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:- 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:- 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:- 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:- All of the mentioned are correct statement

FLIP ROBO

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

   Ans:- Poisson random variables are used to model rates.

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

   Ans:- 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:- 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:- 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:- Normal Distribution is also known as Bell Curve. It approximates many natural phenomenon and so it has developed into a standard of reference for many probability problems. It is the most widely known distribution.

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

Ans:- The nature of missing data is important in determining what treatments can be applied to overcome the lack of data. Data can be missing in the following ways:

a. Missing Completely At Random (MCAR)
b. Missing At Random (MAR)
c. Not Missing At Random (NMAR)

Following are the imputation techniques:

a. Mean or Median Imputation
b. Multivariate Imputation by Chained Equations (MICE)
c. Random Forest

12. What is A/B testing?

Ans:- A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment. It is one of the most prominent and widely used statistical tools. It is a hypothetical testing methodology for making decisions that estimate population parameters based on sample statistics. The population refers to all the customers buying the product, while the sample refers to the number of customers that participated in the test.

FLIP ROBO

13. Is mean imputation of missing data acceptable practice?

Ans:- Yes, mean imputation of missing data is acceptable practice. 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.

14. What is linear regression in statistics?

Ans:- It is a linear approach to modelling the relationship between a scalar response and one or more explanatory known as dependent and independent variables.

15. What are the various branches of statistics?

Ans:- There are mainly two branches of statics.

a. Descriptive statistics
b. Inferential statistics

Descriptive statistics deals with the presentation and collection of data. This is usually the first part of a statistical analysis. It is usually not as simple as it sounds, and the statistician needs to be aware of designing experiments, choosing the right focus group and avoid biases that are so easy to creep into the experiment.

Inferential statistics, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics. In the end, it is the inferences that make studies important and this aspect is dealt with in inferential statistics.

FLIP ROBO