STATISTICS WORKSHEET-4

Question-1 What is central limit theorem and why is it important?

Answer- The central limit theorem says that the sampling distribution of the mean will always be normally distributed, as long as the sample size is large enough. Regardless of whether the population has a normal, Poisson, binomial, or any other distribution, the sampling distribution of the mean will be normal.

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. This allows for easier statistical analysis and inference. For example, investors can use central limits theorem to aggregate individual security performance data and generate distribution of sample means that represent a large population distribution for security returns over a period of time.

Question-2 What is sampling? How many sampling methods do you know?

Answer- sampling is a process in statistical analysing where researchers take a predetermine number of observations from a large population.

The method of sampling depends on the type of analysis being performed, but it may include simple random sampling or systematic sampling.

Question-3 What is the difference between type I and type II error?

Answer- A type I error means rejecting the null hypothesis when it is actually true, while a type II error means failing to reject the null hypothesis when its actually false.

Question-4 What do you understand by the term Normal distribution?

Answer- A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically towards either extreme. The middle of the range is also known as the mean of the distribution.

Question-5 What is correlation and covariance in statistics?

Answer- Covariance is an indicator of the extent to which 2 random variables are dependent on each other. A higher number denotes higher dependency. Correlation is a statistical measure that indicates how strongly two variables are related.

Question-6 Differentiate between univariate, Bivariate, and multivariate analysis?

Answer- Univariate statistics summarize only one variable at a time. Bivariate statistics compare two variables. Multivariate statistics compare more than two variables.

Question-7 What do you understand by sensitivity and how would you calculate it?

Answer- The sensitivity is calculated by dividing the percentage change in output by the percentage change in input.

Question-8 What is hypothesis testing? What is H0 and H1? What is H0 and H1 for two-tail test?

Answer- A statistical hypothesis test is a method of statistical inference used to decide whether the data at hand sufficiently support a particular hypothesis. Hypothesis testing allows us to make probabilistic statements about population parameters.

In hypothesis testing there are two mutually exclusive hypothesis, the null hypothesis (H0) and the Alternative Hypothesis(H1). One of these is the claim to be tested and based on the sampling results the claim will either be supported or not.

Null hypothesis(H0): the null hypothesis here is what currently stated to be true about the population. In our case it will be the average height of student in the batch is 100.

Alternate hypothesis(H1): the Alternate hypothesis is always what is being claimed.

Question-9 What is quantitative data and qualitative data?

Answer- Quantitative data are measures of values or counts and are expressed as numbers.

Quantitative data are data about numeric variables. Qualitative data are measures of types and may be represented by a name, symbol, or a number code.

Question- 10 How to calculate range and interquartile range?

Answer- The range is calculated by subtracting the lowest value from the highest value.

The IQR describe the middle 50% of values when ordered from lowest to highest. To find the interquartile range (IQR), first find the median of the lower and upper half of the data. These values are quartile 1(Q1) and quartile 3(Q3). The IQR is the difference between Q3 and Q1.

Question-11 What do you understand by bell curve distribution?

Answer- A bell curve is a type of graph that is used to visualize the distribution of a set of chosen values across a specified group that tend to have a central, normal values, as peak with low and high extremes tapering off relative symmetrically on either side.

Question- 12 Mention one method to find outliers.

Answer- Sorting method- you can sort quantitative variables from low to high and scan for extremely low or extremely high values. Flag any extreme values that you find. This is a simple way to check whether you need to investigate certain data points before using more sophisticated methods.

Question-13 What is p-value in hypothesis testing?

Answer- The p value is a number, calculated from a statistical test, that describe how likely you are to have found a particular set of observation if the null hypothesis were true. P values are used in hypothesis testing to help decide whether to reject the null hypothesis.

Question-14 What is the Binomial Probability Formula?

Answer- In probability theory and statistics, the binomial distribution with parameters n and p is the discrete probability distribution of the number of successes in a sequence of n independent experiments, each asking a yes-no question, and each with its own Boolean-valued outcome: success or failure.

Question-15 Explain ANOVA and its applications.

Answer- Analysis of variance, or ANOVA, is a statistical method that separates observed variance data into different components to use for additional test. A one-way ANOVA is used for three or more groups of data, to gain information about the relationship between the dependent and independent variables.