Q.1 Define Population

Ans In 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. For example, if we want to have an idea of average per capita (monthly) income of the people in India.

Q.2 Define Sample and Sample size.

Ans A finite subset of statistical individuals in a population is called a **sample** and the number of individuals in a sample is called the **sample size**. 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.

Q.3 Define Parameter and Statistics.

Ans Statistical measures computed from the population observations alone is called parameter, e.g., mean, variance etc. of the population.

Statistical measures computed from the sample observation alone is called statistics, e.g., mean variance etc. of the sample.

Q.4 Explain the meaning of Sampling Error.

Ans 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 utilized 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 the stuff. Thee 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.

The sampling errors cannot be completely eliminated but may be minimized by choosing a proper sample of adequate size and adopting suitable method of estimation.

Other words, "*Sampling error occurs when the sample is not representative of the population. When random sampling techniques are used to select elements for the sample, sampling error occurs by chance. Many times the statistic computed on the sample is not an accurate estimate of the population parameter because the sample was not representative of the population. This result is caused by sampling error. With random samples, sampling error can be computed and analyzed.*"

Q.5 Explain the meaning of Non-sampling Error.

Ans Error other than sampling error in a survey are called non-sampling errors. The error usually arises due to faulty planning, defective schedules or questionnaires, incompleteness and inaccuracy of returns, non-response, compiling errors etc. These errors can be minimized by employing efficient investigators and supervisory staff, full coverage, better management etc. Non-sampling errors are likely to be more widespread in complete enumeration than in a sample survey.

Other words, "*All errors other than sampling errors are non-sampling errors; the many possible non-sampling errors include missing data, recording errors, input processing errors, and analysis errors. Other non-sampling errors result from the measurement instrument, such as errors of unclear definitions, defective questionnaires, and poorly conceived concepts. Improper definition of the frame is a non-sampling error. In many cases, finding a frame that perfectly fits the population is impossible. Insofar as it does not fit, a non-sampling error has been committed.*"

Response errors are also non-sampling errors. They occur when people do not know, will not say, or overstate. Virtually no statistical method is available to measure or control for non-sampling errors. The researcher must eliminate these errors through carefully planning and executing the research study.

Q.6 In what situations sampling is inevitable?

Ans Sampling is inevitable in the following situations:

- ➢ When population is infinite.
- ➢ When the item or unit is destroyed under investigation.
- ➢ When the results are required in a short time.
- ➢ When resources for survey are limited particularly in respect of money and trained persons.
- ➢ When area of survey is wide.

Q.7 What is meant by sampling method?

Ans By sampling method we mean the manner or scheme through which the required numbers of units are selected in a sample from a population.

Q.8 Mention in brief the objective of sampling (need of sampling)

Ans The foremost purpose of sampling is to gather maximum information about the population under consideration at minimum cost, time and human power. This is best achieved when the sample contains all the properties of the population.

Q.9 Define sampling unit.

Ans The constituents of a population which are the individuals to be sampled from the population and cannot be further subdivided for the purpose of sampling at a time are called **sampling units**. For instance, to know the average income per family, the head of the family is a sampling unit. To know the average yield of wheat, each farm owner's field of wheat is a sampling unit.

Q.10 What is meant by sampling frame?

Ans For adopting any sampling procedure it is essential to have a list or a map identifying each sampling unit by a number. Such a list or map is called sampling frame.
A list of voters, a list of house holds, a list of technical persons, areas in a map marked by number for soil surveys, a list of villages in a district, a list of farmer's fields, etc., are a few examples of sampling frame.

Q.11 Define Sample size

Ans The determination of sample size for estimating a mean or proportion is a crucial question. By selecting a sample size lower than the correct size may affect reliability and a higher one will mean more cost and time. The determination of the size of a sample is the most important factor for the purposes of estimation of the value of the population parameters. We have the following formula for it.

