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

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. Which of the following testing is concerned with making decisions using data?

Answer. b) Hypothesis

8. 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?

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. Normal distribution is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. Normal distribution also known as Gaussian distribution. In graphic form, normal distribution appears as a "bell curve". The normal distribution is important in statistics because of its mathematical properties and widespread occurrence in real-world phenomena.

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

Answer. Handling missing data is a crucial step in data analysis to ensure accurate and meaningful results. There are several imputation techniques that can be used to address missing data, depending on the nature of the data and the underlying assumptions. Some common imputation techniques are as follows:

1)Mean/Median Imputation: This helps to replace missing values with the mean or median of the available data for that variable.

2)Mode Imputation: This helps to replace missing categorical values with the mode of the available data for that category.

3)Linear regression Imputation: This predict missing values using a linear regression model based on other variables. This is suitable when there is a correlation between the variable with the missing data and other variables.

4)Multiple imputation: Create multiple imputed datasets by generating plausible values for missing data multiple times.

5)Domain specific imputation: Use domain knowledge to impute missing values based on the characteristics of the data and the context.

12. What is A/B testing?

Answers. A/B testing also known as split testing which is a statistical method used in marketing, product development, and other fields to compare two versions of a web page, advertisement, app or other elements to determine which one performs better. The goal of A/B testing is to identify changes that can improve a certain outcome, such as increasing conversion rates, click through rates or user engagement. A/B testing allows businesses and organizations to make data driven decisions by empirically testing changes before fully implementing them. It helps optimize designs, content and strategies by focusing on what works best for the target.

13. Is mean imputation of missing data acceptable practice?

Answers. Mean imputation of missing data is common and simple method for handling missing values but it has its limitations and potential drawbacks. Whether it is an acceptable practice depends on the context, the nature of the data, and the goals of your analysis. Some advantages are Ease of implementation and Preservation of sample size. Mean imputation might be acceptable in situations where the proportion of missing data is small.

14. What is linear regression in statistics?

Answers. Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. The goal of linear regression is to find the best fitting linear relationships that can be predict the value of the dependent variable based on the values of the independent variables. The linear regression model is fitted to the observed data points using a method that minimizes the sum of the squared differences between the actual and predicted values.

15. What are the various branches of statistics?

Answers. Statistics is a broad and interdisciplinary field that encompasses various branches and subfields, each focusing on different aspects of data analysis, interpretation and application. Below are the some of the branches of statistics:

1)Descriptive statistics: This branch involves summarizing and describing data using measures like mean, median, mode, range, variance, and standard deviation.

2)Inferential Statistics: Inferential statistics deals with making inferences or predictions about a population based on a sample of data.

3)Probability Theory: Probability theory studies randomness and uncertainty.

4)Statistical Inference: Statistical inference involves drawing conclusions or making predictions about a population based on sample data.

5)Biostatistics: Biostatistics focuses on the application of statistical methods to analyze and interpret biological and medical data.

6)Social statistics: Social statistics examines social phenomena and uses statistical methods to analyze and interpret social data.