Sampling & Sampling Distribution

INTRODUCTION

In all the spheres of life (such as Economic, Social and Business) the need for statistical investigation and data analysis is rising day by day. There are two methods of collection of statistical data : (i) **Census Method**, and (ii) **Sample Method.** Under census method, information relating to the entire field of investigation or units of population is collected; whereas under sample method, rather than collecting information about all the units of population, information relating to only selected units is collected. Before we make a detailed study of both the methods, we will explain some basic concepts related to them.

SOME BASIC CONCEPTS

(1) Universe or Population : In statistics, universe or population means an aggregate of items about which we obtain information. A universe or population means the entire field under investigation about which knowledge is sought. For example, if we want to collect information about the average monthly expenditure of all the 2,000 students of a college, then the entire aggregate of 2,000 students will be termed as Universe or Population. A population can be of two kinds (i) Finite and (ii) Infinite. In a finite population, number of items is definite such as, number of students or teachers in a college. On the other hand, an infinite population has infinite number of items *e.g.*, number of stars in the sky, number of water drops in an ocean, number of leaves on a tree or number of hairs on the head.

(2) Sample : A part of population is called sample. In other words, selected or sorted units from the population is known as a sample. In fact, a sample is that part of the population which we select for the purpose of investigation. For example, if an investigator selects 200 students from 2000 students of a college who represent all of them, then these 200 students will be termed as a sample. Thus, sample means some units selected out of a population which represent it.

CENSUS AND SAMPLE METHODS

There are two methods to collect statistical data :

(1) Census Method

(2) Sample Method

(1) Census Method

Census method is that method in which information or data is collected from each and every unit of the population relating to the problem under investigation and conclusions are drawn on their basis. This method is also called as **Complete Enumeration Method.** For example, suppose some information (like Monthly Expenditure, Average Height, Average Weight, etc.) is to be collected regarding 2000 students of a college. For that purpose if we collect data by inquiring each and every student of the college then this method will be called as Census method. In this example, the whole college *i.e.*, all 2000 students will be considered as a population and every student as an individual will be called the unit of the population. Population in India is conducted after every ten years by using census method.

Merits and Demerits of Census Method

Merits

(i) **Reliable and Accurate Data :** Data obtained by census method have more reliability and accuracy because in this method data are collected by contacting each and every unit of the universe.

(ii) **Extensive Information :** This method gives detailed information about each unit of the universe. For example, Indian population census does not only provide the knowledge about the number of persons but also information about their age, occupation, income, education, marital status, etc.

(iii) **Suitability :** This method is more suitable for the population with limited scope and diverse characteristics. Use of this method is also appropriate where intensive study is desired.

Demerits

(i) **More Expensive :** Census method is an expensive one. More money is needed for it as information is collected from each unit of the population. This is why this method is used by Government mostly for very important issues like Census, etc.

(ii) **More Time :** This method involves much time for data collection because data are collected from each and every unit of the population. This results in delay in making statistical inferences.

(iii) **More Labour :** This method of data collection also involves very much labour. For this the enumerators in a large number are required.

(iv) **Not Suitable for Specific Problems :** This method is not suitable relating to certain specific problems and infinite population. For example, if the population is infinite or items of the population are perishable or very complex type, then the census method is not suitable.

(2) Sampling Method

Sampling method is that method in which data is collected from the sample of items selected from population and conclusions are drawn from them. For example, if a study is to be made regarding the monthly expenditure of 2000 students of a college, then instead of collecting information from each student of the college, if we collect information by selecting some students like 100, then this will be called Sampling Method. On the basis of sampling method, it is possible to study the monthly expenditure of all the students of the college. Sampling method has three main stages (i) to select a sample (ii) to collect information from it and (ii) to make inferences regarding the population.

Importance of Sampling Method

In modern times sampling method is an important and popular method of statistical inquiry. Besides economic and business world, this method is widely used in daily life. For example, a housewife comes to know of the coating of the whole lot of rice by observing two-three grains only. A doctor tests the blood of a patient by examining one or two drops of blood only. In the same way, we learn about the quality of a commodity while buying the items of daily use like wheat, rice, pulses, etc. by observing the sample or specimen. In factories, statistical quality controller inspects the quality of items by examining a few units produced. A teacher gets the knowledge about the efficacy of his teaching by putting questions to a few students. In reality, there is scarcely any area left where sampling method is not used.

Merits and Demerits of Sampling Method

Merits

(i) **Saving of Time and Money :** Sampling method is less expensive. It saves money and labour because only a few units of the population are studied.

(ii) **Saving of Time :** In sampling method, data can be collected more quickly as these are obtained from some items of the universe. Thus much time is saved.

(iii) **Intensive Study :** As number of items is less in sampling method, they can be intensively studied.

(iv) **Organisational Convenience :** In this method, research work can be organised and executed more conveniently. More skilled and competent investigators can be appointed.

(v) **More Reliable Results :** If sample is selected in such a manner as it represents totally the universe, then the results derived from it will be more accurate and reliable.

(vi) **More Scientific :** Sampling method is more scientific because data can be inquired with other samples.

(vii) **Only Method :** In some fields where inquiry by census method is impossible, then in such situation, sampling method alone is more appropriate. If the population is infinite or too widespread or of perishable nature, then sampling method is used in such cases.

Demerits

(i) **Less Accurate :** Sampling method has less accuracy because rather than making inquiry about each unit of the universe, partial inquiry or inquiry relating to some selected units only is made.

