STATISTICS WORKSHEET-4

Q1to Q15 are descriptive types. Answer in brief.

1. What is central limit theorem and why is it important?
2. What is sampling? How many sampling methods do you know?
3. What is the difference between type1 and typeII error?
4. What do you understand by the term Normal distribution?
5. What is correlation and covariance in statistics?
6. Differentiate between univariate ,Biavariate,and multivariate analysis.
7. What do you understand by sensitivity and how would you calculate it?
8. What is hypothesis testing? What is H0 and H1? What is H0 and H1 for two-tail test?
9. What is quantitative data and qualitative data?
10. How to calculate range and interquartile range?
11. What do you understand by bell curve distribution ?
12. Mention one method to find outliers.
13. What is p-value in hypothesis testing?
14. What is the Binomial Probability Formula?
15. Explain ANOVA and it's applications.

1: The central limit theorem states that if you have a population with mean and standard deviation and take sufficiently large random samples from the population with replacement, then the distribution of the sample means will be approximately normally distributed

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, as we will see in the next section

2.
sampling is a method when researchers determine a representative segment of a larger population that is then used to conduct a study. Sampling generally comes in two forms — probability sampling and non-probability sampling

3. A type I error (false-positive) occurs if an investigator rejects a null hypothesis that is actually true in the population; a type II error (false-negative) occurs if the investigator fails to reject a null hypothesis that is actually false in the population.

4.
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 toward extreme. The middle of the range is also known as the mean of the distribution.

5.
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.

6.
Univariate statistics summarize only one variable at a time. Bivariate statistics compare two variables. Multivariate statistics compare more than two variables.

7.
Sensitivity = d/(c+d): The proportion of observed positives that were predicted to be positive.

Sensitivity is the metric that evaluates a model's ability to predict true positives of each available category.

8. the theory, methods, and practice of testing a hypothesis by comparing it with the null hypothesis.
Null hypothesis (H0): The null hypothesis here is what currently stated to be true about the population. Alternate hypothesis (H1): The alternate hypothesis is always what is being claimed.
A two-tailed test, in statistics, is a method in which the critical area of a distribution is two-sided and tests whether a sample is greater than or less than a certain range of values