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)

5. **_____ random variables are used to model rates.**
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned
   Ans.(Poisson)

6. **10. 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. 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

Ans.(1)

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.(Outliers cannot conform to the regression relationship)

FLIP ROBO

WORKSHEET

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.(The **Normal Distribution** is a bell-shaped curve that shows how data points are spread around the average value. Most values are close to the average, and fewer are far away. It's used in statistics to describe patterns in data, like test scores or heights.)

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

Ans. To handle missing data, you can:

1. Remove Missing Data: Delete rows or columns with missing values if the amount is small.
2. Impute Missing Data:
   (a) Mean/Median/Mode: Fill missing values with the average (mean), middle value (median), or most common value (mode).
   (b) K-Nearest Neighbors (KNN): Use similar data points to estimate missing values.
   (c) Regression: Predict missing values based on other related data.

12. **What is A/B testing?**

Ans.(A/B testing is a method for comparing two versions of a variable (A and B) to determine which performs better. It involves showing each version to different segments of an audience and measuring which one achieves better results, such as higher conversion rates or user engagement.)

13. **Is mean imputation of missing data acceptable practice?**

Ans.(**Mean imputation** is a common and simple method for handling missing data, but it can introduce bias and reduce variability. It is acceptable for small amounts of missing data or preliminary analysis but may not be ideal for more complex situations. Alternative methods often provide better results.)

14. **What is linear regression in statistics?**

Ans.(Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. It aims to find the best-fitting line or hyperplane that predicts the dependent variable based on the independent variables. It helps in making predictions and understanding relationships between variables.)

15. **What are the various branches of The main branches of statistics are:**

Ans.

1. **Descriptive Statistics** : Summarizes and describes data using measures like mean, median, and standard deviation.

2. Inferential Statistics : Makes predictions or inferences about a population based on a sample, using techniques like hypothesis testing and confidence intervals.

3. Probability Theory : Provides the mathematical foundation for statistical inference, focusing on probability distributions and stochastic processes.

4. Bayesian Statistics : Updates probabilities and predictions based on prior knowledge and new data.

5. Mathematical Statistics : Develops theoretical properties and methods for statistical analysis.

6. Applied Statistics : Applies statistical methods to real-world problems in various fields.

7. Experimental Design : Plans experiments to ensure valid and reliable results.

8. Multivariate Statistics : Analyzes data with multiple variables simultaneously.

9. Time Series Analysis : Analyzes data collected over time to identify trends and make forecasts.

10. Nonparametric Statistics : Analyzes data without assuming a specific distribution.

15. What are the various branches of statistics?