(ii) **Wrong Conclusions :** If method of selecting a sample is not unbiased or proper caution has not been taken, then results are definitely misleading.

(iii) **Less Reliable :** Compared to census method, there is more likelihood of the bias of the investigator, which makes the results less reliable.

(iv) **Need of Specified Knowledge :** This is a complex method as specialised knowledge is required to select a sample.

(v) **Not Suitable :** If all units of a population are different from one another, then sampling method will not prove to be much useful.

Difference between Census and Sample Method

The main difference between the census method and the sampling method are as follows :

(i) **Scope :** In census method, all items relating to a universe are investigated whereas in sampling method only a few items are inquired.

(ii) **Cost :** Census method is expensive from the point of view of time, money and labour whereas Sampling method economises on them.

(iii) **Field of Investigation :** Census method isused in investigations with limited field whereas sampling method is used for investigations with large field.

(iv) **Homogeneity :** Census method is useful where units of the population are heterogeneous whereas sampling method proves more useful where population units are homogenous.

(v) **Type of Universe :** In such fields where study of each and every unit of the universe is necessary, census method is more appropriate. On the contrary, when population is infinite or vast or liable to be destroyed as a result of complete enumeration, then sampling method is considered to be more appropriate.

SAMPLING METHODS

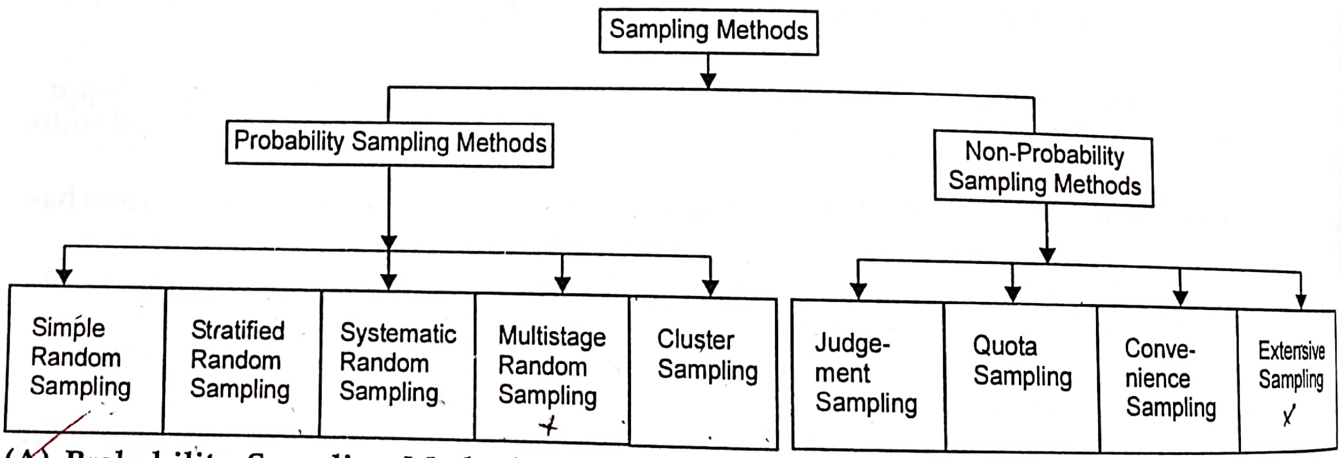
The method of selecting a sample out of a given population is called sampling. In other wodrs, sampling denotes the selection of a part of the aggregate statistical material with a view of obtaining information about the whole. Now a days, there are various methods of selecting sample from a population in accordance with various needs.

(A) Probability Sampling Methods :

(1) Simple Random Sampling
(2) Stratified Random Sampling
(3) Systematic Random Sampling
(4) Multistage Random Sampling
(5) Cluster Sampling

(B) Non-Probability Sampling Methods :

(6) Judgement Sampling
(7) Quota Sampling
(8) Convenience Sampling
(9) Extensive Sampling



(A) Probability Sampling Methods

Probability sampling methods are such methods of selecting a sample from the population in which all units of the universe are given equal chances of being included in the sample.

There are various variants of probability sampling methods, which are given below :

(1) Simple Random Sampling : Simple random sampling is that method in which each item of the universe has an equal chance of being selected in the sample. Which item will included in the sample and which not, such decision is not made by an investigator on his will but selection of the units is left on chance. According to random sampling, there are two methods of selecting a random sample:

(i) Lottery Method : In this method, each unit of the population is named or numbered which is marked on separate piece of paper. Such chits are folded and put into some urn or bag. Thereafter as many chits are made selected by some person as many units are to be included in a sample.

(ii) Tables of Random Numbers : Some experts has constructed random number tables. These tables help in selection of a sample. Of all such various tables, Tippett's Tables are most famous and are in use. Tippett has constructed a four-digit table of 10,400 numbers by using numbers as many as 41,600. In this method, first of all, all the items of a population are written serially. Thereafter by making use of Tippett's tables, in accordance with the size of the sample, numbers are selected. The selection of a sample with the help of Tippett's table can be made clear by an example :

An Extract of Tippett's Table

2952	6641	3992	9792	7979	5911
3170	5524	4167	9525	1545	1396
7203	4356	1300	2693	2370	7483
3408	2762	3563	6107	6913	7691
0560	5246	1112	9025	6008	8127

For example, suppose 12 units are to be chosen out of 5000 units. With Tippett's table, to decide such units, firstly 5000 units will be serially ordered from 1 to 5000 and then from Tippett's table, 12 numbers will be chosen from the beginning which are less than 5000. These 12 numbers are follows :

2952	4167	4356	2370
3992	1545	1300	3408
3170	1396	2693	2762

The items of such serial numbers will be included in the sample. If units of the population are less than 100, then 4 digit random numbers will be made compact into two digit numbers, and then such two digit numbers will be selected. Like as to select 6 units out of 60 units, the units with serial numbers 29, 39, 31, 41, 15 and 13 will be included in the sample.

