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)

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)

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)

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: (c)

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

   ANS: (c)

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

   ANS: (true (a))

7. Which of the following testing is concerned with making decisions using data?
   c) Probability
   d) Hypothesis

e) Causal

f) None of the mentioned

ANS: (d(Hypothesis))

7. 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

   ANS: (c)

8. 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)

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

9. What do you understand by the term Normal Distribution?
10. How do you handle missing data? What imputation techniques do you recommend?
11. What is A/B testing?
12. Is mean imputation of missing data acceptable practice?
13. What is linear regression in statistics?
14. What are the various branches of statistics?

Q9: 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 graphical form, the normal distribution appears as a "bell curve".

Q10: How do you handle missing data what imputation techniques do you recommend?

ANS:

Some of the various data imputation techniques are:

1. Next or Previous Value.
2. K Nearest Neighbors.
3. Maximum or Minimum Value.
4. Missing Value Prediction.
5. Most Frequent Value.
6. Average or Linear Interpolation.
7. (Rounded) Mean or Moving Average or Median Value.
8. Fixed Value.

Q11: What is A/B testing?

ANS:

Website A/B testing (copy, images, colors designs, calls to action), which splits traffic between two versions—A and B. You monitor visitor actions to identify which version yields the highest number of 1) conversions or 2) visitors who performed the desired action.

Q12: Is mean imputation of missing data acceptable practice?

ANS:

Mean imputation does not preserve the relationships among variables. 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.

Q13: 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.

Q14: What are the various branches of statistics?

ANS:

There are three real branches of statistics: data collection, descriptive statistics and inferential statistics. Let us look at these concepts in a little more detail. Data collection is all about how the actual data is collected.

FLIP ROBO