STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Bernoulli random variables take (only) the values 1 and 0.
   a) True
   b) False
   Ans- a) True

2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?
   a) Central Limit Theorem
   b) Central Mean Theorem
   c) Centroid Limit Theorem
   d) All of the mentioned
   Ans- a) Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?
   a) Modeling event/time data
   b) Modeling bounded count data
   c) Modeling contingency tables
   d) All of the mentioned
   Ans- a) Modeling event/time data

4. Point out the correct statement.
   a) The exponent of a normally distributed random variables follows what is called the log- normal distribution
   b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent
   c) The square of a standard normal random variable follows what is called chi-squared distribution
   d) All of the mentioned
   Ans- c) The square of a standard normal random variable follows what is called chi-squared distribution

5. _____random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned
   Ans- c) Poisson

6. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   b) False
   Ans- b) False

7. Which of the following testing is concerned with making decisions using data?
   a) Probability
   b) Hypothesis
   c) Causal
   d) None of the mentioned
   Ans – b) Hypothesis

8. Normalized data are centered at_____and have units equal to standard deviations of the original data.

a) 0
b) 5
c) 1
d) 10
Ans- a) 0

9. Which of the following statement is incorrect with respect to outliers?
   a) Outliers can have varying degrees of influence
   b) Outliers can be the result of spurious or real processes
   c) Outliers cannot conform to the regression relationship
   d) None of the mentioned
   Ans- c) Outliers cannot conform to the regression relationship

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What do you understand by the term Normal Distribution?
Ans- 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 either extreme. The middle of the range is also known as the mean of the distribution.

11. How do you handle missing data? What imputation techniques do you recommend?
Ans- Missing data can be handle in various ways. But imputation techniques, we can get good results. Imputation is the process of substituting an estimate for missing values and analyzing the entire data set as if the imputed values were the true observed values.

I recommend, Regression and Multiple imputation techniques for handling missing data.

12. What is A/B testing?
Ans- A/B testing is a powerful tool for optimizing performance and improving user experience. It allows for data-driven decision making and helps to ensure that resources are being used effectively. However, it is important to design A/B tests carefully to ensure that they are unbiased and that the results are valid.

13. Is mean imputation of missing data acceptable practice?
Ans- Yes, but only when amount of data is small and don't have distort values.

14. What is linear regression in statistics?
Ans- Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. It assumes that the relationship between the variables is linear, which means that a change in one variable is proportional to a change in the other variable.

15. What are the various branches of statistics?
Ans- Data Collection, Descriptive and Inferential statistics.

FLIP ROBO