Merits

(i) There is no possibility of personal prejudice in this method. In other words, this method is free from personal bias.

(ii) Under this method, every unit of the universe gets the equal chance of being selected.

(iii) The use of this method saves time, money and labour.

Demerits

(i) If sample size is small, then sample is not adequately represented.

(ii) If universe is very small, then this method is not suitable.

(iii) If some items of the universe are so important that their inclusion in the sample is very essential, then this method will not be appropriate.

(iv) This method will not be appropriate when population has units with diverse characteristics.

(2) Stratified Random Sampling : This method is used when units of the universe are heterogeneous rather than homogeneous. Under this method, first of all units of the population are divided into different *strata* in accordance with their characteristics. Thereafter by using random sampling, sample items are selected from each stratum. For example, if 150 students are to be selected out of 1500 students of a college, then firstly the college students will be divided into three groups on the basis of Arts, Commerce and Science. Suppose there are 500, 700, 300 students respectively in three faculties and 10% sample is to be taken, then on the basis of random sampling 50, 70 and 30 students respectively will be selected by using random sampling. Thus, this method assumes equal representation to each class or group and all the units of the universe get equal chance of being selected in the sample.

Merits

(i) There is more likelihood of representation of units in this method.

(ii) Comparative study on the basis of facts at different strata is possible under this method.

(iii) This method has more accuracy.

Demerits

(i) This method has limited scope because this method can be adopted only when the population and its different strata are known.

(ii) There can be the possibility of prejudice if the population is not properly stratified.

(iii) If the population is too small in size, it is difficult to stratify it.

(3) Systematic Random Sampling: In this method, all the items of the universe are systematically arranged and numbered and then sample units are selected at equal intervals. For example, if 5 out of 50 students are to be selected for a sample, then 50 students would be numbered and systematically arranged. One item of the first 10 would be selected at random. Subsequently, every 10th item from the selected number will be selected to frame a sample. If the first selected number is 5th item, then the subsequent numbers would be 15th, 25th, 35th and 45th.

Merits

(i) It is a simple method. Samples can be easily obtained by it.

(ii) This method involves very little time in sample selection and results are almost accurate.

Demerits

(i) In this method, each unit does not stand the equal chances of being selected because only the first unit is selected on random sampling basis.

(ii) If all the units are different in characteristics, then results will not be appropriate.

(4) Multistage Random Sampling : When sampling procedure passes through many stages, then it is known as multi-stage sampling. In this method, firstly the entire universe or population is divided into stages or substages. From the each stage some units are selected on random sampling basis. Thereafter these units are subdivided and on the basis of random sampling again some sub-units are selected. Thus, this goes on with sub-division further and selection on. For example, for the purpose of a study regarding Adult Education in Haryana State, first some districts will be selected on random basis. Thereafter out of the selected districts, some tehsils and out of tehsils, some villages or towns may be thus selected, further out of the villages or towns, some neighbourhood, or wards and out of the wards, some households will be selected from whom the inquiry will be made concerning the problem at hand.

Merits

(i) This is the best method of studying a universe or population on regional basis.

(ii) This method is suitable for those problems where decisions on the basis of sample alone can not be taken.

Demerits

(i) This method requires many tests to correctly estimate the level of accuracy which involves a lot of time and labour.

(ii) In this method, level of estimated accuracy level is predecided which does not seem logical.

(5) Cluster Sampling : In this method, simply the universe is divided into many groups called cluster and out of which a few clusters are selected on random basis and then the clusters are complete enumerated. This method is usually applied in industries like as in pharmaceutical industry, a machine produces medicines tablets in the batches of hundred each, then for quality inspection, a few randomly selected batches are examined.

(B) Non-Probability Sampling Methods

Non-probability sampling methods are those methods in which selection of units is made on the basis of convenience or judgement of the investigator rather than on the basis of probability or chance. In such methods, selection of units is made in accordance with the specific objectives and convenience of the investigator.

(6) Judgement Sampling : Under this method, the selection of the sample items depends exclusively on the judgement of the investigator. In other words, the investigator exercises his judgement in the choice and includes those items in the sample which he thinks are most typical of the universe with regard to the characteristics under study. For example, if a sample of 20 students is to be selected from a class of 80 students for analysing the spending habits of the 10 students, the investigator would select 20 students, who in his opinion are representative of the class.

Merits

(i) This method is less expensive.
(ii) This method is very simple and easy.
(iii) This method is useful in those fields where almost similar units exist or some units are too important to be left out of the sample.

Demerits

(i) There is greater chance of the investigator's own prejudice in this method.
(ii) This method is not very accurate and reliable.

