus statistic which may be regarded as an estimate of the parameter, obtained from the sample, is a function of the sample values only. It may be pointed out that a statistic, as it is based on sample values and as there are multiple choices of the samples that can be drawn from a population, varies from sample to sample. The determination or the characterisation of the variation (in the values of the statistic obtained from different samples) that may be attributed to chance or fluctuations of sampling is one of the fundamental problems of the sampling theory.



Scanned with CamScanner

*Remark (Unbiased Estimate).* A statistic $t = t(x_1, x_2, \ldots, x_n)$, a function of the sample values $x_1, x_2, \ldots, x_n$ is an unbiased estimate of population parameter $\theta$, if $E(t) = \theta$, *i.e.*, if $E(Statistics)$ = Parameter, then statistic is said to be an *unbiased estimate* of the parameter.

**7·2·1. Sampling Distribution.** The number of possible samples of size $n$ that can be drawn from a finite population of size $N$ is ${}^N C_n$. (If $N$ is large or infinite, then we can draw a large number of such samples.) For each of these samples we can compute a statistic, say '$t$' ..., *e.g.*, mean, variance, etc., which will obviously vary from sample to sample. The aggregate of the various values of the statistic under consideration so obtained (one from each sample), may be grouped into a frequency distribution which is known as the *sampling distribution* of the statistic. Thus, we can have the sampling distribution of the sample mean $\bar{x}$, the sample variance, etc.

**7·2·2. Standard Error.** The standard deviation of the sampling distribution of a statistic is known as its *Standard Error.* The standard errors (S.E.) of some of the well-known statistics are given in *Table 7·1*, where $n$ is the sample size, $\sigma^2$ the population variance, $P$ the population proportion and $Q = 1 - P$.

TABLE 7·1 : STANDARD ERRORS OF STATISTIC

| *S. No.* | *Statistic* | *Standard Error* |
|---|---|---|
| 1. | $\bar{x}$ | $\sigma/\sqrt{n}$ |
| 2. | Observed sample proportion '$p$' | $\sqrt{PQ/n}$ |
| 3. | Sample standard deviation $s$ | $\sqrt{\sigma^2/2n}$ |
| 4. | $s^2$ | $\sigma^2\sqrt{2/n}$ |
| 5. | Quartiles | $1{\cdot}36263\ \sigma/\sqrt{n}$ |
| 6. | Median | $1{\cdot}25331\ \sigma/\sqrt{n}$ |
| 7. | '$r$' = sample correlation coefficient | $(1-\rho^2)/\sqrt{n}$<br>$\rho$, population correlation coeff. |
| 8. | $\mu_3$ | $\sigma^3\sqrt{96/n}$ |
| 9. | $\mu_4$ | $\sigma^4\sqrt{96/n}$ |
| 10. | Coefficient of variation ($V$) | $\frac{V}{\sqrt{2n}}\sqrt{\left(1+\frac{2V^2}{104}\right)} \approx \frac{V}{\sqrt{2n}}$ |

***Utility of Standard Error.*** S.E. plays a very important role in the large sample theory and forms the basis of the testing of hypothesis. If $t$ is any statistic, then for large samples :

$$Z = \frac{t - E(t)}{\sqrt{V(t)}} \sim N(0, 1)$$

$$\Rightarrow \quad Z = \frac{t - E(t)}{\text{S.E.}(t)} \sim N(0, 1)$$

Thus, if the discrepancy between the observed and the expected (hypothetical) values of the statistic is greater then 1·96 times the S.E., the hypothesis is, rejected at 5% level of significance. Similary, if

$$|\, t - E(t)\,| \le 1{\cdot}96 \times \text{S.E.}(t),$$

the diviation is not regarded significant at 5% level of significance. In other words, the diviation, $t - E(t)$, could have arisen due to fluctuations of sampling and the data do not provide us any evidence against the null hypothesis which may, therefore, be accepted at 5% level of significance. Similary we can discuss the significance of the difference at 1% level of significance.

The magnitude of the standard error gives an index of the precision of the estimate of the parameter. The reciprocal of the standard error is taken as the measure of reliability or precision of the sample.



