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

   Answer: 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

   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: d) All of the mentioned

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: b) False

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: 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

   Answer: c) Outliners 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:

The normal distribution, also known as the Gaussian or standard normal distribution, 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. Values are equally likely to plot either above or below the mean. It is a probability distribution that (roughly) describes many common datasets in the real world. It is the most common type of distribution, and it arises naturally in statistics through random sampling techniques.

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

Answer:

Techniques for Handling the Missing Data are:

a. List wise or case deletion.
b. Pairwise deletion.
c. Mean substitution.
d. Regression imputation.
e. Last observation carried forward.
f. Maximum likelihood.
g. Expectation-Maximization.
h. Multiple imputation.

Recommended imputation techniques are:

i. Complete Case Analysis (CCA):- This is a quite straightforward method of handling the Missing Data, which directly removes the rows that have missing data i.e. we consider only those rows where we have complete data or data is not missing.
ii. Arbitrary Value Imputation.
iii. Frequent Category Imputation.

12. What is A/B testing?

Answer:

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 a hypothetical testing methodology for making decisions that estimate population parameters based on sample statistics.

13. Is mean imputation of missing data acceptable practice?

Answer:

Yes, mean imputation is considered data acceptable practice since it ignores feature correlation.

14. What is linear regression in statistics?

Answer:

In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. 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 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?

Answer:

There are three real branches of statistics, namely:

1. **Data collection** – It is all about how the actual data is collected.

2. **Descriptive statistics** – It is the part of statistics that deals with presenting the data we have.

3. **Inferential statistics** – It is the aspect that deals with making conclusions about the data.