(7) Quota Sampling : In this method, the investigators are assigned definite quotas according to some criteria. They are instructed to obtain the required number to fill in each quota. The investigators select the individuals (*i.e.*, sample items) to collect information on their personal judgements within the quotas. When all or a part of the whole quota is not available or approachable, the quota is completed by supplementing new responds. Quota sampling is a type of judgement sampling.

Merits

(i) In this method, there is greater chance of important units being included.
(ii) Statistical inquiry is more organised in this method on account of the units of the quotas being fixed.

Demerits

(i) Possibility of prejudice shall remain.
(ii) There is greater likelihood of sampling error in this method.

(8) Convenience Sampling : In this type of non-probability sampling, the choice of the sample is left completely to the convenience of the investigator. The investigator obtain a sample according to his convenience. For example, a book publisher selects some teachers conveniently on the basis of the list of the teachers from the college prospectus and gets feedback from them regarding his publication. This method is less expensive and more simple but is unscientific and unreliable. This method results in more dependence on the enumerators. This method is appropriate for sample selection where the universe or population is not clearly defined or list of the units is not available or sample units are not clear in themselves.

(9) Extensive Sampling : In this method, sample size is taken almost as big as the population itself like 90% the section of the population. Only those units are left out for which data collection is very difficult or almost impossible. Due to very large sample size, the method has greater level of accuracy. Intensive study of the problem becomes possible but this method involves heavy resources at disposal.

SAMPLING AND NON-SAMPLING ERRORS

The choice of a sample though may be made with utmost care, involves certain errors which may be classified into two types : (i) **Sampling Errors,** and (2) **Non-Sampling Errors.** These errors may occur in the collection, processing and analysis of data.

(1) Sampling Errors

Sampling errors are those which arise due to the method of sampling. Sampling errors arise primarly due to the following reasons:

(1) Faulty selection of the sampling method.
(2) Substituting one sample for the sample due to the difficulties in collecting the sample.
(3) Faulty demarcation of sampling units. (limits)
(4) Variability of the population which has different characteristics.

(2) Non-Sampling Errors

Non-sampling errors are those which creep in due to human factors which always varies from one investigator to another. These errors arise due to any of the following factors :

(1) Faulty planning.
(2) Faulty selection of the sample units.
(3) Lack of trained and experienced staff which collect the data.
(4) Negligence and non-response on the part of the respondent.
(5) Errors in compilation.
(6) Errors due to wrong statistical measures.
(7) Framing of a wrong questionnaire.
(8) Incomplete investigation of the sample survey.

Imp.

Basic Concepts of Sampling

Sampling Distribution: The purpose of selecting and studying a sample from the population is to estimate or make inference about some population characteristics. In this process, the knowledge of the sampling distribution is of vital importance.

Some Important Terms:

The following terms are widely used in the study of the sampling distribution:

(1) Parameters: Any statistical measures computed from the population data is known as parameter. Thus, population mean, population standard deviation, population variance, population proportion, etc., are all parameters. Parameters are denoted by the Greek letters such as μ, σ^2, σ and P.

(2) Statistics: Any statistical measure computed from sample data is known as statistic. Thus, sample mean, sample standard deviation, sample variance, sample proportion, etc., are all statistics. Statistics are denoted by Roman letters such as $\overline{X}$, s, s^2 and p.

(3) Sampling with and without replacement: Sampling is a procedure of selecting a sample from the population. Sampling may be done with or without replacement. Sampling where each unit of a population may be chosen more than once is called sampling with replacement. If each unit cannot be chosen more than once, it is called sampling without replacement. In case of sampling with replacement, the total number of possible samples each of size n drawn from a population of size N is N^n. But if the sampling is without replacement, the total number of possible samples will be $^{N}C_{n} = m$ (say).

Sampling Distribution of a Statistic

Sampling distribution constitutes the theoretical basis of statistical inference and is of considerable importance in business decision making. Sampling distribution of a statistic is the frequency distribution which is formed with various values of a statistic computed from different samples of the same size drawn from the same population. Suppose we draw all possible samples of size n from the population (N) with or without replacement. For each possible sample drawn from the population, we compute a statistic such as mean, median, standard deviation, variance, etc. The set of all possible values of a statistic is then classified and grouped into a frequency distribution (or probability distribution). The distribution so obtained is called the sampling distribution of a statistic. We could have various sampling distribution depending upon the nature of the statistic we have computed. If, for instance, the particular statistic computed is the sample mean, the distribution is called sampling distribution of mean. If, we compute variance of each sample, then it is called the sampling distribution of variance. Similarly, we could have sampling distributions of proportion, median, standard deviation, etc.

An Important Property of Sampling Distribution

An important property of the sampling distribution of a statistic is that if random samples of large size ($n > 30$) are taken from a population which may be normally distributed or not, then the sampling distribution of the statistic will approach a normal distribution.

Standard Error of a Statistic

The standard deviation of the sampling distribution of a statistic is known as the standard error of a statistic. As there are various types of sampling distributions, we could have various types of standard errors depending on the nature of sampling distribution. The standard deviation of the sampling distribution of means is called the standard error of the means. In sampling theory, instead of using the term standard deviation for measuring variation, we use a new term called standard error of mean. The standard error of mean measure the extent to which the sample mean differ from the population mean. Thus, **the basic difference between the standard deviation and standard error of mean is that the former measures the extent to which the individual items differ from the central value and the latter measures the extent to which individual sample mean differ from the population mean.** Like the standard error of the means, we could have standard error of the median, standard deviation, proportion, variance, etc.