Scanned with CamScanner

## 7·3. THE PRINCIPAL STEPS IN A SAMPLE SURVEY

The main steps involved in the planning and execution of a sample survey may be grouped somewhat arbitrarily under the following heads :

**1. *Objectives of the Survey.*** The first step is to define in clear and concrete terms, the objectives of the survey. It is generally found that even the sponsoring agency is not quite clear in mind as to what it wants and how it is going to use the results. The sponsors of the survey should take care that these objectives are commensurate with the available resources in terms of money, manpower and the time limit required for the availability of the results of the survey.

**2. *Defining the Population to be Sampled.*** The population, *i.e.*, the aggregate of objects (animate or in-animate) from which sample is chosen should be defined in clear and unambiguous terms. For example, in sampling of farms clear-cut rules must be framed to define a farm regarding shape, size., etc., keeping in mind the border-line cases so as to enable the investigator to decide in the field without much hesitation whether or not to include a given farm in the population.

But practical difficulties in handling certain segments of the population may point to their elimination from the scope of the survey. Consequently, for reasons of practicability or convenience the population to be sampled (*the sampled population*) is different, in fact more restricted, than the population for which results are wanted (*the target population*).

**3. *The Frame and Sampling Units.*** The population must be capable of division into what are called *sampling units* for purposes of sample selection. The sampling units must cover the entire population and they must be distinct, unambiguous and non-overlapping in the sense that every element of the population belongs to one and only one sampling unit. For example, in socio-economic survey for selecting people in a town, the sampling unit might be an individual person, a family, a household or a block in a locality.

In order to cover the population decided upon, there should be some list, map or other acceptable material, called the *frame,* which serves as a guide to the population to be covered. The construction of the frame is often one of the major practical problems since it is the frame which determines the structure of the sample survey. The lists which have been routinely collected for some purpose, are usually found to be incomplete or partly illegible or often contain an unknown amount of duplication. Such lists should be carefully scrutinised and examined to ensure that they are free from these defects and are up-to-date. If they are not up-to-date, they should be brought up-to-date before using them. A good frame is hard to some by and only good experience helps to construct a good frame.

**4. *Data to be collected.*** The data should be collected keeping in view the objectives of the survey. The tendency should not be to collect too many data some of which are never subsequently examined and analysed. A practical method is to chalk out an outline of the tables that the survey should produce. This would help in eliminating the collection of irrelavent information and ensure that no essential data are omitted.

**5. *The Questionnaire or Schedule.*** Having decided about the type of the data to be collected, the next important part of the sample survey is the construction of the *questionnaire* (to be filled in by the respondent) or schedule of enquiry (to be completed by the interviewer) which requires skill, special technique as well as familiarity with the subject-matter under study. The questions should be clear, brief, corroborative, non-offending, courteous in tone, unambiguous and to the point so that not much scope of



Scanned with CamScanner

guessing is left on the part of the respondent or interviewer. Suitable and detailed instructions for filling up the questionnaire or schedule should also be prepared.

6. ***Method of Collecting Information.*** The two methods commonly employed for collecting data for human populations are :

(i) ***Interview Method.*** In this method, the investigator goes from house to house and interviews the individuals *personally*. He asks the questions one by one and fills up the schedule on the basis of the information supplied by the individuals.

(ii) *Mailed Questionnaire Method.* In this method, the questionnaire is mailed to the individuals who are required to fill it up and returns it duly completed.

Whether the data should be collected by interview method or mail questionnaire method or by physical observation has to be decided keeping in view the costs involved and the accuracy aimed at. Although mail surveys are less costly, there is scope for considerable non-response. Moreover mail method is practicable only among the educated people who are really interested in the particular survey being conducted. On the other hand, interview method costs more and there are interviewer errors also but without investigators the data collected may be worthless. In cases where data are to be collected by observations, the method of measurement, the type of measuring equipment or instrument, etc., are to be decided.

7. ***Non-respondents.*** Quite often (due to practical difficulties), the data cannot be collected for *all* the sampled units. For example, the selected respondent may not be available at his place when the investigator goes there or he may fail or even refuse to give certain information when contacted. This incompleteness, called *non-response,* obviously tends to change the results. Such cases of non-response should be handled with caution in order to draw unbiased and valid conclusions. Procedures will have to be devised to deal with those who do not furnish information. The reasons for non-response should be recorded by the investigator.

