

STATISTICS WORKSHEET-4

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

Ans: The Central Limit Theorem (CLT) states that the distribution of a sample mean that approximates the normal distribution, as the sample size becomes larger, assuming that all the samples are similar, and no matter what the shape of the population distribution is. The central limit theorem holds for the sample of size greater than or equal to 30. This theorem is very important for testing hypotheses in statistical analysis.

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

Ans: Sampling methods are the ways to choose people from the population to be considered in a sample survey. Samples can be divided based on following criteria.

1. Probability samples - In such samples, each population element has a known probability or chance of being chosen for the sample.
2. Non-probability samples - In such samples, one can not be assured of having known probability of each population element.

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

Ans: In statistics, a **Type I error** is a false positive conclusion, while a **Type II error** is a false negative conclusion.

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

Ans: Normal 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 graphical form, the normal distribution appears as a "bell curve".

5. What is correlation and covariance in statistics?

Ans: Covariance and correlation are two terms that are opposed and are both used in statistics and regression analysis. Covariance shows you how the two variables differ, whereas correlation shows you how the two variables are related.

6. Differentiate between univariate, Biivariate, 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.

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

Ans: The sensitivity is calculated by dividing the percentage change in output by the percentage change in input

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

Ans: A statistical hypothesis is an assertion or conjecture concerning one or more populations. Hypothesis testing is formulated in terms of two hypotheses: • H0: the null hypothesis; • H1: the alternate hypothesis. **Null hypothesis (H0):** The null hypothesis here is what currently stated to be true about the population.

9. What is quantitative data and qualitative data?

Ans: Quantitative data are measures of values or counts and are expressed as numbers. Quantitative data are data about numeric variables (e.g. how many; how much; or how often). Qualitative data are measures of 'types' and may be represented by a name, symbol, or a number code.

10. How to calculate range and interquartile range?

Ans: The IQR describes the middle 50% of values when ordered from lowest to highest. To find

11. What do you understand by bell curve distribution ?

Ans: A bell curve (means normal distribution) 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 relatively symmetrically on either side.

12. Mention one method to find outliers.

Ans: You can convert extreme data points into z scores that tell you how many standard deviations away they are from the mean. If a value has a high enough or low enough z score, it can be considered an outlier. As a rule of thumb, values with a z score greater than 3 or less than -3 are often determined to be outliers.

13. What is p -value in hypothesis testing?

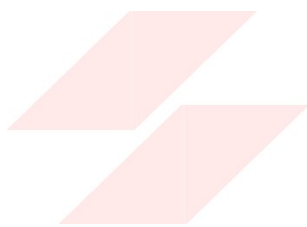
Ans: The p **value** is a number, calculated from a statistical test, that describes how likely you are to have found a particular set of observations if the null hypothesis were true. P values are used in hypothesis testing to help decide whether to reject the null hypothesis. The smaller the p value, the more likely you are to reject the null hypothesis.

14. 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 = $nCx * p^x * (1-p)^{(n-x)}$.

15. Explain ANOVA and its applications.

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



FLIP ROBO