Very short questions

1. Describe the type I and II error. (Describe error in testing of hypothesis)

Type I error:

The rejection of null hypothesis, when it is true, is called type I error. The probability of a type I error is denoted by α. This α is also known as size of the critical region.

α=P (type I error)

=P (reject H_0 when H_0 is true)

Type II error

If false null hypothesis H_0 is accepted, it is said to be type II error. The probability of this type of error is denoted by β.

β=P (type II error)

= P (accept H_0 when H_0 is false)

2. Explain qualities of a good estimator.

A good estimator is one which is as close to the true value of the parameter as possible. A good estimator has following four properties.

- **Unbiasedness:** If the expected value of sample statistic is equal to parametric value, then the estimator is said to be unbiased, otherwise biased.
  i.e., E(t) =θ
- **Consistency:** if the sample size increases, the value of sample statistic (t_n) becomes very nearer to the value of population parameter θ then it is called consistent estimator.
  i.e., $t_n \rightarrow \theta$ as $n \rightarrow$□
- **Sufficiency:** an estimator containing all the information contained in the sample regarding the parameter θ is a sufficient estimator.
- **Efficiency:** let t_1 and t_2 be two consistent estimators of parameter θ. Then the estimator t_1 is said to be more efficient estimator than t_2 if variance of t_1 is less than variance of t_2.
  i.e., var (t_1) < var (t_2).

3. What are the advantages of stem and leaf display?

- Concise representation of data
- Shows range, minimum and maximum, gaps and cluster, and outlier easily.
- Can handle extremely large data sets.

4. Write down the area properties of normal distribution.

The area properties of normal distribution are as follows:

- The area under the normal probability curve between the ordinates at X= μ ± σ is 0.6826.
  i.e., P (μ-σ < X < μ + σ) = 0.6826. This is same as, the interval μ ± σ covers 68.26% of observations.
- The area under the normal probability curve between the ordinates at X= μ ± 2σ is 0.9544.

i.e., P ($\mu-2\sigma < X < \mu + 2\sigma$) = 0.9544. This is same as, the interval $\mu \pm 2\sigma$ covers 95.44% of observations.

- The area under the normal probability curve between the ordinates at $X= \mu \pm 3\ \sigma$ is 0.9974. i.e., P ($\mu-3\sigma < X < \mu + 3\sigma$) = 0.9974. This is same as, the interval $\mu \pm 3\sigma$ covers 99.74% of observations.

5. State the necessary conditions to apply binomial distribution.

Binomial distribution is widely used probability distribution of discrete random variable. The binomial distribution holds under the following conditions:

- A Binomial experiment consists of n trials repeated under the same conditions.
- Each trail has only two outcomes – Success and Failure.
- The repeated trials are independent.
- The probability of success remains constant for each trial.

6. State the necessary conditions that Poisson distribution is obtained from binomial distribution.

Poisson distribution is a limiting case of binomial distribution under the following conditions:

- The number of trials n is indefinitely large i.e., $n \rightarrow \square$.
- The probability of success for each trial is very small i.e., $p \rightarrow 0$.
- The mean np is finite constant i.e., $np=\lambda$ (finite)
- Events are independent.

Under these four conditions, the binomial probability function tends to probability function of the Poisson distribution given by

$P(X=x) = p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$, x= 0, 1, 2, …

7. Define descriptive and inferential statistics.

Descriptive Statistics

In the descriptive Statistics, the Data is described in a summarized way. The summarization is done from the sample of the population using different parameters like Mean or standard deviation. Descriptive Statistics are a way of using charts, graphs, and summary measures to organize, represent, and explain a set of Data.

- Data is typically arranged and displayed in tables or graphs summarizing details such as histograms, pie charts, bars, or scatter plots.

Inferential Statistics

In the Inferential Statistics, we try to interpret the Meaning of descriptive Statistics. After the Data has been collected, analyzed, and summarized we use Inferential Statistics to describe the Meaning of the collected Data.

- Inferential Statistics are intended to test hypotheses and investigate relationships between variables and can be used to make population predictions.
- Inferential Statistics are used to draw conclusions and inferences, i.e., to make valid generalizations from samples.