8. ***Selection of Proper Sampling Design.*** The size of the sample ($n$), the procedure of selection and the estimation of the population parameters along with their margins of uncertainty are some of the important statistical problems that should receive the most careful attention.

A number of designs (plans) for the selection of a sample are available and a judicious selection will guarantee good and reliable estimates. For each sampling plan, rough estimates of sample size $n$ can be obtained for a desired degree of precision. The relative costs and time involved should also be considered before making a final selection of the sampling plan.

9. ***Organisation of Field Work.*** It is absolutely essential that the personnel should be thoroughly trained in locating the sample units, recording the measurements, the methods of collection of required data before starting the field work. The success of a survey to a great extent depends upon the reliable field work. It is very necessary to make provisions for adequate supervisory staff for inspection after field work.

From practical point of view a *small pretest*, (*i.e.*, trying out the questionnaire and field methods on a small scale) has been found to be immensely useful. It always helps to decide upon effective method of asking questions and results in the improvement of the questionaire. Moreover, it might disclose certain problems and troubles that will otherwise be quite serious on a large-scale survey such as "the cost and the time may far exceed the available money and stipulated period."



Scanned with CamScanner

**11. *Summary and Analysis of the Data.*** The analysis of the data may be broadly classified into the following heads :

(*a*) *Scrutiny and editing of the data.* An initial quality check should be carried out by the supervisory staff while the investigators are in the field. Accordingly, the schedules should be thoroughly scrutinised to examine the plausibility and consistency of the data obtained. The scrutiny or editing of the completed questionnaires will help in amending recording errors or in eliminating data that are obviously erroneous and inconsistent

(*b*) *Tabulation of data.* Before carrying out the tabulation of the data, we must decide about the procedure for tabulation of the data which are incomplete due to non-response to certain items in the questionnaire and where certain questions are deleted in editing process. The method of tabulation, *viz.*, hand tabulation or machine tabulation, will depend upon the quantity of the data. For large-scale survey, machine tabulation will obviously be much quicker and economical. For a large-scale sample survey, the use of *code numbers* for qualitative variables is essential for machine tabulation. With simple questionnaires, the answers can sometimes be precoded, *i.e.*, entered in a manner in which they can be conveniently or routinely transferred to mechanical equipment such as personal computers, etc. Finally, the tables that lead to the estimates are prepared.

(*c*) *Statistical analysis.* After the data has been properly scrutinised, edited and tabulated, a very careful statistical analysis is to be made. Different methods of estimation may be available for the same data. Appropriate formulae should then be used to provide final estimates of the required information. Efforts should be made to keep the procedure free from errors.

(*d*) *Reporting and conclusions.* Finally, a report incorporating detailed statement of the different stages of the survey should be prepared. In the presentation of the results, it is good practice to report the technical aspect of the design, *viz.*, the types of the estimators used along with the amount of error to be expected in the most important estimate.

**12. *Information gained for Future Surveys.*** Any completed survey is helpful in providing a note of caution and taking lessons from it for designing future surveys. The information gained from any completed sample in the form of the data regarding the means, standard deviations and the nature of the variability of the principal measurements tougher with the cost involved in obtaining the data serves as a potential guide for improved together sampling. Moreover, in any complex survey, the things usually do not go exactly as planned. Any completed sample may serve as a lesson to the organisers for future surveys in recognising and rectifying the mistakes committed in the execution of the survey.

## 7·4. PRINCIPLES OF SAMPLE SURVEY

The theory of sampling is based on the following important principles :

**1.** *Principle of Statistical Regularity.* This principle has its origin in the mathematical theory of probability. According to **King,** "*The law of statistical regularity lays down that a moderately large number of items chosen at random from a large group are almost sure on the average to possess the characteristics of the large group.*" This principle stresses the desirability and importance of selecting the sample at random so that each and every unit in the population has an equal chance of being selected in the sample.

