STATISTICS WORKSHEET-4

Q1to Q15 are descriptive types. Answer in brief.

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

A. Central limit theorem is a statistical theory which states that when the large sample size is having a finite variance, the samples will be normally distributed and the mean of samples will be approximately equal to the mean of the whole population

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

A. Selecting or picking a small portion from a large population to represent the population for research is called a sampling.So, Sample is a subset of a Population.

Sampling Methods:

Probability Sampling:

- Random
- Systematic
- Stratified
- Clustering

Non- Probabilistic:

- Convenience
- Judgement
- Snowball
- Quota

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

A . Type 1 error is when null hypothesis is true but is rejected

Type 2 error is when null hypothesis is false but we fail to reject it

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

A. 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 graphical form, the normal distribution appears as a "bell curve".

5. What is correlation and covariance in statistics?

A. Correlation is a statistical measure that indicate how strongly two variables are related. It lies between -1 to +1

Covariance is a measure indicating the extent to which two random variables change in tandem.It lies between -∞ to + ∞

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

A. Univatiate : This type of data consists of only one variable. The analysis of univariate data is thus the simplest form of analysis since the information deals with only one quantity that changes. It does not deal with causes or relationships and the main purpose of the analysis is to describe the data and find patterns that exist within it

Bivariate: This type of data involves two different variables. The analysis of this type of data deals with causes and relationships and the analysis is done to find out the relationship among the two variables

Multivariate: When the data involves three or more variables, it is categorized under multivariate. It is similar to bivariate but contains more than one dependent variable. The ways to perform analysis on this data depends on the goals to be achieved

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

A. Sensitivity measures how many positives were predicted correctly from Total (Actual)positives.

Sensitivity = Total Positives / True Positives + False Negitives

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

A. Hypothesis testing is a statistical method that is used in making a statistical decision using experimental data. Hypothesis testing is basically an assumption that we make about a population parameter. It evaluates two mutually exclusive statements about a population to determine which statement is best supported by the sample data.

Null hypothesis(H0): Null hypothesis is a general statement that there is no relationship between two measured cases or no relationship among groups

Alternative hypothesis(H1): The alternative hypothesis is the hypothesis that is contrary to the null hypothesis.

9. What is quantitative data and qualitative data?

A. Quantitative data is anything that can be counted or measured; it refers to numerical data. Qualitative data is descriptive, referring to things that can be observed but not measured—such as colors or emotions.

10. How to calculate range and interquartile range?

A. Range = The range measures the difference between the minimum value and the maximum value in a dataset

Range = Maximum value – Minimum value

Interquartile Range = The interquartile range measures the difference between the first quartile (25th percentile) and third quartile (75th percentile) in a dataset. This represents the spread of the middle 50% of values

Interquartile Range = 3rd Quartile – 1st Quartile

11. What do you understand by bell curve distribution ?

A. Bell curve is defined as a graphical depiction of a normal probability distributionwhose standard deviations from the mean form a bell-shaped curve.

12. Mention one method to find outliers.

A. Visual approaches such as histogram, scatter plot, and boxplot are the easiest method to detect outliers

One more extensively used methods to find outliers is IQR.

13. What is p-value in hypothesis testing?

A. The P-value method is used in Hypothesis Testing to check the significance of the given Null Hypothesis. Then, deciding to reject or support it is based upon the specified significance level or threshold. The P in P-value stands for Probability. The lower the p-value, the higher the chances are for Rejecting the Null Hypothesis

14. What is the Binomial Probability Formula?

A. Binomial Probability Formula :P (X) = n!/r!(n-r)!* p^x * q^(n – x)

Where,

n = Total number of trials

x = Total number of successful trials

p = probability of success in a single trial

q = probability of failure in a single trial = 1-p

15. Explain ANOVA and it’s applications

A. Analysis of variance (ANOVA) is a statistical technique that is used to check if the means of two or more groups are significantly different from each other. ANOVA checks the impact of one or more factors by comparing the means of different samples.

Applications of ANOVA:

ANOVA can be applied in various fields like medicine , marketing , finance etc.