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.

Ans-1 Central limit theorm said if we are trying to find mean from the different subset of some population then the mean should be identicle/similar.

Ans-2 Sampeling means to collect few data in form of subset from the data set and analyze that data. there are 8 types of sampeling in machine learning.

Ans-3 Type-1 errors are false positive errors and Type-2 errors are false negative errors.
Type 1 means- You are suffering from covid, but you actually don't
Type-2 means- You are not suffering from covid but you actually do

Ans-4 Normal distribution actually not possible in real life .Normal distribution is when mean =0 and std= 1.We call this normal distribution curve or bell curve.

Ans-5 Covariance menas when two varibles vary to each others while coorelation determines the strength and direction of relationship .

Ans-6 univariate describes analysis of one variable, Bivariate analysis took relationship of two v ariable and multivariate defines between more than 2 variables.

Ans-7 Sensitive is measurment of how machine learning model perform.we can say this true positive result.calculate by divide TP by (TP+FN)

Ans-8 Hypothesis Testing- Just the assumption of any particular data based on the evidances. we read at any conclusion.
H0- Null Hypothesis
H1- Alternative Hypothesis

pvalue>0.5 when H0 is true and H1 is false
pvalue<0.5 when H1 is true and H0 is false

Ans-9 As per name shows Quantitative data is analysed using statistical analysis and qualitative data is collected by interviewing and observing.

Ans-10 Interquartile range = Upper Quartile – Lower Quartile = Q -3 – Q -1
Range = Highest Value – Lowest Value

Ans-11 Bell curve distribution is just a normal distribution where mean=0 and std=1. Normal distribution is not possible in real time situation.

Ans-12 IQR method IQR=Q3-Q1 and Zscore method Zscore= (x-mean)/STD