An immediate derivation from the principle of statistical regularity is the *Principle of Inertia of Large Numbers which states that, "other things being equal, as the sample size*



Scanned with CamScanner

*increases, the results tend to be more reliable and accurate."* This is because in dealing with large numbers the variations in the component parts tend to balance each other and consequently the variation in the aggregate result is likely to be insignificant. For example, in a coin-tossing experiment, the results will be approximately 50% heads and 50% tails provided the experiment is performed a fairly large number of times.

2. *Principle of Validity*. By the validity of a sample design we mean that it should enable us to obtain valid tests and estimates about the parameters of the population. The samples obtained by the technique of probability sampling satisfy this principle.

3. *Principle of Optimisation*. This principle impresses upon obtaining optimum results in terms of *efficiency and cost* of the design with the resources at our disposal. The reciprocal of sampling variance of an estimate provides a measure of its efficiency while a measure of the cost of the design is provided by the total expenses incurred in terms of money and man hour. The principle of optimisation consists in :

(*i*) achieving a given level of efficiency at minimum cost, and

(*ii*) obtaining maximum possible efficiency with given level of cost.

## 7·5. SAMPLING AND NON-SAMPLING ERRORS

The errors involved in the collection, processing and analysis of a data may be broadly classified under the following two heads :

(*i*) *Sampling Errors*, and (*ii*) *Non-sampling Errors.*

(*i*) ***Sampling Errors.*** Sampling errors have their origin in sampling and arise due to the fact that only a part of the population (*i.e.*, sample) has been used to estimate population parameters and draw inferences about the population. As such the sampling errors are absent in a complete enumeration survey.

Sampling biases are primarily due to the following reasons :

1. *Faulty selection of the sample.* Some of the bias is introduced by the use of defective sampling technique for the selection of a sample, *e.g.*, purposive or judgment sampling in which the investigator deliberately selects a representative sample to obtain certain results. This bias can select a representative sample to obtain certain results. This bias can be overcome by strictly adhering to a simple random sample or by selecting a sample at random subject to restrictions which while improving the accuracy are of such nature that they do not introduce bias in the results.

2. *Substitution.* If difficulties arise in enumerating a particular sampling unit included in the random sample, the investigators usually substitute a convenient member of the population. This obviously leads to some bias since the characteristics processed by the substituted unit will usually be different from those possessed by the unit originally included in the sample.

3. *Faulty demarcation of sampling units.* Bias due to defective demarcation of sampling units is particularly significant in area surveys such as agricultural experiments in the field or crop cutting survey, etc. In such surveys, while dealing with border line cases, it depends more or less on the discretion of the investigator whether to include them in the sample or not.

4. *Constant error due to improper choice of the statistics for estimating the population parameters.* For example, if $x_1, x_2, \ldots, x_n$ is a sample of independent observations, then the



Scanned with CamScanner

sample variance $s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/n$ as an estimate of the population variance $\sigma^2$ is biased whereas the statistic $\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$, is an unbiased estimate of $\sigma^2$

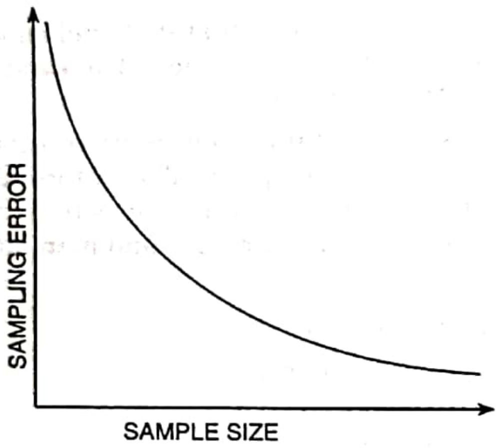

**Fig. 7·1**

***Remark.*** Increase in the sample size (*i.e.*, the number of units in the sample) usually results in the decrease in sampling error. In fact, in many situations this decrease in sampling error is inversely proportional to the square root of the sample size as illustrated in *Fig. 7·1.*

(*ii*) ***Non-sampling Errors.*** As distinct from sampling errors which are due to the inductive process of inferring about the population on the basis of a sample, the non-sampling errors primarily arise at the stages of observation, ascertainment and processing of the data and are thus present in both the complete enumeration survey and the sample survey. Thus, *the data obtained in a complete census, although free from sampling errors, would still be subject to non-sampling errors whereas data obtained in a sample survey should be subject to both sampling and non-sampling errors.*