8.Describe the difference between descriptive and inferential statistics.

Descriptive Statistics	Inferential Statistics
Concerned with the describing the target population.	Make inferences from the sample and generalize them to the population.
Organize, analyze, and present the data in a meaningful manner.	Compares, test and predicts future outcomes.
Final results are shown in form of charts, tables and graphs.	Final result is the probability scores.
Describes the data which is already known.	Tries to make conclusion about the population that is beyond the data available.
Tools- measures of central tendency and measure of dispersion.	Tools- hypothesis tests, analysis of variance etc.

9. Write down the methods of collecting primary data.

The various methods with which the primary data can be collected are as follows:

- **Direct personal interview:** In this method, the investigator personally interviews the respondent either directly or through phone or through any electronic media. This method is suitable when the scope of investigation is small and greater accuracy is needed.
- **Indirect oral interview:** The indirect method is used in cases where it is delicate or difficult to get the information from the respondents due to unwillingness or indifference. The information about the respondent is collected by interviewing the third party who knows the respondent well.
- **Local correspondents' method:** In this method, the investigator appoints local agents or correspondents in different places. They collect the information on behalf of the investigator in their locality and transmit the data to the investigator or headquarters.
- **Mailed questionnaire method:** In this method, a questionnaire is prepared. These questionnaires are addressed to individual informants and sent by post. They are requested to answer the questions and post back to the investigator.
- **Schedule sent through enumerators:** In this method, the trained enumerators or interviewers take the schedules themselves, contact the informants, get replies, and fill them in their own handwriting. This method is suitable when the respondents include illiterates.

10. Define secondary data and discuss different sources of secondary data.

Secondary data is collected and processed by some other agency, but the investigator uses it for his study. They can be obtained from published sources such as government reports, documents, newspapers, books written by economists or from any other source. Secondary data can be classified under the following two headings:

1) Published sources:

- Reports and publications of ministers, departments of the government.
- Reports and publications of reputed INGO's such as UBDP, ADB, WHO, IMF etc.
- Reports and publications of reliable NGO's, journals etc.

2) Unpublished sources

- Records maintained by government offices.
- Records maintained by research institutions, research scholars etc.
- Records updated by the department institutions for their internal purpose.

11. What are the limitation of statistics?

The limitation of statistics are as follows:

- Statistics does not deal with individuals.
- Statistics does not study qualitative phenomena.
- Statistical laws are not exact.
- Statistics can be misused.

12. Discuss the important of statistics on business and industry.

Statistics focuses on the study and manipulation of data, as well as gathering, documenting, reviewing, analyzing, and drawing conclusions from it. Statistical methods are widely used in business and trade solutions such as financial analysis, market research and manpower planning. Every business establishment irrespective of the type must adopt statistical techniques for its growth. They estimate the trend of prices, buying and selling, importing, and exporting of goods using statistical methods and past data. In any business enterprise, statistical methods can be used for three major purposes. Among them are:

- **Operational planning:** This may be done for special projects or for the recurring activities of a firm over a given period of time.

- **Setting standards:** For instance, setting out standards for the size of employment, sales volume, product quality specifications, and production output.

- **Control:** This involves comparing actual production with a norm or target established earlier. When production falls short of the target, it provides remedial measures so that such a lapse does not happen again.

13. What do you mean by data? What are the differences between primary and secondary data?

Data is a collection of discrete or continuous values that convey information, describing the quantity, quality, fact, statistics, other basic units of meaning, or simply sequences of symbols that may be further interpreted.

Difference between Primary data and Secondary data

Parameters of Comparison	**Primary Data**	**Secondary Data**
Definition	It is the crude form of all the data.	It is a refined form of data.
Source	It can be collected using various methods like interviews, experiments, etc.	It can be obtained from the internet, journals, etc.
Authenticity	It is very authentic in relation to the topic concerned.	It may be biased. It depends on the biases of the researcher.
Cost of collection	It is very costly to collect such data.	It costs very little or nothing.
Purpose	The primary purpose of the data is to add new knowledge.	It is a manipulated form of data and just tells the same story from a different perspective.