Utility of Standard Error : The standard error is used in a large number of problems which are discussed as follows :

(1) Reliability of a Sample : The standard error gives an idea about the reliability and precision of a sample. That is, it indicates how much the estimated value differs from the observed values. The greater the standard error, the greater is the deviation between the estimated and observed values and lesser is the reliability of a sample. The smaller the standard error, the smaller is the deviation between the estimated and observed values and greater is the reliability of a sample.

(2) Tests of Significance : The standard error is also used to test the significance of the various results obtained from small and large samples. In case of large sample, if the difference between the observed and the expected value is greater than 1.96 standard error, then we reject the hypothesis at 5% and conclude that sample differs widely from the population. But if the difference between the observed and the expected value is greater than 2.58 S.E. (Standard error), then we reject the null hypothesis at 1% and conclude that the sample differs widely from the population.

(3) To determine the confidence limits of the unknown population mean : The standard error enables us in determining the confidence limits within which a population parameter is expected to lie with a certain degree of confidence. The confidence limits of the unknown population mean μ are given by.

Large Sample	Small Sample
95% confidence limits for μ	95% confidence limit for μ
$\bar{x} - 1.96$ S.E. and $\bar{x} + 1.96$ S.E.	$\bar{x} \pm t_{.05}$ S.E.
99% confidence limits for μ	99% confidence limits for μ
$\bar{x} - 2.58$ S.E. and $\bar{x} + 2.58$ S.E.	$\bar{x} \pm t_{01}$ S.E.

SAMPLING DISTRIBUTION OF MEANS

It is an important sampling distribution widely used in the sampling theory. Draw all possible samples of size n with or without replacement from population of size N with mean μ and variance σ^2. For each possible sample drawn from the population, we compute the mean $\bar{x}$ of each sample. The mean will vary from sample to sample. The set of all possible means obtained from different samples is called the sampling distribution of means.

Properties : The following are the important properties of the sampling distribution of means:

(i) The mean of the sampling distribution of means is equal to the population mean (μ). Symbolically,

$$\mu_{\bar{x}} = \mu \qquad \text{or} \qquad E(\bar{x}) = \mu$$

This property can be proved as follows :

Let $x_1, x_2, \ldots, x_n$ represent a random sample (with replacement) of size n from a finite population of size N having its mean μ and variance σ^2, then

$$\bar{x} = \frac{x_1 + x_2 + \ldots x_n}{n}$$

$$E(\bar{x}) = E\left[\frac{\Sigma x}{n}\right] = E\left[\frac{x_1 + x_2 + \ldots + x_n}{n}\right]$$

$$= \frac{1}{n}\{E(x_1) + E(x_2) + \ldots + E(x_n)\}$$

$$= \frac{1}{n}\{\mu + \mu + \mu + \ldots + \mu\} = \frac{1}{n} \cdot n\,\mu = \mu$$

Thus, the mean of the sampling distribution of means is equal to the population mean.

(2) The standard error of the sampling distribution of means is obtained as :

$$\text{S.E.}_{\bar{x}} \text{ or } \sigma_{\bar{x}} = \frac{\text{S.D. of Population}}{\sqrt{\text{Size of the sample}}} = \frac{\sigma}{\sqrt{n}}$$

This property can be proved as follows :

$$\text{Var}(\bar{x}) = \text{Var}\left(\frac{\Sigma x}{n}\right) = \text{Var}\left(\frac{x_1 + x_2 + \ldots x_n}{n}\right)$$

$$= \frac{1}{n^2}[\text{Var}(x_1) + \text{Var}(x_2) + \ldots \text{Var}(x_n)]$$

$$= \frac{1}{n^2}[\sigma^2 + \sigma^2 + \ldots \sigma^2]$$

$$= \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

where, σ^2 is the population variance, x is the sample.

Because $n > 1$, obviously, $\frac{\sigma^2}{n} < \sigma^2 \Rightarrow V(\bar{x}) <$ Population variance.

$$\therefore \quad S.E._{\bar{x}} = \sigma_{\bar{x}} = \sqrt{Var(\bar{x})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

This formula holds only when sampling is with replacement.

Note : When the population is finite and the samples are drawn without replacement, then $S.E._{\bar{x}}$ is obtained as :

$$S.E._{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

(3) The sampling distribution of means is approximately a normal distribution with mean μ and variance σ^2, provided the sample is large ($n > 30$).

(4) The following formula is used to find the probability of the sampling distribution of means.

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

Let us illustrate the concept of sampling distribution of means by the following example :

Example 1. **Consider a population consisting of three values : 2, 5 and 8. Draw all possible samples of size 2 with replacement from the population. Construct sampling distribution of means. Also find the mean and standard error of the distribution.**

Solution. The population consists of three values.The total number of possible samples of size 2 drawn with replacement are $N^n = 3^2 = 9$. All possible random samples and their sample mean are shown is the following table :

Sample No.	Sample Values	Sample Mean $\bar{x}$
1.	(2, 2)	$\frac{1}{2}(2+2) = 2$
2.	(5, 2)	$\frac{1}{2}(5+2) = 3.5$
3.	(8, 2)	$\frac{1}{2}(8+2) = 5$
4.	(2, 5)	$\frac{1}{2}(2+5) = 3.5$
5.	(5, 5)	$\frac{1}{2}(5+5) = 5.0$
6.	(8, 5)	$\frac{1}{2}(8+5) = 6.5$