Q.12 Calculate Sample size for Estimating a Mean

Ans In order to determine the sample size for estimating a population mean, the following factors must be known:

(i) The descried confidence level.

(ii) The permissible sampling error $E = \overline{X} - \mu$

(iii) The standard deviation σ .

After having known the above mentioned three factors, the size of sample mean n is given by

$$n = \left(\frac{\sigma Z}{E}\right)^2$$

Example:1 It is know that the population standard deviation in waiting time for L.P.G. gas cylinder in Bhuj is 15 days. How large a sample should be chosen to be 95% confident, the waiting time is within 7 days of true average.
Solution: The required sample size is

$$n = \left(\frac{\sigma Z}{E}\right)^2 = \left(\frac{15 \times 1.96}{7}\right)^2 = 17.64 \approx 18$$

Q.13 Calculate Sample size for Estimating a Proportion

Ans In this case we must know the following three factors
(i) The desired confidence level

(ii) The permissible sampling error E =difference between the estimate from the sample

p and P to be estimated = $p - P$

(iii) The estimated true proportion of success.

The sample size n is given by

$$n = \frac{Z^2 pq}{E^2}, where \quad q = 1 - p$$

Example:2 Mr. X wants to determine on the basis of sample study the mean time required to complete a certain job so that he may be 95% confident that the mean may remain within ± 2 days of the true mean. As per the available records the population variance is 64 days. How large should the sample be for his study?

Solution: Here $\sigma = \sqrt{64} = 8$. Also Z is $N(0,1)$.

It is given that $P\left[|\bar{x} - \mu| < 2\right] = 0.95$ (I)

But $$P\left[(\bar{x} - \mu) < 1.96 \frac{\sigma}{\sqrt{n}}\right] = 0.95 \quad \text{(II)}$$

From (I) and (II) we get

$$\frac{1.96 \times \sigma}{\sqrt{n}} = 2 \Rightarrow n = \left(\frac{1.96 \times 8}{2}\right)^2 = (7.84)^2 = 61.46 \approx 62$$

Q.14 Which factors are responsible for the size of a sample?

Ans The size of a sample depends upon the following factors:

- ➢ The purpose for which the sample is drawn.
- ➢ The heterogeneity of the sampling units in the population. More is the heterogeneity; larger is the size of the sample.
- ➢ Resources available for the study in terms of time and money.
- ➢ Number of technical person and/or equipment available.
- ➢ Precision of estimates required is an important factor in determining the size of sample. For greater precision usually a large sample is preferred.

Q.15 Explain Sampling Distribution

Ans The number of possible samples of size that can be drawn from a finite population of sizes is (If 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, 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 , the sample variance etc.

Q.16 Explain Sampling Distribution of Mean

Ans In the inferential statistics process, a researcher selects a random sample from the population, computes a statistic on the sample, and reaches conclusions about the population parameter from the statistic. In attempting to analyze the sample statistic, it is essential to know the distribution of the statistic. So far we studied several distributions, including the binomial distribution, the Poisson distribution, the hypergeometric distribution, the uniform distribution, the normal distribution, and the exponential distribution.

In this section we explore the sample mean $\bar{x}$ as the statistic. The sample means is one of the more common statistics used in the inferential process. To compute and assign the probability of occurrence of a particular value of a sample mean, the researcher must know the distribution of the sample means. One way to examine the distribution possibilities is to take a population with a particular distribution, randomly select samples of a given size, compute the sample means, and attempt to determine how the means are distributed.

Suppose a small finite population consists of only N = 8 numbers:

54 55 59 63 64 68 69 70

Suppose we take all possible samples of size n = 2 from this population with replacement. The result is the following pairs of data.