14. What are the differences between population and sample?

Basic for comparison	Population	Sample
Meaning	Population refers to the collection of all elements possessing common characteristics, that comprises universe.	Sample means a subgroup of the members of population chosen for participation in the study
Includes	Each and every unit of the group	Only a handful units of population
Characteristics	Parameter	Statistic
Data collection	Complete enumeration or census	Sample survey or sampling
Focus on	Identifying the characteristics	Making inferences about population.

15. Define mutually exclusive and independent events.

Mutually exclusive events: Two or more events are said to be mutually exclusive, when the occurrence of any one event excludes the occurrence of other event. Mutually exclusive events cannot occur simultaneously. E.g., Thus, if a coin is tossed, either the head can be up, or tail can be up; but both cannot be up at the same time.

Independent event: A set of events is said to be independent, if the occurrence of any one of them does not, in any way, affect the Occurrence of any other in the set. For example, when we toss a coin twice, the result of the second toss will in no way be affected by the result of the first toss.

16. Define the term sample space and exhaustive number of events.

Sample space: The set or aggregate of all possible outcomes is known as sample space. For example, when we roll a die, the possible outcomes are 1, 2, 3, 4, 5, and 6; one and only one face come upwards. Thus, all the outcomes— 1, 2, 3, 4, 5 and 6 are sample space. And each possible outcome or element in a sample space called sample point.

Exhaustive number of events: The total number of possible outcomes of a random experiment is called exhaustive events. The group of events is exhaustive, as there is no other possible outcome. Thus, tossing a coin, the possible outcome is head or tail; exhaustive events are two. Similarly throwing a die, the outcomes are 1, 2, 3, 4, 5 and 6. In case of two coins, the possible number of outcomes are 4 i.e. (2^2), i.e., HH, HT TH and TT.

17. What is confidence interval? What do you understand by 95% confidence level?

The **confidence interval** is the range of values that you expect your estimate to fall between a certain percentage of the time if you run your experiment again or re-sample the population in the same way. It is given by,

C.I. = $\bar{X} \pm Z_\alpha$ * S.E. ($\bar{X}$)

95% confidence level means, you are confident that 95 out of 100 times the estimate will fall between the upper and lower values specified by the confidence interval.

18. Describe the importance of sampling.

- It helps to collect vital information more quickly and it helps to make estimates of the characteristics of the total population in a shorter time.
- Sampling can save time and money.
- Sampling techniques often increases the accuracy of the data. With small samples it become easier to check the accuracy of the data.
- If the population is too large, or hypothetical sampling is the only method to be used.

19. What are the difference between Null hypothesis and Alternative hypothesis?

Difference between Null and Alternative hypothesis

Null hypothesis	**Alternative hypothesis**
A null hypothesis represents the hypothesis that there is **"no relationship" or "no association" or "no difference" between two variables.**	An alternative hypothesis is the opposite of the null hypothesis where **we can find some statistical importance or relationship between two variables.**
In case of null hypothesis, researcher tries to invalidate or reject the hypothesis.	In an alternative hypothesis, the researcher wants to show or prove some relationship between variables.
It is an assumption that specifies a possible truth to an event where there is **absence of an effect.**	It is an assumption that describes an alternative truth where there is **some effect** or some difference.
Null hypothesis is a statement that **signifies no change**, no effect, and no differences between variables.	Alternative hypothesis is a statement that **signifies some change, some effect, and some differences** between variables.
If null hypothesis is true, any discrepancy between observed data and the hypothesis is only due to chance.	If alternative hypothesis is true, the observed discrepancy between the observed data and the null hypothesis is not due to chance.
A null hypothesis is denoted as H_0.	An alternative hypothesis is denoted as H_1 or H_A.
Example of null hypothesis: There is no association between use of oral contraceptive and blood cancer. $H_0: \mu = 0$	**Example of an alternative hypothesis:** There is no association between use of oral contraceptive and blood cancer. $H_A: \mu \neq 0$