FLIP ROBO

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

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

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

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 dealt with in a variety of ways.
Another common strategy among those who pay attention is imputation.
Mean imputation.
Substitution.
Hot deck imputation.
Cold deck imputation.
Regression imputation.
Stochastic regression imputation
Interpolation and extrapolation

12. What is A/B testing?
ANS- A/B testing also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics.

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 scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.
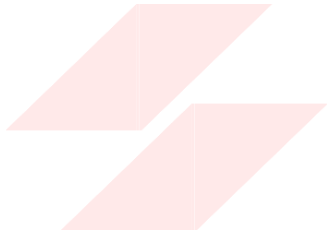
Second, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

14. What is linear regression in statistics?
ANS- Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

15. What are the various branches of statistics?
ANS- The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are employed in scientific analysis of data and both are equally important for the student of statistics.

FLIP ROBO