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: - (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

   ANS: - (D) All of the mentioned

FLIP ROBO

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: - The Normal Distribution known as Standard Normal Distribution or Gaussian Distribution It is the probability distribution that plots all of its values in a symmetrical fashion, and most of the results are situated around the probability's mean. The values are equally likely to plot either above or below the mean. It is more common to show up as a model for the lifespan for a product like a lightbulb or the output of standardized tests like height or weight are often estimated with normal distributions.

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

    ANS: - First of all we need to use deletion methods to eliminate missing data the detection methods only work for certain datasets where participants have missing fields than use regression analysis to systematically eliminate data. The important techniques is to use data imputation. This is one of the most common method of implementing values when dealing with missing data.

12. What is A/B testing?

    ANS: - A/B testing known as split testing refers to a randomized experimentation process wherein two or more versions of variables are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drive business metrics. Essentially A/B testing eliminates all the guess work out of website optimization and enables experience optimizers to make data-backed decisions.

13. Is mean imputation of missing data acceptable practice?

    ANS: - The process of replacing null values in a data collection with the data's mean is known as mean imputation. Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following example – we have a table with age and fitness score, and an eight year old has a missing fitness if we average the fitness scores of people between the age of 15 and 80 the eight year old will appear to have a significantly greater fitness level than he actually does.

14. What is linear regression in statistics?

    ANS: - Linear regression quantifies the relationship between one or more predictor variable and one outcome variable. Linear regression is commonly used for predictive analysis and modeling. The variable you want to predict is called the dependent. The variable you are using to predict the other variables value called the independent variable.

15. What are the various branches of statistics?

    ANS: - There are two branches descriptive statistics and inferential statistics.

    (1) Descriptive Statistics: - The branch of statistics that focuses on collecting, summarizing, and presenting a set of data.
    (2) Inferential Statistics: - The branch of statistics that analyzes sample data to draw conclusions about a population.

FLIP ROBO