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- C(The square of a standard normal random variable follows what is called chi-squared distribution.)

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- A(True)

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(Outliers 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- A normal distribution is a continuous probability distribution that is symmetrical on both sides of the mean. It has a bell-shaped curve and is defined by its mean and standard deviation. In a normal distribution, most of the observations cluster around the central peak and probabilities for values further away from the mean taper off equally in both directions.

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

Answer- Handling missing data involves several strategies:

Deletion Methods: Removing records with missing values (complete case analysis).

Imputation Methods: Filling in missing values with estimated ones, such as mean imputation, median imputation, or using more advanced methods like multiple imputation or predictive modeling.

Recommendation: Multiple imputation is generally recommended as it accounts for the uncertainty associated with the missing data by creating multiple datasets and combining the results.

12. What is A/B testing?

Answer- A/B testing is a statistical method used to compare two versions of a variable to determine which one performs better. It involves randomly assigning participants into two groups: the control group (A) which receives the current version, and the experimental group (B) which receives the new version. The performance of both versions is then compared using statistical analysis to decide if there is a significant difference.

13. Is mean imputation of missing data acceptable practice?

Answer- Mean imputation is a simple technique where missing values are replaced with the mean of the observed values. While it is easy to implement, it can reduce variability in the data and potentially bias the results. More advanced methods like multiple imputation or using predictive models are often preferable.

14. What is linear regression in statistics?

Answer- Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. The model assumes a linear relationship between the variables, and the goal is to estimate the coefficients that minimize the difference between the observed values and the values predicted by the

linear model.

15. What are the various branches of statistics?

Answer- Statistics can be broadly divided into two main branches:

- **Descriptive Statistics**: Involves summarizing and describing the features of a dataset.
- **Inferential Statistics**: Involves making inferences and predictions about a population based on a sample of data.

Other specialized branches include:

- **Applied Statistics**: Application of statistical techniques to real-world problems.
- **Theoretical Statistics**: Development of new statistical methods and theories.
- **Mathematical Statistics**: Focuses on the theoretical foundations of statistical inference.
- **Biostatistics**: Application of statistics to biological and health sciences.
- **Econometrics**: Application of statistical methods to economic data and problems.