1.What is statistics?

**Statistics** is the study of the collection, analysis, interpretation, presentation, and organization of data

2.What is the difference between inferential and descriptive statistics?

**Descriptive Statistics:** Data gathered about the group and drawing conclusions about the same group

**Inferential Statistics**: Data gathered from the sample and the statistics generated to reach conclusions about the population from which the sample is taken.

3.what are quantitative and qualitative data?

**Quantitative data** are measures of values or counts and are expressed as numbers. Quantitative data refers to numerical data

**Qualitative data** are measures of ‘types’ and may be represented by a name, symbol, or a number code. Qualitative data is also known as categorical data.

4.what is population and sample?

**Population**: A set of similar items or events which is interest for some question or experiment.

**Sample**: A finite subset of the population that is selected to represent the entire group. A sample is usually selected because the population is too large or costly to study the entire items.

5.what are measures of central tendency?

**Mean**: it is calculated by taking average of numbers

**Median**: it is the central value of a given set of data.it divides the items in two equal parts.

If there is no middle element take average of two middle items in the set.

**Mode** : most occurring element in the set of data.

6.what are measures of dispersion?

**Dispersion** refers distribution of set of data. measure of dispersion shows variation of data from one another

Range, inter quartile range, standard deviation, variance

7.What is range?

**Range**: it is the difference between maximum and minimum values from set of data.it shows how dispersed the data.

Range= max-min

8.what is inter quartile range?

**IQR**: it is a statistical dispersion which is spread of data.

Divide data in four equal parts in ascending order.

We use three values q1,q2,q3

Q2 is median of data

Q1 is middle number between smallest no and median

Q3 is middle number between largest no and median

IQR=Q3-Q1

9.what is upper whisker value and lower whisker value ?

**Upper whisker value:** the boundary of the upper whisker is the maximum value of the data set

**Lower whisker value**: the boundary of the lower whisker is the minimum value of the data set

Upper whisker= Q3 + (1.5 \* IQR)

Lower whisker= Q1 – (1.5 \* IQR)

10.What is Outliers?

An **outlier** is a data point that lies outside the overall pattern of a dataset, significantly differing from other observations*.* this values results in skewing and wrong conclusion in our studies.

11.what is variance and standard deviation?

**Variance** measures how far each number in a dataset is from the mean. it is the average squared deviation from the mean.

**Standard deviation** is just the **square root** of variance.  
It tells us the **average distance** of each number from the mean.

12. what are inliers?

 A data point that is within the normal distribution of other values.

13. what is normal distribution?

**Normal distribution** also known as the Gaussian distribution, is a bell-shaped frequency distribution curve. Most of the data values in a normal distribution tend to cluster around the mean.

14.what are skewness?

**Skewness** quantifies a asymmetry in a symmetrical bell curve or normal distribution, in a set of data. If the curve is shifted to the left or to the right, it is said to be skewed.

15.what is kurtosis?

**Kurtosis** The degree of tailedness of a distribution is measured by kurtosis.It tells us the extent to which the distribution is more or less outlier-prone (heavier or light-tailed) than the normal distribution.

16. what is sampling?

**Sampling** is inferential statistics taking a subset of larger population to estimate characteristics of the whole population from that sample.

17.Different types of sampling?

**Simple random sampling** – taking random individuals everyone have equal probability

**Systematic sampling** – taking individuals at specific interval from total population.

**Stratified sampling** – population is divided in to groups then select a random sample from each group.

**Cluster sampling**- population is divided in to clusters(groups) then instead of selecting individual take a random cluster.

18. Types of quantitative data?

**Discreate Data:** The data that have fixed value is called **discreate data**, discreate data can easily be counted.

**Continuous Data**: The data that has no fixed value and has a range of data is called continuous data.

19.Types of qualitative data?

**Nominal data**: It includes named or labeled data and does not consider numerical values.

**Ordinal data**: it is a type of categorical data where the values have a meaningful order or ranking.

20.what is correlation and covariance?

**Correlation** is a standardized measure of the strength and direction of the linear relationship between two variables.

It takes values between -1 to +1, wherein values close to +1 represents strong positive correlation and values close to -1 represents strong negative correlation.

Positive corelation-strong,weak

Negative corelation-strong,weak

Zero corelation

**Covariance** is a statistical measure that indicates the direction of the linear relationship between two variables.

Positive covariance

Negative covariance

Zero covariance