STATISTICS WORKSHEET-4

Q1to Q15 are descriptive types. Answer in brief.

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

   Answer: The CLT is a statistical theory that states that - if you take a sufficiently large sample size from a population with a finite level of variance, the mean of all samples from that population will be roughly equal to the population mean.

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

Answer: There are two primary types of sampling methods that you can use in your research:

- 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.

3. What is the difference between type1 and typeII error?

   Answer : If Ho is true but rejected its type1 error and when Ho is false and is accepted it is typeII error.

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

   Answer: If Data points lies between -1 to +1 standard deviation it is considered to be a normal distribution.

5. What is correlation and covariance in statistics?

   Answer : When one feature is dependent or similar to another it is called correlation and vise a versa covariance means how different two variables are.

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

Answer : The analysis done on 1 variable is univariate, 2 variable is Biavariate and on multiple variables is called multivariate analysis.

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.

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

Answer: The procedure to decide whether the sample data are agreeable or consistent with the null hypothesis is called statistical hypothesis testing or simply hypothesis testing or test of significance.

Ho is null hypothesis and H1 is alternative hypothesis.

9. What is quantitative data and qualitative data?

Answer : The data which represents numbers is quantitative data and the data which represents quality of something is called qualitative data.

10. How to calculate range and interquartile range?

    Answer : The IQR describes the middle 50% of values when ordered from lowest to highest. To find the interquartile range (IQR), first find the median (middle value) 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.

11. What do you understand by bell curve distribution ?
Answer : The normal distribution which has all data points between -1 to +1 Standard Deviation is also called bell curve distribution.

12 Mention one method to find outliers.
Answer: IQR, Z-Score

13. What is p-value in hypothesis testing?

- A p-value is a statistical measurement used to validate a hypothesis against observed data.
- A p-value measures the probability of obtaining the observed results, assuming that the null hypothesis is true.
- The lower the p-value, the greater the statistical significance of the observed difference.
- A p-value of 0.05 or lower is generally considered statistically significant.

14. What is the Binomial Probability Formula?
Answer : variable that represents the number of successes in a series of nn trials. The probability mass function (PMF) gives the probability that XX equals a specific value. For instance, P(X=1)P(X=1) represents the probability of exactly one success. With this notation, we can write the binomial probability formula, which gives the probability of kk successes out of nn trials.
P(X=k)=(nk)pkqn−k

15. Explain ANOVA and it's applications.
Answer : Analysis of variance, or ANOVA, is a statistical method that separates observed variance data into different components to use for additional tests. 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. It can help in many areas of real world problems.