7.	(2, 8)	$\frac{1}{2}(2+8)=5.0$
8.	(5, 8)	$\frac{1}{2}(5+8)=6.5$
9.	(8, 8)	$\frac{1}{2}(8+8)=8.0$

On the basis of the means ($\bar{x}$) of all the 6 possible samples, the sampling distribution of means is given below :

Sample Means ($\bar{x}$)	f	$f\bar{x}$	$d=\bar{x}-\mu_{\bar{x}}$	d^2	fd^2
2	1	2	− 3	9	9
3.5	2	7	− 1.5	2.25	4.50
5.0	3	15	0	0	0
6.5	2	13	1.5	2.25	4.50
8.0	1	8	+ 3	9.0	9.0
	$\Sigma f=9$	$\Sigma f\bar{x}=45$			$\Sigma fd^2=27$

Mean of the Sampling Distribution of Means

$$\mu_{\bar{x}}=\frac{\Sigma f\bar{x}}{\Sigma f}=\frac{45}{9}=5$$

Variance of the Sampling Distribution of Means

$$\text{Var}(\bar{x})=\frac{\Sigma f(\bar{x}-\mu_{\bar{x}})^2}{\Sigma f}=\frac{\Sigma fd^2}{\Sigma f}=\frac{27}{9}=3$$

Hence, $\text{S.E.}_{\bar{x}}=\sigma_{\bar{x}}=\sqrt{3}=1.732$

Aliter : The sampling distribution of means can also be written in terms of probability as :

Sample Means ($\bar{x}$)	2	3.5	5.0	6.5	8.0
Probability (p)	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

Since 3.5 occurs twice, its probability of occurence is $\frac{2}{9}$, 5 occurs thrice, its probability of occurrence is $\frac{3}{9}$ and 6.5 occurrence is $\frac{2}{9}$. Each of the other sample mean occurs only once with probability $\frac{1}{9}$.

Mean of the Sampling Distribution of Means

$$E(\bar{x})=\Sigma p\bar{x}=2\times\frac{1}{9}+3.5\times\frac{2}{9}+5\times\frac{3}{9}+6.5\times\frac{2}{9}\times 8.0\times\frac{1}{9}$$

$$=\frac{1}{9}\cdot[2+7+15+13+8]=\frac{45}{9}=5$$

Variance of the Sampling Distribution of Means

$$\text{Var}(\bar{x})=E(\bar{x}^2)-[E(\bar{x})]^2$$

$$E(\bar{x}^2) = 2^2 \times \frac{1}{9} + 3.5^2 \times \frac{2}{9} + 5^2 \times \frac{3}{9} + 6.5^2 \times \frac{2}{9} + 8^2 \times \frac{1}{9}$$
$$= \frac{1}{9} \cdot [4 + 25 + 75 + 84.5 + 64]$$
$$= \frac{1}{9} [252.5] = 28.055$$
$$\text{Var}(\bar{x}) = E(\bar{x}^2) - [E(\bar{x})^2] = 28.055 - 25 = 3.055 = 3$$

Hence, $$\text{S.E.}_{\bar{x}} = \sqrt{\text{Var}(\bar{x})} = \sqrt{3} = 1.732$$

Example 2. **Construct a sampling distribution of the sample means from the following population :**

Population Unit :	1	2	3	4
Observation :	22	24	26	28

when random sample of size 2 are taken from it without replacement. Also find the mean and standard error of the distribution.

Solution. The population consists of four values (22, 24, 26, 28). The total number of possible sample of size 2 drawn without replacement are ${}^4C_2 = 6$. All the possible random samples and their sample means are shown in the table given below :

Sample No.	Sample Values	Sample Mean $\bar{x}$
1.	(22, 24)	$\frac{1}{2}(22+24) = 23$
2.	(22, 26)	$\frac{1}{2}(22+26) = 24$
3.	(22, 28)	$\frac{1}{2}(22+28) = 25$
4.	(24, 26)	$\frac{1}{2}(24+26) = 25$
5.	(24, 28)	$\frac{1}{2}(24+28) = 26$
6.	(26, 28)	$\frac{1}{2}(26+28) = 27$

On the basis of the means ($\bar{x}$) of all the 6 samples without replacement, the sampling distribution of mean is given below:

Sampling Distribution of Means without Replacement

Sample Means ($\bar{x}$)	f	$f\bar{x}$	$d = \bar{x} - \mu_{\bar{x}}$	d^2	fd^2
23	1	23	−2	4	4
24	1	24	−1	1	1
25	2	50	0	0	0
26	1	26	1	1	1
27	1	27	2	4	4
	$\Sigma f = 6$	$\Sigma f\bar{x} = 150$			$\Sigma fd^2 = 10$

Mean of the Sampling Distribution of Means

$$\mu_{\bar{x}} = \frac{\Sigma f\bar{x}}{\Sigma f} = \frac{150}{6} = 25$$

Variance of the Sampling Distribution of Means

$$\text{Var}(\bar{x}) = \frac{\Sigma f(\bar{x}-\mu_x)^2}{\Sigma f} = \frac{\Sigma fd^2}{\Sigma f} = \frac{10}{6} = \frac{5}{3}$$

Hence, $$S.E._{\bar{x}} = \sigma_x = \sqrt{\text{Var}\,\bar{x}} = \sqrt{\frac{5}{3}} = 1.29$$

Aliter : The sampling distribution of means can also be written in terms of probability as below :

Sample Means ($\bar{x}$)	23	24	25	26	27
Probability (p)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Since, 25 occurs twice, its probability of occurence is $\frac{2}{6}$. Each of the other sample means occurs only once with probability $\frac{1}{6}$.

