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

Answer: 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

Answer: 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

Answer:

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

Answer: 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

Answer:

6. Usually replacing the standard error by its estimated value does change the CLT.

a) True

b) False

Answer: 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

Answer: 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

Answer: 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

Answer: c) Outliers cannot conform to the regression relationship

WORKSHEET

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

10. What do you understand by the term Normal Distribution?

Answer: The normal distribution is also known as a probability bell curve. It is symmetric about the mean and indicates that values near the mean occur more frequently than the values that are farther away from the mean.

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

Answer:Missing data can be handled either by data dropping or mean/mode imputation techniques to fill the null values .

For best result we can use knn imputer or Iterative imputer

Knn imputer:Knn imputer will try to find the relation with other columns and impute the data according the relation with other columns.

Iterative imputer:This method treat other columns which doesnot have nulls as feature and train on them and treat the null columns as label .Finally it will predict the NAN data and impute.

12. What is A/B testing?

Answer:

13. Is mean imputation of missing data acceptable practice?

Answer:It depends on size of the dataset , for example if the data set is too large then we can use mean imputation as it will not cause much difference if the missing values are less however if data set is low the it will impact the outcome if we proceed with mean imputation.

14. What is linear regression in statistics?

Answer: Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable we want to predict is called the dependent variable. The variable we are using to predict the other variable's value is called the independent variable.

15. What are the various branches of statistics?

Answer: There are two main branches of statistics which are descriptive statistics and inferential statistics.