STATISTICS WORKSHEET-4

Q1to Q15 are descriptive types. Answer in brief.

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

Ans=The Central Limit Theorem is important for statistics because it allows us to safely assume that the sampling distribution of the mean will be normal in most cases. This means that we can take advantage of statistical techniques that assume a normal distribution.

1. What is sampling? How many sampling methods do you know?

Ans=There are two types of sampling methods: Probability sampling involves random selection, allowing you to make strong statistical inferences about the whole group. Non-probability sampling involves non-random selection based on convenience or other criteria, allowing you to easily collect data.

1. What is the difference between type1 and type II error?

Ans=Type 1 error, in statistical hypothesis testing, is the error caused by rejecting a null hypothesis when it is true. Type II error is the error that occurs when the null hypothesis is accepted when it is not true. Type I error is equivalent to false positive. Type II error is equivalent to false negative.

1. What do you understand by the term Normal distribution?

Ans= Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.

1. What is correlation and covariance in statistics?

Ans=Correlation is a measure used to represent how strongly two random variables are related to each other.  Covariance indicates the direction of the linear relationship between variables. Correlation on the other hand measures both the strength and direction of the linear relationship between two variables.

1. Differentiate between univariate ,Biavariate,and multivariate analysis.

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

1. What do you understand by sensitivity and how would you calculate it?

Ans=A sensitivity analysis determines how different values of an independent variable affect a particular dependent variable under a given set of assumptions. In other words, sensitivity analyses study how various sources of uncertainty in a mathematical model contribute to the model's overall uncertainty.

Find the percentage change in the output and the percentage change in the input. The sensitivity is calculated by dividing the percentage change in output by the percentage change in input.

1. What is hypothesis testing? What is H0 and H1? What is H0 and H1 for two-tail test?

Ans=Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter. The methodology employed by the analyst depends on the nature of the data used and the reason for the analysis. Hypothesis testing is used to assess the plausibility of a hypothesis by using sample data.

 H1: The hypothesis that we are interested in proving. Null hypothesis: H0: The complement of the alternative hypothesis.

 null hypothesis is that the mean is equal to x. A two-tailed test will test both if the mean is significantly greater than x and if the mean significantly less than x.

1. What is quantitative data and qualitative data?

Ans=Quantitative data is information about quantities, and therefore numbers, and qualitative data is descriptive, and regards phenomenon which can be observed but not measured, such as language.

1. How to calculate range and interquartile range?

Ans=Order the data from least to greatest.

Find the median.

Calculate the median of both the lower and upper half of the data.

The IQR is the difference between the upper and lower medians.

1. What do you understand by bell curve distribution ?

Ans=The term "bell curve" is used to describe a graphical depiction of a normal probability distribution, whose underlying standard deviations from the mean create the curved bell shape. A standard deviation is a measurement used to quantify the variability of data dispersion, in a set of given values around the mean.

1. Mention one method to find outliers.

Ans=Boxplots, histograms, and scatterplots can highlight outliers. Boxplots display asterisks or other symbols on the graph to indicate explicitly when datasets contain outliers. These graphs use the interquartile method with fences to find outliers**.**

1. What is p-value in hypothesis testing?

Ans=The P value, or calculated probability, is the probability of finding the observed, or more extreme, results when the null hypothesis (H 0) of a study question is true – the definition of 'extreme' depends on how the hypothesis is being tested.

1. What is the Binomial Probability Formula?

Ans=Binomial probability refers to the probability of exactly x successes on n repeated trials in an experiment which has two possible outcomes (commonly called a binomial experiment). If the probability of success on an individual trial is p , then the binomial probability is nCx⋅px⋅(1−p)n−x .

1. Explain ANOVA and it’s applications.

Ans=A common approach to figure out a reliable treatment method would be to analyse the days it took the patients to be cured. We can use a statistical technique which can compare these three treatment samples and depict how different these samples are from one another. Such a technique, which compares the samples on the basis of their means, is called ANOVA.

Grand Mean

Hypothesis

Between Group variability

Within Group variability

F-Statistic