Mean of the Sampling Distribution of Means

$$E(\bar{x}) = \Sigma p\bar{x} = \frac{1}{6}\times 23 + \frac{1}{6}\times 24 + \frac{2}{6}\times 25 + \frac{1}{6}\times 26 + \frac{1}{6}\times 27$$

$$= \frac{1}{6}\cdot[23+24+50+26+27] = \frac{150}{6} = 25$$

Variance of the Sampling Distribution of Means

$$\text{Var}(\bar{x}) = E(\bar{x}^2) - [E(\bar{x})]^2$$

$$E(\bar{x}^2) = 23^2\times\frac{1}{6} + 24^2\times\frac{1}{6} + 25^2\times\frac{2}{6} + 26^2\times\frac{1}{6} + 27^2\times\frac{1}{6}$$

$$= \frac{1}{6}\cdot[529+576+1250+676+729]$$

$$= \frac{3760}{6} = 626.17$$

$$\text{Var}(\bar{x}) = E(\bar{x}^2) = [E(\bar{x})]^2 = 626.17 - 625 = 1.67$$

Hence, $$S.E._{\bar{x}} = \sigma_{\bar{x}}\ \sqrt{\text{Var}(\bar{x})} = \sqrt{1.67} = 1.29$$

Example 3. **A population consists of four elements : 3, 7, 11, 15. Consider all possible samples of size two which can be drawn with replacement from this population. Find (i) the population mean μ (ii) the population variance σ^2 (iii) the mean of the sampling distribution of means (iv) standard error (or S.D.) of the sampling distribution of means. Verify (iii) and (iv) by using (i) and (ii) and by use of suitable formula.**

Solution. (i) μ = population mean $= \frac{\Sigma X}{N} = \frac{3+7+11+15}{4} = \frac{36}{4} = 9$

(ii) σ^2 = population variance $= \frac{\Sigma (X-\mu)^2}{N} = \frac{(-6)^2+(-2)^2+(2)^2+(6)^2}{4} = \frac{80}{4} = 20$

$\therefore \quad \sigma = \text{S.D.} = \sqrt{20}$

(iii) All possible random samples of size two with replacement is $N^n = 4^2 = 16$ and their sample means are shown in the following table :

Sample No.	Sample Values	Sample Mean $\bar{x}$	Sample No.	Sample Values	Sample Mean $\bar{x}$
1.	(3, 3)	3	9	(11, 3)	7
2.	(3, 7)	5	10	(11, 7)	9
3.	(3, 11)	7	11	(11, 11)	11
4.	(3, 15)	9	12	(11, 15)	13
5.	(7, 3)	5	13	(15, 3)	9
6.	(7, 7)	7	14	(15, 7)	11
7.	(7, 11)	9	15	(15, 11)	13
8.	(7, 15)	11	16	(15, 15)	15

On the basis of the mean ($\bar{x}$) of all the 16 samples with replacement, the sampling distribution of $\bar{x}$ can be written as :

Sample Means ($\bar{x}$)	Frequency (f)	$f\bar{x}$	$d = \bar{x} - \mu_{\bar{x}}$	d^2	fd^2
3	1	3	− 6	36	36
5	2	10	− 4	16	32
7	3	21	− 2	4	12
9	4	36	0	0	0
11	3	33	+ 2	4	12
13	2	26	+ 4	16	32
15	1	15	+ 6	36	36
	$\Sigma f = 16$	$\Sigma f\bar{x} = 144$			$\Sigma fd^2 = 160$

Mean of the Sampling Distribution of Means

$$\mu_{\bar{x}} = \frac{\Sigma f\bar{x}}{\Sigma f} = \frac{144}{16} = 9$$

Variance of the Sampling Distribution of Means

$$\text{Var}(\bar{x}) = \frac{\Sigma f(\bar{x} - \mu_{\bar{x}})^2}{\Sigma f} = \frac{160}{16} = 10$$

Hence $\qquad \text{S.E.}_{\bar{x}} = \sigma_{\bar{x}} = \sqrt{\text{Var}(\bar{x})} = \sqrt{10}$

Using the formula, $\mu_{\bar{x}} = \mu$ and $V(\bar{x}) = \frac{\sigma^2}{n}$, we get the mean of the sampling distribution of means $= \mu_{\bar{x}} = \mu = 9$ and variance of the sampling distribution of means $= \frac{\sigma^2}{n} = \frac{20}{2} = \frac{\sigma^2}{2} = 10.$

Hence, the results of (iii) and (iv) are verified by using the results of (i) and (ii).

Example 4. A population consists of the following elements :

2, 4, 5, 8, 11

Find:

(a) How many different samples of size 3 are possible when sampling is done without replacement.

(b) List all of the possible different samples.

(c) Compute the mean of each of the samples given in part (b).

(d) Find the sampling distribution of sample mean $\overline{X}$.

(e) If all the elements are euqally likely, compute the value of the population mean μ.

Solution. The population consists of five elements (2, 4, 5, 8, 11).

(a) The total number of possible samples of size 3 drawn without replacement are $^5c_3 = 10$.

(b) All the possible different samples and their sample means are shown in the following table.