(54,54)	(55,54)	(59,54)	(63,54)	(64,54)	(68,54)	(69,54)	(70,54)
(54,55)	(55,55)	(59,55)	(63,55)	(64,55)	(68,55)	(69,55)	(70,55)
(54,59)	(55,59)	(59,59)	(63,59)	(64,59)	(68,59)	(69,69)	(70,59)
(54,63)	(55,63)	(59,63)	(63,63)	(64,63)	(68,63)	(69,63)	(70,63)
(54,64)	(55,64)	(59,64)	(63,64)	(64,64)	(68,64)	(69,64)	(70,64)
(54,68)	(55,68)	(59,68)	(63,68)	(64,68)	(68,68)	(69,68)	(70,68)
(54,69)	(55,69)	(59,69)	(63,69)	(64,69)	(68,69)	(69,69)	(70,69)
(54,70)	(55,70)	(59,70)	(63,70)	(64,70)	(68,70)	(69,70)	(70,70)

The means of each of these samples follow.

54	54.5	56.5	58.5	59	61	61.5	62
54.5	55	57	59	59.5	61.5	62	62.5
56.5	57	59	62	61.5	63.5	64	64.5
58.5	59	61	63	63.5	65.5	66	66.5
59	59.5	61.5	63.5	64	66	66.5	67
60	62.5	63.5	65.5	66	68	68.5	69
61.5	62	64	66	66.5	68.5	69	69.5
62	62.5	64.5	66.5	67	69	69.5	70

However, the sample means for samples taken from these populations appear to be approximately normally distributed, especially as the sample sizes become larger.

Q.17 Explain CENTRAL LIMIT THEOREM

Ans If samples of size n are drawn randomly from a population that has a mean of μ and a standard deviation of σ , the sample means $\bar{x}$ are approximately normally distributed for sufficiently large sample sizes $(n \geq 30)$ regardless of the shape of the population distribution. If the population is normally distributed, the sample means are normally distributed for any size sample.

From mathematical expectation, it can be shown that the mean of the sample is the population mean.

$$\mu_{\bar{x}} = \mu$$

and the standard deviation of the sample means (called the standard error of the mean) the standard deviation of the population divided by the square root of the sample size.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

The central limit theorem creates the potential for applying the normal distribution to many problems when sample size is sufficiently large. Sample means that have been computed for random samples drawn from normally distributed populations are normally distributed. However, the real advantage of the central limit theorem comes when sample data drawn from populations not normally distributed or from populations of unknown shape also can be analyzed by using the normal distribution because the sample means are normally distributed for sufficiently large sample sizes. [Note: The actual form of the central limit theorem is a limit function of calculus. As the sample size increases to infinity, the distribution of sample means literally becomes normal in shape.]

The central limit theorem states that sample means are normally distributed regardless of the shape of the population for large samples and for any sample size with normally distributed populations. Thus, sample means can be analyzed by using z scores.

The formula to determine z scores for individual values from a normal distribution:

$$z = \frac{\bar{x} - \mu}{\sigma}$$

If sample means are normally distributed, the z score formula applied to sample means would be

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

This result follows the general pattern of z scores: the difference between the statistic and its mean divided by the statistic's standard deviation. In this formula, the mean of the statistic of interest is $\mu_{\bar{x}}$, are the standard deviation of the statistic of interest is $\sigma_{\bar{x}}$, sometimes referred to as the standard error of the mean. To determine $\mu_{\bar{x}}$, the researcher would randomly draw out all possible samples of the given size from the population, compute the sample means, and average them. This task is virtually impossible to accomplish in any realistic period of time. Fortunately $\mu_{\bar{x}}$ equals the population mean, μ, which is easier to access. Likewise, to determine directly the value of $\sigma_{\bar{x}}$, the researcher would take all possible samples of a given size from a population, compute the sample means, and determine the standard deviation of sample means. This task also is practically impossible. Fortunately, $\sigma_{\bar{x}}$ can be computed by using the population standard deviation divided by the square root of the sample size.

As sample size increases, the standard deviation of the sample means becomes smaller and smaller because the population standard deviation is being divided by larger and larger values of the square root of n. The ultimate benefit of the central limit theorem is a practical, useful version of the z formula for sample means:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Q.18 Explain different sampling methods in brief.

