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.
   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?
   Answer: a) Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?
   Answer:
   b) Modeling bounded count data

4. Point out the correct statement.
   Answer: d) All of the mentioned

5. _____ random variables are used to model rates.
   Answer: c) Poisson

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

7. 1. Which of the following testing is concerned with making decisions using data?
   Answer: b) Hypothesis

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.
   Answer: a) 0

9. Which of the following statement is incorrect with respect to outliers?
   Answerc) Outliers cannot conform to the regression relationship.

Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

- Normal distribution is a probability distribution that is symmetric around its mean, forming a bell-shaped curve. It is characterized by its mean and standard deviations.
- Symmetry: Mean, median, and mode are all equal.
- Bell-Shaped Curve: Higher frequencies of data are near the mean, tapering off as values move away from the mean.
- It's widely used in statistics to model natural and social phenomena.

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

- To deal with missing data, you can:
  - Remove Data:
  - Delete Rows: Remove rows with missing values.
  - Delete Columns: Remove columns with too many missing values.
  - Imputation Techniques:
  - Mean/Median/Mode Imputation: Replace missing values with the average, middle, or most common value.
  - Forward/Backward Fill: Use nearby values to fill in gaps.
  - K-Nearest Neighbors (KNN): Fill in missing values based on similar data points.
  - Multiple Imputation: Create several filled datasets and combine them.
  - Regression Imputation: Use regression to predict missing values.
  - Interpolation: Estimate missing values based on data trends.
- Choose a method based on the type of data, how much is missing, and its effect.

12. What is A/B testing?
Answer 12:

- A/B testing is a method for comparing two versions of something to determine which one performs better. It involves:
  - Creating Two Versions of the data, Develop two variations, A and B, of a webpage, app feature, or marketing campaign etc.
  - Splitting the Audience, Randomly divide your audience into two groups. One group sees version A, and the other sees version B.
  - Measuring Performance, Compare the feedback of version based on specific metrics, such as click-through rates, conversion rates, or user engagement.
  - Analyzing Results, Determine which version achieved better results and use that version to optimize performance.
- A/B testing helps make data-driven decisions to improve outcomes by evaluating changes incrementally.

Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.

13. Is mean imputation of missing data acceptable practice?
Answer 13:
While mean imputation can be useful for small amounts of missing data, it's important to consider more advanced imputation techniques for a more accurate analysis, especially when dealing with larger or more complex datasets.

14. What is linear regression in statistics?
Answer 14:
Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. The goal is to find the best-fitting line (or hyperplane, in the case of multiple independent variables) that predicts the dependent variable based on the independent variables.

15. What are the various branches of statistics?
Answer 15:
Statistics includes branches like descriptive, inferential, probability theory, Bayesian, multivariate, non-parametric, experimental design, time series, spatial, and statistical learning, each focusing on different aspects of data analysis and interpretation.