Sample No.	Sample Values	Sample Mean $\bar{x}$
1.	(2, 4, 5)	$\frac{1}{3}(2+4+5) = 3.67$
2.	(2, 4, 8)	$\frac{1}{3}(2+4+8) = 4.67$
3.	(2, 4, 11)	$\frac{1}{3}(2+4+11) = 5.67$
4.	(2, 5, 8)	$\frac{1}{3}(2+5+8) = 5.0$
5.	(2, 5, 11)	$\frac{1}{3}(2+5+11) = 6.0$
6.	(2, 8, 11)	$\frac{1}{3}(2+8+11) = 7.0$
7.	(4, 5, 8)	$\frac{1}{3}(4+5+8) = 5.67$
8.	(4, 5, 11)	$\frac{1}{3}(4+5+11) = 6.67$
9.	(4, 8, 11)	$\frac{1}{3}(4+8+11) = 7.67$
10.	(5, 8, 11)	$\frac{1}{3}(5+8+11) = 8.0$

In the above table, we have 10 possible samples of size 3 without replacement. Since, 5.67 occurs twice, its probability of occurrence is $\frac{2}{10}$. Each of the other sample means occur only once with probability $\frac{1}{10}$.

Sampling distribution of means $\bar{x}$ (*i.e.*, the probability distribution of sample mean $\bar{x}$) is given below :

Sampling Distribution of $\bar{x}$

Sample Mean $\bar{x}$	3.67	4.67	5	5.67	6	6.67	6	7.67	8.0
Probability (p)	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

(e) Population consists of the values (2, 4, 5, 8, 11). Since, each value occurs equally likely, the probability of occurence of each value is $\frac{1}{5}$. Hence,

Population Values (X) :	2	4	5	8	11
Probability (p) :	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

$$\text{Population Mean} = \mu = 2 \times \frac{1}{5} + 4 \times \frac{1}{5} + 5 \times \frac{1}{5} + 8 \times \frac{1}{5} + 11 \times \frac{1}{5}$$

$$= \frac{1}{5} \cdot [2 + 4 + 5 + 8 + 11] = \frac{30}{5} = 6$$

LAW OF LARGE NUMBERS AND CENTRAL LIMIT THEOREM

Law of Large Numbers and the Central Limit Theorem both serve the basis for the development of sampling distribution of a statistic.

Law of Large Numbers : The law of large numbers states that as the sample size increases, the sample mean would be closer and closer to the population mean. It does not gurantee that if the sample size is increased sufficiently, the sample mean will be equal to the population mean. There are two implications of the law of large numbers (i) the difference between sample mean and population mean can be reduced by increasing the sample size, and (ii) variation from one sample mean to another sample mean (of the same size) also decreases as the size of the sample increases.

Central Limit Theorem : It is widely used in the field of estimation and inference. This theorem states that if we select random sample of large size n from any population with mean μ and standard deviation σ and compute the mean of each sample, then the sampling distribution of mean $\bar{x}$ approaches normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. This is true even if the population itself is not normal. The utility of this theorem is that it requires virtualy no conditions on the distribution pattern of the population.

QUESTIONS

1. Distinguish between population and sample. Discuss the relative merits and demerits of census and sampling Methods.
2. Explain the various methods of sampling. Also discuss their relative merits and demerits.
3. Explain the simple random sampling technique and state when it is used.
4. (a) What is a random sample ? Discuss the various methods of drawing a random sample.
   (b) Distinguish between sampling and non-sampling errors.

5. What is meant by sampling distribution of a statistic? Also, define standard error of a statistic.

OR

Discuss briefly the concept of sampling distribution of an estimator.

6. Distinguish between :

   (i) Population and Sample (ii) Parameters and Statistics

   (iii) Sampling with and without replacement.

7. Write a note on "Sampling Distribution of Means."

8. A population consists of five numbers (2, 3, 6, 8, 11). Draw all possible random samples of size 2 which can be drawn with replacement from this population. Construct a sampling distribution of means. Also, find the mean and standard error of the distribution.

9. A population consists of four numbers (3, 7, 11, 15). Consider all possible samples of size 2 which can be drawn without replacement from this population. Construct the sampling distribution of means. Also find the mean and standard error of the distribution.

10. A population consists of the following five elements :

$$3, 5, 9, 11, 17$$

Find :

(a) How many different samples of size 3 are possible when sampling is done without replacement ?

(b) List all of the possible different samples.

(c) Compute the sample mean for each of the samples given in part (b).

(d) Find the sampling distribution of the sample mean $\overline{x}$. Use a probability histogram to graph the sampling distribution of $\overline{x}$.

(e) If all five population values are equally likely, compute the value of the population mean, μ.

11. A population consists of numbers : 2, 3, 6, 8, 11. (i) Enumerate all possible samples of size 2 which can be drawn from this population (without replacement) (ii) Calculate the mean of the sampling distribution of means and show that the mean of the sampling distribution of means is equal to the population mean. (iii) Calculate the variance of the sampling distribution of means and show that it is less than that population variance.

12. (a) Show that the mean of the sampling distribution of means is equal to the population mean *i.e.*, $E(\overline{x}) = \mu$.

    (b) Derive the variance of the sampling distribution of the sample mean. Is it more than population variance ?

13. Give statements of Law of Large Numbers and Central Limit Theorem. Discuss their significance in sampling theory.

❀❀❀