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) False

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. b) Modeling bounded count 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

Answer. a) The exponent of normally distributed random variables follows what is called the log- normal distribution.

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

Answer. c) Poisson

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

Answer. a) True

7. 1. 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. 4. 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. b) 5

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

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.

It 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

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

Answer.

Regression/ Classification imputation. In this method, we can train regression and classification for numerical or categorical missing data column respectively and then let the model predict the missing value. Can implement this method is KNN. because it is. It takes distance between two data points in diemensional vector space into account. this method is also referred to as nearest neighbor imputation.

12.What is A/B testing?

Answer.

A/B testing the refers to the experiment where two or more variations of the same web page are compared against each other by displaying them to real time visitors to determine which one performs better for a given goal.

13.Is mean imputation of missing data acceptable practice?

Answer.

Yes it will replace the null values with the mean value of the feature. It is used for numerical features only.

14.What is linear regression in statistics?

Answer.

It is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called dependent variable. The variable you are using to predict the other variables value is called independent variable.

15.What are the various branches of statistics?

Answer.

Descriptive statistics and Inderential statistics