STATISTICS WORKSHEET-1

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

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

Ans:- 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 table
   d) All of the mentioned

Ans:- B) Modeling bounded count data

4. Point Out 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:- D) All of the mentioned

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 can be 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:- 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 graph form, normal distribution will appear as a bell curve.

11. How do you handle missing data? What imputation techniques do you recommend?

Ans:-1. Mean or Median Imputation. When data is missing at random, we can use list-wise or pair-wise deletion of the missing observations. ...

2.Multivariate Imputation by Chained Equations (MICE) MICE assumes that the missing data are Missing at Random (MAR). ...

3.Random Forest.

Imputation Techniques

- Complete Case Analysis(CCA):- This is a quite straightforward method of handling the Missing Data, which directly removes the rows that have missing data i.e we consider only those rows where we have complete data i.e data is not missing. ...
- Arbitrary Value Imputation. ...
- Frequent Category Imputation.

12. What is A/B testing?

Ans:- A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

13. Is mean imputation of missing data acceptable practice?

Ans:- True, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased. ... Since most research studies are interested in the relationship among variables, mean imputation is not a good solution.

14. What is linear regression in statistics?

Ans:- In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables)

15. What are the various branches of statistics

Ans:- Data Collection, Descriptive Statistics, Inferential Statistics.