STATISTICS WORKSHEET-4

Q1to Q15 are descriptive types. Answer in brief.

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

⇨ Central limit theorem say that the 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.

⇨ it allows one to assume that the sampling distribution of the mean will be normally-distributed in most cases.

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

⇨ Sampling means where we can divide our dataset into no. of small groups to use for research.

⇨ simple random sampling or systematic sampling.

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

➔If Ho is true and we reject Ho it is called type1 error and in if Ho is false and we fail to reject Ho it called as typeII error.

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

➔Normal distribution mean where mean=0 and the standard deviation =1 or symmetric distribution or left and right distribution

5. What is correlation and covariance in statistics?

➔ Relation between pair of random variables and change in single variable reciprocated by an equivalent change in another variable called covariance ,it take the value between -infinity to +infinity.

➔Correlation means a measure which determine the change in one variable due to change in other variable, it takes the value between -1 to 1.

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

➔univariate means where we plot each single frequency class or plot of single variable.

Bivariate means the plot between two variables.

Multivarinate means the plot between multipal variable.

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

➔it is measure of form the total no of + results, how many +ve were correctly predicted by the model.

Calculated by TP/(TP+FN).

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

➔ The testing is concerned with making decisions using data called as hypothesis testing.

H0 is an Null hypothesis where dicision always leads to stats assumption dosent change and H1 is an alternate hypothesis where decisions leads to opposite of Ho.

Our null hypothesis is that the mean is equal to x. A two-tailed test will test both if the mean is significantly greater than x and if the mean significantly less than x.and H1 is opposite to Ho.

9. What is quantitative data and qualitative data?

➔Quantitative data can be counted, measured, and expressed using numbers. Qualitative data is descriptive and conceptual. Qualitative data can be categorized based on traits and characteristics.

10. How to calculate range and interquartile range?

➔ The range is calculated by subtracting the lowest value from the highest value.

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 ?

➔ A bell curve 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.

➔Z-score is also known as standard score gives us an idea of how far a data point is from the mean. It indicates how many standard deviations an element is from the mean. In order to use a z-score, we need to know the population mean (μ) and also the population standard deviation (σ).

13. What is p-value in hypothesis testing?

➔ It is close related to confidance intervals we start with Ho and H1. Then we check wether the data rejects the Ho or fail to reject H1. Pvalue represent the probability of coefficient actually being zero.

14. What is the Binomial Probability Formula?

➔ 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 $nCx \cdot px \cdot (1-p)n-x$

15. Explain ANOVA and it's applications

➔It is used to compare difference of means among more than 2 groups. It does this by looking a variance in the data & where that variance found, specifically, ANOVA compare the amount of variance between the group with the amount of variance within group. Null hypothesis typically is that all means are equal.