**Interview questions for Descriptive statistics:**

What is the use of mean in statistics?

Mean is used to find the average of the given collection of data set. It is equal to the sum of all the values in the group of data divided by the total number of values.

1. What are the types of mean?

The types of the mean are:

Arithmetic Mean, Weighted Mean, Geometric Mean, and Harmonic Mean.

1. What is the use of a weighted mean?

Weighted mean are useful in our daily life (e.g., assignments, exams, projects, etc.)

The weighted mean is used to calculate the average of a given data set where the weight of each observation

1. What is mean?

Mean is the average of a collection of values. We can calculate the mean by dividing the sum of all observations by the number of observations.

1. Is arithmetic mean and weighted mean the same?

No, both are different. However, if all the observation carry the same weight then the weighted means generally behave in a similar fashion to arithmetic means.

1. Can you calculate Geometric Mean based on Arithmetic Mean?

Geometric Mean cannot be based on Arithmetic Mean as GM calculates the nth root of the product of ‘n’ numbers. The only thing you can say is that the geometric mean is smaller or equal to the arithmetic.

1. How is Mode useful?

The mode is useful when the most common item, A mode is a data functionality used for segregating data on a set parameter of repeating frequency.

1. Can there be NO Mode?

Yes, if the given data set does not have any value repeating itself more than once, it is no -mode set.  For example, the dataset: 1, 3, 5, 8, 7, 12, 16, 18, does not have any repeating number, and hence it does not have any mode.

1. How many modes can be there in a data set?

A data set can be multimodal or no mode. It all depends on how many different values repeat most frequently.

1. What are the types of Mode?

There are three types of modes: Bimodal, trimodal, and multimodal.

1. Does mode refer to the value that repeats maximum times?

Yes, a mode is a value that has the highest frequency in a data set.

1. How is the median calculated?

Arrange the number in numerical order.

* The number in the middle is the median. If there are two middle numbers, add them and divide by two.

1. How can we calculate a two number median?

Add the two intermediates and divide by 2 if there is an equal number of digits. The median would be the outcomes.

1. What is the best route for the medians to be found?

Arrange all numbers in ascending order to find the median, and function in the middle by marking off number at either end. If there are many data, add 1 to the number of

individual items and then split it by 2 to decide which data elements will be median.

1. How can you find a wide range of numbers in the median?

In an attempt to discover the median, arrange the data from smallest to largest first. Find the middle number then.

1. How is the median age calculated?

Subtract the smaller sum of the range from the larger number to determine the median, and split the figure by two. And apply this result to the range’s smaller total.

1. What is the meaning of standard deviation?

A low value of standard deviation is an indication of the data being close to the mean, and a high value indicates that the data is spread to extreme ends.

1. How are they different in population and sample in inferential statistics?

* We calculate the statistics using the sample.
* Using these sample statistics, we make conclusions about the population.

1. What is the relationship between mean and Median in a normal distribution?

In a normal distribution, the mean is equal to the median. To know if the distribution of a dataset is normal, we can just check the dataset's mean and median.

1. What is skewness?

Skewness measures the lack of symmetry in a data distribution. It measures the deviation of the given distribution of a random variable from a symmetric distribution. Skewed data cannot be used to create a normal distribution.

1. What is harmonic mean?

Harmonic means are terms that are between any two non-consecutive terms of a harmonic sequences. There is no general formula for the sum of any number of terms in HP.Harmonic Mean is the reciprocal of the arithmetic mean of the reciprocals

1. What is kurtosis?

Kurtosis is used to describe the statistical measure that defines one tail of distribution versus the other. It is actually the measure of outlier present in the distribution. A high value of kurtosis represents large amount of outliers being present in data.we have to either add more data into the dataset or remove the outliers.

1. What are quantitative data and qualitative data?

* Quantitative data is also known as numeric data.
* Qualitative data is also known as categorical data.

1. What are left-skewed and right -skewed distribution?

A left-skewed distribution is one where the left tail is longer than that of the right tail. Here, it is important to note that the mean < median < mode.

Similarly, a right-skewed distribution is one where the right tail is longer than the left one. But, here the mean > median > mode.

1. What is the meaning of covariance?

Covariance can be defined as a measure of how much two random variables vary together. The systematic relation is determined between a pair of random variables to see if the change in one will affect the other variable in the pair or not.

1. What is the difference between the 1st quartile, the 2nd quartile, and the 3rd quartile?

The boundary or edge of these portions are called quartiles.

* + The lower quartile (Q1) is the 25th percentile
  + The middle quartile (Q2) is the 50th percentile.
  + The upper quartile (Q3) is the 75th percentile.

1. Name and explain a few methods/techniques used in Statistics for analyzing the data?

Arithmetic mean:

* The **arithmetic mean** is the average of a series of numbers.
* It works well when the distribution is symmetric and there are no outliers.
* The total of all the values divided by the size of the data set. It is the most commonly used statistic of position.

Median:

* The median is the middle value in a data set. As you might guess, in order to calculate the middle, you need:

**–  first** listing the data in a numerical order  
 **– second**, locating the value in the middle of the list

Mode:

* The value that occurs the most often in a data set.
* It is rarely used as a central tendency measure

Standard deviation:

Standard Deviation is a measure of how much your data is spread out in statistics.

1. Explain about statistics branches?

The two main branches of statistics are descriptive statistics and inferential statistics.

* Descriptive statistics:

Descriptive statistics is a branch of Statistics. It uses analytical methods which provide the math to model and predict variation.It uses graphical method to help making numbers visible for communication purposes.

* Inferential statistics:

Inferential Statistics conclude from data that are subject to random variation, such as observation errors and sample variation.

1. What is the difference between the range and interquartile range?

 Range: the difference between the highest and lowest values.

 Interquartile range: the range of the middle half of a distribution.

1. What is the median and standard deviation of a distribution are 50 and 5 respectively, if each item is increased by 4?

Median will change if the observation are changes but standard deviation is unaffected by the origin. so, here the median will go up by 4 and S.D will remain same.

1. The change in which of the following terms does not affect the standard deviation?

Change in origin does not affect the standard deviation, whereas standard deviation is affected by scale.

1. Why we need 5 number summary?

Low extreme (minimum)

Lower quartile (q1)

Median

Upper quartile(q3)

Upper extreme(maximum)

1. What is the benefit of using box plot?

Shows the 5-number summary pictorially can be used to compare group of histograms.

1. What is the central limit theorem and why is it important?

Central limit theorem performs a significant part in statistical inference. The CLT is vital in statistics for two main reason -the normality assumption and the precision of the estimates.

1. What is exactly variance?

 Variance in probability theory and statistics is a way to measure how far a set of number is spread out. Moreover, we can describe how much a random variable differs from its expected value.

1. **Why is variance important?**

 It is tremendously important as a means to visualize and understand the data being considered. Besides, statistics in a sense were created to represent the data in two or three number.

1. When is the coefficient of skewness positive or negative?

If the co-efficient of skewness is a positive value then the distribution is positively skewed and when it is a negative value, then the distribution is negatively skewed.

1. What do skewness and kurtosis indicate?

Skewness:

Skewness is a measure of a distribution’s symmetry about its mean.it can be positive, negative or zero.

Kurtosis:

Kurtosis tells us the heaviness of both tails of a distribution. The kurtosis of a univariate normal distribution is three.

1. **How do you calculate the variance?**

 We can define variance as the average of the squared differences from the mean. Besides, for calculating the variance follow these steps:

Firstly, work out the mean (simple average of the mean)

Next, subtract the mean and square the result for each number.

After that, work out the average of those squared differences.