Non-sampling errors can occur at every stage of the planning or execution of census or sample survey. The preparation of an exhaustive list of all the sources of non-sampling errors is a very difficult task. However, a careful examination of the major phases of a survey (complete or sample) indicates that some of the more important non-sampling errors arise from the following factors :

**1.** *Faulty Planning or Definitions.* The planning of a survey consists in explicitly stating the objectives of the survey. These objectives are then translated into (*i*) a set of definitions of the characteristics for which data are to be collected, and (*ii*) into a set of specifications for collecting, processing and publishing. Here the non-sampling errors can be due to :

(*a*) Data specification being inadequate and inconsistent with respect to the objectives of the survey.

(*b*) Error due to location of the units and actual measurement of the characteristics, errors in recording the measurements, errors due to ill-designed questionnaire, etc.

(*c*) Lack of trained and qualified investigators and lack of adequate supervisory staff.

**2.** *Response Errors.* These errors are introduced as a result of the responses furnished by the respondents and may be due to any of the following reasons :

(*i*) *Response errors may be accidental.* For example, the respondent may misunderstand a particular question and accordingly furnish improper information un-intentionally.

(*ii*) *Prestige bias.* An appeal to the pride or prestige of person interviewed may introduce yet another kind of bias, called *prestige bias* by virtue of which he may upgrade his education, intelligence quotient, occupation, income, etc., or downgrade his age, thus resulting in wrong answers.



Scanned with CamScanner

(*iii*) *Self-interest*. Quite often, in order to safeguard one's self-interest, one may give incorrect information, *e.g.*, a person may give an underestimate of his salary or production and an over-statement of his expenses or requirements, etc.

(*iv*) *Bias due to interviewer*. Sometimes the interviewer may affect the accuracy of the response by the way he asks questions or records them. The information obtained on suggestions from the interviewer is very likely to be influenced by interviewer's beliefs and prejudices.

(*v*) *Failure of respondent's memory*. One source of error which is common to most of the methods of collecting informations is that of '*recall*'. Many of the questions in surveys refer to happenings or conditions in the past and there is a problem both of remembering the event and associating it with the correct time period.

3. *Non-response Biases*. Non-response biases occur if full information is not obtained on all the sampling units. In house-to-house survey, non-response usually results if the respondent is not found at home even after repeated calls, or if he/she is unable to furnish the information on all the questions or if he/she refuses to answer certain questions. Therefore, some bias is introduced as a consequence of the exclusion of a section of the population with certain peculiar characteristics, due to non-response.

4. *Errors in Coverage*. If the objectives of the survey are not precisely stated in clear cut terms, this may result in (*i*) the inclusion in the survey of certain units which are not to be included, or (*ii*) the exclusion of certain units which were to be included in the survey under the objectives. For example, in a census to determine the number of individuals in the age group, say, 20 years to 50 years, more or less serious errors may occur in deciding whom to enumerate unless particular community or area is not specified and also the time at which the age is to be specified.

5. *Compiling Errors*. Various operations of data processing such as editing and coding of the responses, tabulation and summarising the original observations made in the survey are a potential source of error. Compilation errors are subject to control *though verification, consistency check*, etc.

6. *Publication Errors*. Publication errors, *i.e.*, the errors committed during presentation and printing of tabulated results are basically due to two sources. The first refers to the mechanics of publication—the proofing error and the like. The other, which is of more serious nature, lies in the failure of the survey organisation to point out the limitations of the statistics.

***Remarks*** **1.** In a sample survey, non-sampling errors may also arise due to defective frame and faulty selection of sampling units.

2. It is obvious that the non-sampling errors are likely to be more serious in a complete census as compared to a sample survey since in a sample survey the non-sampling errors can be reduced to a greater extent by employing qualified, trained and experienced personnel, better supervision and better equipment for processing and analysing relatively smaller data as compared to a complete census.