Ans 1. Random Sampling or Probability Sampling Method.
2. Non-Random Sampling or Non-Probability Sampling Method.

RANDOM SAMPLING OR PROBABILITY SAMPLING METHOD

B. SIMPLE RANDOM SAMPLING: It is the method of selection of a sample in such a way that each and every number of population or universe has an equal chance or probability of being included in the sample. Random sampling can be carried our in two ways: (i) Lottery Method (ii) Random Number Method.

C. RESTRICTED RANDOM SAMPLING: It is three types (i) Stratified Sampling (ii) Systematic Sampling (iii) Multi-Stage Sampling.

(i) STRATIFIED SAMPLING: In stratified random sampling, the population is divided into strata (groups) before the sample is drawn. Strata are so designed that they do not overlap. An elementary unit from each stratum is drawn at random and the units so drawn constitute a sample. Stratified sampling is suitable in those cases where the population is heterogeneous but there is homogeneity within each of the groups or strata.

(ii) SYSTEMATIC SAMPLING: In this method every elementary unit of the population is arranged in order and the sample units are distributed at equal and regular intervals. In other words, a sample of suitable size is obtained (from the orderly arranged population) by taking every unit say tenth unit of the population, one of the first units in this ordered arrangement is chosen at random and the

sample is computed by selecting every tenth unit (say) from the rest of the lot. If the first unit selected is 4, then the other units constituting the sample will be 14, 24, 34, 44 and so on.

(iii)MULTISTAGE SAMPLING: In this sampling method, sample of elementary units is selected in stages. Firstly a sample of cluster is selected and from among them a sample for elementary units is selected. It is suitable in those case where population size is very big and it contains a large number of units.

NON-RANDOM SAMPLING OR NON-PROBABILITY SAMPLING METHOD

(i)PURPOSIVE SAMPLING: Purposive sampling is the method of sampling by which a sample is drawn from a population based entirely on the personal judgment of the investigator. It is also known as JUDGEMENT SAMPLING or DELIBERATE SAMPLING. Randomness finds no place in it and so the sample drawn under this method cannot be subjected to mathematical concepts used in computing sampling error.

(ii)CLUSTER SAMPLING: Cluster Sampling involves arranging elementary items in a population into heterogeneous subgroups that are representative of the overall population. One such group constitutes a sample for study.

(iii)QUOTA SAMPLING: In quota sampling method, quotas are fixed according to the basic parameters of the population determined earlier and each field investigator is assigned with quotas of number of elementary units to be interviewed.

(iv)CONVENIENCE SAMPLING: In convenience sampling, a sample is obtained by selecting convenient population elements from the population.

(v)SEQUENTIAL SAMPLING: In sequential sampling a number of sample lots are drawn one after another from the population depending on the results of the earlier samples drawn from the same population. Sequential sampling is very useful in Statistical Quality Control. If the first sample is acceptable, then no further sample is drawn. On the other hand if the lot is completely unacceptable, it is rejected straightway. But if the initial lot is of doubtful and marginal character falling in the bond ling between the acceptance and rejection limits, a second sample is drawn and if need be a third sample of bigger size may be drawn in order to arrive at a decision on the final acceptance or rejection of the lot. Such sampling can be based on any of the random or non random method of selection.

Cluster(or Area)Sampling

Cluster (or area) sampling is a fourth type of random sampling. Cluster (or area) sampling Involves dividing the population into non overlapping areas or clusters. However, in contrast to stratified random sampling where strata are homogeneous, cluster sampling identifies clusters that tend to be internally heterogeneous. In theory, each cluster contains a wide variety of elements, and the cluster is a miniature, or microcosm, of the population. Examples of clusters are towns, companies, homes, colleges, areas of a city and geographic regions. Often clusters are naturally occurring groups of the population and are already identified, such as states or Standard Metropolitan Statistical Areas. Although area sampling usually refers to clusters that are areas of the population, such as geographic regions and cities, the terms cluster sampling and area sampling are used interchangeably in this text.

