**1 What is the difference between population and sample in statistics?**

A population is the entire group that you want to draw conclusions about. It can mean a group containing elements of anything which is to be studied & data that possesses a common trait. , such as objects, events etc.

A sample is the specific group that the data is collected from. It is a subset of the entire population thus the size of the sample is always less than the size of the population.

**2.Define the term 'mean' in statistics and provide the formula for its calculation.**

Mean/ average is the measure of central tendency. In statistics, the mean is the average of a data set. It is calculated by adding all the values in the data set and dividing by the number of values in the set.

Mean is denoted as a bar on x or xˉ.

Dataset is X = {x1, x2, x3, . . ., xn}

The number of values = n

xˉ={x1 + x2 + x3+ . . . +xn}/ n

xˉ=∑xi/n​​

**3.Explain the concept of standard deviation and its significance in statistics.**

The standard deviation is a measure of the amount of variation or dispersion of a set of values. Standard deviation helps us understand how much individual values in a data set differ from the mean, thus it is used to assess the consistency or variability of a set of data.

A low standard deviation indicates that the values tend to be close to the mean of the set, while a high standard deviation indicates that the values are spread out over a wider range.

**4.In a given data set, what is the interquartile range (IQR) and how is it calculated?**

It is a measure of dispersion based on the lower and upper quartile. It tells the range of the middle half of a data set.

To calculate the interquartile range-

Calculate the median of the entire dataset.

Find the median of the lower half of the dataset (Lower Quartile) and the median of the upper half of the dataset (Upper Quartile).

The interquartile range is the difference between upper and lower quartiles.

Interquartile range = Upper Quartile – Lower Quartile = Q­3 – Q­1

Q1 – Lower Quartile

Q3 – Upper Quartile

**5.What is the purpose of conducting a hypothesis test in statistics?**

Hypothesis tests are used to provide direction for research by identifying the expected outcome.

It assess whether a difference between two samples represents a real difference between the populations from which the samples were taken. A null hypothesis of 'no difference' is taken as a starting point, and we calculate the probability that both sets of data came from the same population.

Hypothesis testing helps us estimate the sampling error and factor it into the test results, and then make conclusions about the population.

**6.What is the difference between correlation and causation in statistics? Provide an example to illustrate.**

Causation indicates that one event is the result of the occurrence of the other event; i.e. there is a causal relationship between the two events.

A correlation is simply a relationship where action of one element relates to action other , but one event doesn't necessarily cause the other event to happen. These variables change together but not due to each other.

Example of correlation without causation is the relationship between ice cream sales and fires. Both tend to increase during the summer, but one does not cause the other.

Example of causation is the relationship between smoking and lung cancer. Studies have shown that smoking causes an increased risk of developing lung cancer.

**7.Explain the concept of p-value in statistical hypothesis testing and its significance**

A p-value measures the probability of obtaining the observed results, assuming that the null hypothesis is true, that is, how likely it is that your set of observations could have occurred under the null hypothesis. The lower the p-value, the greater the statistical significance of the observed difference.

If a p-value reported from a t test is less than 0.05, then that result is said to be statistically significant. If a p-value is greater than 0.05, then the result is insignificant.

**8.Provide a detailed explanation of the Central Limit Theorem and its importance in statistics.**

The central limit theorem says that the sampling distribution of the mean will always be normally distributed, as long as the sample size is sufficiently large.

The central limit theorem is useful when analysing large data sets because it allows one to assume that the sampling distribution of the mean will be normally-distributed in most cases. Thus allowing for easier statistical analysis.

**9.Define Type I and Type II errors in the context of hypothesis testing**

In statistics, a Type I error is a false positive conclusion, that is, rejecting the null hypothesis when it's actually true.

While a Type II error is a false negative conclusion, that is, not rejecting the null hypothesis when it’s actually false. This does not mean that the null hypothesis is accepted, rather failing to conclude there was an effect when there actually was.

**10.What is the difference between inferential and descriptive statistics?**

Descriptive statistics state facts and proven outcomes from a population, whereas inferential statistics analyse samplings to make predictions about larger populations.

Descriptive Statistics helps to organize, analyse, and present the data in a meaningful way. Inferential statistics allows allow us to compare data, and make hypotheses and predictions with it.

Descriptive Statistics is used to describe a situation but inferential statistics is used to explain the chance of occurrence of an event.

Descriptive Statistics can be achieved with the help of charts, graphs, etc. Inferential statistics can be achieved by probability.

**11.Explain the concept of z-score in statistics and its applications.**

A z-score tells us the number of standard deviations a value is from the mean of a given distribution. If a Z-score is 0, it indicates that the data point's score is identical to the mean score, negative z-scores indicate the value lies below the mean. positive z-scores indicate the value lies above the mean.

Z-scores are used to quickly and easily determine how far above or below the mean a data point is.

* It helps to identify outliers.
* Used to determine if there is a significant difference between the mean of two independent samples.