It has already been pointed out that usually sampling error decreases with increase in sample size. On the other hand, as the sample size increases, the non-sampling error tends to increase. Accordingly as sample size increases, the behaviour of non-sampling error is likely to be opposite to that of sampling error.

3. Quite often, the non-sampling error in a complete census is greater than both the sampling and non-sampling errors taken together in a sample survey. Obviously in such situations sample survey is to be preferred to complete enumeration survey.


Scanned with CamScanner

## 7·6. SAMPLING *vs* COMPLETE CENSUS

The *main merits* of sampling technique over the complete enumeration survey may be outlined as follows :

**1.** *Less time.* There is considerable saving in time and labour since only a part of the population has to be examined. The sampling results can be obtained more rapidly and the data can be analysed much faster since relatively fewer data have to be collected and processed.

**2.** *Reduced Cost of the Survey.* Sampling usually results in reduction in cost in terms of money and in terms of man hours. Although the amount of labour and the expenses involved in collecting information are generally greater per unit of sample than in complete enumeration, the total cost of the sample survey is expected to be much smaller than that of the complete census. Since in most of the cases our resources are limited in terms of money and the time within which the results of the survey should be obtained, it is usually imperative to resort to sampling rather than complete enumeration.

**3.** *Greater Accuracy of Results.* The results of a sample survey are usually much more reliable than those obtained from a complete census due to the following reasons :

(*i*) It is always possible to determine the extent of the sampling errors, and

(*ii*) The non-sampling errors due to factors such as training of field workers, measuring and recording observations, location of units, incompleteness of returns, biases due to interviewers, etc. are likely to be of a serious nature in complete census than in a sample survey. In a sample survey non-sampling errors can be controlled more effectively by employing more qualified and better trained personnel, better supervision and better equipment for processing and analysis of relatively limited data. Moreover, it is easier to guard against incomplete and inaccurate returns. There can be a follow-up in case of non-response or incomplete response. Effective control of non-sampling errors more than compensates the errors in the estimates due to sampling. As such more sophisticated statistical techniques can be employed to obtain relatively more reliable results.

**4.** *Greater Scope.* Sample survey has generally greater scope as compared with complete census. The complete enumeration is impracticable, rather inconceivable if the survey requires a highly trained personnel and more sophisticated equipment for the collection and analysis of the data. Since sample survey saves in time and money, it is possible to have a thorough and intensive enquiry because a more detailed information can be obtained from a small group of respondents.

**5.** If the population is too large, as for example, of trees in a jungle, we are left with no way but to resort to sampling.

**6.** If testing is destructive, *i.e.,* if the quality of an article can be determined only by destroying the article in the process of testing, as for example :

(*i*) testing the quality of milk or chemical salt by analysis,

(*ii*) testing the breaking strength of chalks,

(*iii*) testing of crackers and explosives,

(*iv*) testing the life of an electric tube or bulb, etc.,

complete enumeration is impracticable and sampling technique is the only method to be used in such cases.



Scanned with CamScanner

**7.** If the population is hypothetical, as for example in coin-tossing problem where the process may continue indefinitely (any number of times), sampling method is the only scientific method of estimating the parameters of the universe.

*Remarks* **1.** ***Prof. R. A. Fisher*** (1950) in a report of "The Sub-commission on Statistical Sampling of the United Nations" sums the advantages of sampling techniques over complete census in the following four words : *Adaptability, Speed, Economy* and *Scientific approach.*

**2.** From practical point of view, it has been seen that the method of random sampling with suitable adaptation of stratification of the universe, if it is heterogeneous or the technique of multistage random sampling if there are clearly demarcated stages, gives fairly good results, often better than those obtained by a complete census.

## 7·7. LIMITATIONS OF SAMPLING

The advantages of sampling over complete census as enumerated above can be derived only if

(*i*) the sampling units are drawn in a scientific manner,

(*ii*) appropriate sampling technique is used, and

(*iii*) the sample size is adequate.

Sampling theory has its own limitations and problems which may be briefly outlined are as follows :

**1.** Proper care should be taken in the planning and execution of the sample survey, otherwise the results obtained might be inaccurate and misleading.

**2.** Sampling theory requires the services of trained and qualified personnel and sophisticated equipment for its planning, execution and analysis. In the absence of these, the results of the sample survey are not trustworthy.

