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
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
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
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
5. _____ random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   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
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
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

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

10. What do you understand by the term Normal Distribution?

The Normal Distribution is a bell-shaped, symmetric probability distribution for real-valued variables. It's defined by its mean and standard deviation. It's widely used in statistics due to the central limit theorem, which states that the average of many samples of a random variable converges to a normal distribution.

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

Handling missing data involves techniques like deletion, mean/median/mode imputation, prediction models, multiple imputation, and LOCF/NOCB for time-series data. The choice depends on the data and situation. It's crucial to conduct a missing data analysis before deciding on the technique.

12. What is A/B testing?

A/B testing is a research method used to compare two versions of a variable to determine which performs better. It's used in fields like marketing and web design, showing version 'A' (control) and 'B' (modified) to different user segments simultaneously. Performance is measured with metrics like click-through or conversion rates. Decisions are made based on actual user behaviour.

13. Is mean imputation of missing data acceptable practice?

While mean imputation is a simple method for handling missing data, it's not usually the best practice. It can lead to biased estimates, underestimate variances, distort relationships among variables, and is sensitive to outliers. Better methods, like multiple imputation or predictive modeling, are often recommended.

14. What is linear regression in statistics?

Linear regression is a statistical model predicting the relationship between a dependent variable and one or more independent variables. It uses the equation y = mx + b to find the best fit line, minimizing the difference between predicted and actual values. It's used for prediction and understanding variable relationships.

15. What are the various branches of statistics?
16.

Statistics has two main branches: Descriptive Statistics, which organizes and presents data, and Inferential Statistics, which makes predictions or generalizations about a population based on a sample.