Q.19 Write the characteristic of a good sample.

Ans
- ➢ The sample must be representative of the population, i.e. all the characteristics of the population should be in the sample.
- ➢ Each unit of the population must be independent. The selection of any unit of the population should not depend upon the selection of any other unit.
- ➢ All units of the sample should be selected during the same period of time.
- ➢ No unit of the population should be favoured, at the time of selecting a sample.
- ➢ All the units of the sample should be selected under identical conditions.
- ➢ The number of units of the sample should be adequate. If less number of units are selected, better accuracy cannot be achieved. If more units are selected, the expenditure, time etc. will be more. Hence taking into consideration, the size of the population, the amount to be spent in sampling, and the accuracy to be maintained, the size of the sample should be decided.

The reliable information about a population can be obtained if the sample is truly representative. In most of the statistical investigation sampling method is used instead of the entire population survey.

Q.20 Write the advantages of sampling.

Ans ➢ As few units are to be examined, detailed information can be obtained. Thus, the standard of accuracy increases.

➢ As small numbers of units are to be examined, the survey work can be completed in less time. If the result is required within a given period of time, sampling is useful.

➢ The method is less expensive compared to the population survey.

➢ As few persons are to be employed for survey work experts can be selected and the results become more reliable.

➢ In destructive type of testing i.e. during inspection, the unit is required to be destroyed; the sampling method is only the way out.

➢ Sometimes the results obtained from a population survey are required to be tested. Sampling can be used in such circumstances.

➢ Time and energy can be saved by sample survey.

➢ When the population is large, the method of sampling is useful to draw inferences about it.

Thus, there are many advantages of sampling. Statisticians therefore regard this method very useful. **The main aim of a sample survey is to obtain reliable information about the population in less time and at a lower cost.**

Q.21 Write the difference between population study and sample study.

Ans

	Population study	Sample study
1.	In population study all units are examined and hence it requires more time.	In sample study few units are examined and hence it requires less time.
2.	The cost of survey is more.	The cost of survey is less.
3.	As more observations are to be studied, the accuracy cannot be maintained.	As less number of observations are to be studied, the desired accuracy can be maintained.
4.	When the study is of destructive nature, population study cannot be employed	This method can be employed even if the study is of destructive nature.
5.	As more persons are to be employed in the survey work experts may not be available.	As few persons are to be employed the experts can be appointed.
6.	As all observations are to be studied, detail study cannot be done.	As few observations are to be studied detail study can be done.
7.	When the field of inquiry is very large this method is labourious and difficult.	The sample survey is less labourious and relatively easy.
8.	When the complete information is needed this method is to be used.	When the complete information of all units is required, this method cannot be used.

Q.22 Write the limitation of the Sampling.

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

➢ 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.

➢ 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.

Q.23 Explain the meaning of simple random sampling with replacement and without replacement.

Ans If the units are selected or drawn one by one in such a way that a unit drawn at a time is replaced back to the population before the subsequent draw, it is known as simple random sampling with replacement (SRSWR). In this type of sampling from a population of size , the probability of selection of a unit at each draw remains . In SRSWR, a unit can be included more than once in a sample. Therefore, if the required sample size is , the

effective sample size is sometimes less than due to the inclusion of one or more units more than once.

With the idea that effective sample size should be adhered to, the simple random sampling without replacement (SRSWOR) is adopted. In this method a unit selected once is not included in the population at any subsequent draw. Hence, the probability of drawing a unit from a population of units at draw is .

In simple random sampling, the probability of selection of any sample of size from a population consisting of units remains the same, is the number of all possible samples.

Q.24 Define Standard Error.

Ans Standard error is the standard deviation of the sampling distribution of an estimator.