**3.** However, if the information is required about each and every unit of the universe, there is no way but to resort to complete enumeration. Moreover, if time and money are not important factors or if the universe is not too large, a complete census may be better than any sampling method.

## 7·8. TYPES OF SAMPLING

The technique or method of selecting a sample is of fundamental importance in the theory of sampling and usually depends upon the nature of the data and type of enquiry. The procedures of selecting a sample may be broadly classified under the following three heads :

(*i*) Subjective or judgement sampling,

(*ii*) Probability sampling, and

(*iii*) Mixed sampling

**7·8·1. Subjective (or Purposive or Judgment) Sampling.** In this sampling, the sample is selected with definite purpose in view and the choice of the sampling units depends entirely on the discretion and judgment of the investigator. This sampling suffers from drawbacks of favouritism and nepotism depending upon the beliefs and prejudices of the investigator and thus does not give a representative sample of the population. For example, if an investigator wants to give the picture that the standard of living has increased in the city of New Delhi, he may take individuals in the sample from the posh localities like Defence Colony, South Extension, Golf Link, Jor Bagh, Chanakyapuri, etc., and ignore the localities where low income group and middle class families live.


Scanned with CamScanner

This sampling method is seldom used and cannot be recommended for general use since it is often biased due to element of subjectiveness or the part of the investigator. However, if the investigator is experienced and skilled and this sampling is carefully applied, then judgment samples may yield valuable results.

**7·8·2. Probability Sampling.** Probability sampling is the scientific method of selecting samples according to some laws of chance in which *each unit in the population has some definite pre-assigned probability of being selected in the sample.* The different types of probability sampling are :

(*i*) Where each unit has an *equal chance* of being selected.

(*ii*) Sampling units have different probabilities of being selected.

(*iii*) Probability of selection of a unit is proportional to the sample size.

**7·8·3. Mixed Sampling.** If the samples are selected partly according to some laws of chance and partly according to a fixed sampling rule (no assignment of probabilities), they are termed as *mixed samples* and the technique of selecting such samples is known as *mixed sampling.*

The different types of sampling as given above have a number of variations, some of which may be listed below :

(*i*) *Simple Random Sampling*
(*ii*) *Stratified Random Sampling*
(*iii*) *Systematic Sampling*
(*iv*) *Multistage Sampling*
(*v*) *Quasi Random Sampling*
(*vi*) *Area Sampling*
(*vii*) *Simple Cluster Sampling*
(*viii*) *Multistage Cluster Sampling*
(*ix*) *Quota Sampling*

## 7·9. SIMPLE RANDOM SAMPLING (S.R.S.)

It is the technique of drawing a sample in such a way that *each unit of the population has an equal and independent chance of being included in the sample*.

In this method, an equal probability of selection is assigned to each unit of the population at the first draw. It also implies an equal probability of selecting any unit from the available units at subsequent draws.

Thus in S.R.S. from a population of $N$ units, the probability of drawing any unit at the first draw is $1/N$, the probability of drawing any unit in the second draw from among the available $(N-1)$ units, is $1/(N-1)$, and so on.

Let $E_r$ be the event that any specified unit is selected at the $r$th draw. Then

$P(E_r)$ = Prob. [that the specified unit is not selected in anyone of the previous $(r-1)$ draws and then selected at the $r$th draw]

$$\therefore \quad P(E_r) = \prod_{i=1}^{r-1} P\text{ (It is not selected at } i\text{th draw.)}$$

$\times P$[It is selected at $r$th draw given that it is not selected at the previous $(r-1)$ draws]

(By compound probability theorem, since draws are independent)

$$\therefore \quad P(E_r) = \prod_{i=1}^{r-1} \left[1-\frac{1}{N-(i-1)}\right] \times \frac{1}{N-(r-1)} = \prod_{i=1}^{r-1} \left(\frac{N-i}{N-i+1}\right) \times \frac{1}{N-r+1}$$



Scanned with CamScanner

**Source** :


1. S.C. Gupta and V.K. Kapoor: Fundamental of Applied Statistics –Sultan Chand & Sons, Fourth Edition, 2015.