WORKSHEET

STATISTICS WORKSHEET-4

Q1to Q15 are descriptive types. Answer in brief.

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

The central limit theorem states that the sampling distribution of the mean approaches a normal distribution, as the sample size increases.

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.

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

In data analysis, sampling is the practice of analyzing a subset of all data in order to uncover the meaningful information in the larger data set. There are 2 types of sampling: probability sampling and non probability sampling.

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

A type I error (false-positive) occurs if an investigator rejects a null hypothesis that is actually true in the population.

While, a type II error (false-negative) occurs if the investigator fails to reject a null hypothesis that is actually false in the population.

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

The normal distribution is a continuous probability distribution that is symmetrical around its mean with most values near the central peak

5. What is correlation and covariance in statistics?

Correlation is a statistical measure (expressed as a number) that describes the size and direction of a relationship between two or more variables. A correlation between variables, however, does not

automatically mean that the change in one variable is the cause of the change in the values of the other variable.

Covariance is a measure of the relationship between two random variables and to what extent, they change together

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

Univariate analysis is the simplest of the three analyses where the data you are analyzing is only one variable.

Bivariate analysis is where you are comparing two variables to study their relationships.

Multivariate analysis is similar to Bivariate analysis but you are comparing more than two variables.

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

The sensitivity is called the true positive rate (TPR) and is the proportion of samples that are genuinely positive that give a positive result using the test in question.

Sensitivity=(true positives)/(true positives+false negatives)

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

Hypothesis Testing is a type of statistical analysis in which you put your assumptions about a population parameter to the test.

Ho: null hypothesis

H1: alternative hypothesis

For a two tailed test:

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

9. What is quantitative data and qualitative data?

Quantitative data is the value of data in the form of counts or numbers where each data set has a unique numerical value.

Qualitative data is defined as the data that approximates and characterizes. Qualitative data can be observed and recorded.

10. How to calculate range and interquartile range?

Range is calculated from difference between the maximum and the minimum value in a given set of observations.

Interquartile range is calculated by finding the difference between the third quartile and the first quartile.

11. What do you understand by bell curve distribution ?

The normal distribution is often called the bell curve because the graph of its probability density looks like a bell. It is also known as called Gaussian distribution.

12. Mention one method to find outliers.

Zscore is a method used to find the outliers

13. What is p-value in hypothesis testing?

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

14. What is the Binomial Probability Formula?

binomial probability = nCx*px*(1−p)n−x .

15. Explain ANOVA and it’s applications.

Analysis of Variance (ANOVA) is a statistical formula used to compare variances across the means (or average) of different groups.