From a population of units, samples of size can possibly be drawn from the population. If the sampling units are distinct, each sample will give more or less a different estimate of a parameter. In this way, the estimates themselves will follow a distribution. The standard deviation of the estimates obtained from different possible samples is called standard error (SE). May it be standard error of the sample mean , the standard deviation , etc. The standard errors of some of the well known statistics are given below, where is the sample size, the population variance, the population proportion and .

S. No.	Statistics	Standard Errror
1		
2	Observed sample proportion	
3	Sample standard deviation	
4		
5		being the population correlation coefficient

Q.25 Write merits of Simple Random Sampling

Ans
- ➢ As all the members of the population have equal chance of being selected in the sample, the selection is without any partiality or prejudice.
- ➢ Random sample is a representative sample. As the size of the sample increase, the reliability of the result also increases.
- ➢ The error committed about the result of the population from the results of the sample can be estimated.
- ➢ The reliability of the results can also be checked by this method.

Q.26 Write limitations of Simple Random Sampling

Ans
- ➢ In this method a list of all the units of the population is required. In absence of such a list, the method cannot be used.
- ➢ Simple random sampling can give reliable results only when the population is homogeneous.
- ➢ The work of preparing slip or giving numbers is tedious and tiresome particularly when the population is large.
- ➢ At times, the results obtained by the method are biased.
- ➢ When the population is very small, this method fails to give reliable results.

Q.27 Write merits of Stratified Random Sampling

Ans
- ➢ As population is divided into different strata, all the parts of the population are adequately represented.
- ➢ Administrative convenience increases in this type of sampling. The work of drawing samples in different strata can be assigned to different persons.
- ➢ Each stratum being internally homogeneous even a small sample from that stratum will give reliable information about the stratum.
- ➢ Sampling problems differ in different parts of the population. By appointing suitable persons, accurate information can be obtained from all the parts of the population.
- ➢ If different standards of accuracy are required for different strata, this method is convenient.

Q.28 Write limitations of Simple Random Sampling

Ans
- ➢ It is always difficult to divide the population into homogeneous strata.
- ➢ The results obtained by this method cannot be reliable if stratification is not proper.
- ➢ The mathematics involved in estimating the population characteristics from the sample results is difficult in this method.

➢ Simple random samples are to be drawn from different strata. In absence of required number of efficient persons, the desired standard of accuracy cannot be attained.

Q.29 Write merits of Systematic Sampling

Ans ➢ Systematic sampling is operationally more convenient than simple random sampling or stratified random sampling. Time and work involved is also relatively much less.

➢ Systematic sampling may be more efficient the simple random sampling provided the frame (the list from which sample units are drawn) is arranged wholly at random. The most common approach to randomness is provided by alphabetical lists such as names in telephone directory although even these may have certain non-random characteristics.

Q.30 Write limitations and disadvantages of Systematic Sampling

Ans ➢ Systematic samples are not random samples.

➢ If is not a multiple of , then

(i) the actual sample size is different from that required, and

(ii) sample mean is not an unbiased estimate of the population mean.

However, these disadvantages can be overcome by adopting a technique known as circular systematic sampling.

➢ In systematic sampling, we have

However, it is not possible to obtain an unbiased estimate of this variance. This obviously is a great drawback, since one important requirement for adopting any sampling method is that it should provide an estimate of the sampling error.

➢ Systematic sampling may yield highly biased estimates if there are periodic features associated with the sampling interval, i.e., if the frame(list) has a periodic feature and is equal to or a multiple of the period.

Q.31 Write comparisons Systematic Sampling with stratified sampling.

Ans ➢ Systematic sampling resembles stratified sampling in the sense that groups of units look like strata but no criterion has been considered in the formation of groups that ensure homogeneity.

➢ No independent samples are drawn from each group.

➢ In systematic sampling we have only one sample from the whole population.

The above three points clearly reveal that systematic sampling is quite different from stratified sampling.