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

2. Which of the following theorem states that the distribution of averages of iid variables, properlynormalized, 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

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

4. Point out the correct statement.
   a) The exponent of a normally distributed random variables follows what is called the log- normaldistribution
   b) Sums of normally distributed random variables are again normally distributed even if the variablesare dependent
   c) The square of a standard normal random variable follows what is called chi-squareddistribution
   d) All of the mentioned
   Answer: D

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

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

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

8. 4. Normalized data are centered at___and have units equal to standard deviations of theoriginal data.
   a) 0
   b) 5
   c) 1
   d) 10
   Answer: A

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

10and 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, or Gaussian distribution, is a symmetric probability distribution centered around the mean, with most data points near the mean. It forms a bell-shaped curve when plotted.

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

Answer: Handling missing data can be approached by various methods including:

- **Deletion Methods:** Removing data points with missing values.
- **Mean/Median/Mode Imputation:** Replacing missing values with the mean, median, or mode of the data.
- **Regression Imputation:** Using regression analysis to estimate and replace missing values.
- **Multiple Imputation:** Creating multiple complete datasets and combining results for more robust estimates.
- **K-Nearest Neighbors (KNN) Imputation:** Using the mean or median value of the K-nearest neighbors to fill in missing data.

12. What is A/B testing?

Answer: A/B testing, or split testing, compares two versions of a webpage or app to determine which performs better. It collects user interaction data and uses statistical analysis to identify the version that yields better results, such as higher conversion rates.

1.

13. Is mean imputation of missing data acceptable practice?

Answer: Mean imputation is a straightforward and frequently used method, but it has its drawbacks. It can lead to bias and reduce variability. Although it might be suitable for minor amounts of missing data, more sophisticated techniques such as multiple imputation or model-based methods are usually advised for greater accuracy

14. What is linear regression in statistics?

Answer: Linear regression models the relationship between a dependent variable and one or more independent variables. It estimates coefficients of a linear equation to predict the dependent variable's value, represented by a straight line minimizing differences between observed and predicted values.

15. What are the various branches of statistics?
Answer:

Branch of Statistics	**Description**
Descriptive Statistics	Summarizing and describing the features of a dataset.
Inferential Statistics	Making predictions or inferences about a population based on a sample of data.
Probability Theory	Studying the likelihood of events occurring.
Regression Analysis	Modeling the relationships between variables.
Multivariate Statistics	Analyzing data that involves multiple variables simultaneously.
Time Series Analysis	Analyzing data points collected or recorded at specific time intervals.
Non-parametric Statistics	Methods that are not based on parameterized families of probability distributions.

FLIP ROBO