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?

Answer:-

A Normal Distribution is a symmetric probability distribution that forms a bell-shaped curve. Its key properties include:

1. **Symmetry**: The curve is symmetrical around its mean.

2. **Bell-Shaped Curve**: It has a peak at the mean and tapers off gradually.

3. **Mean, Median, Mode**: These are all at the same point, the center.

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

Answer :-

Missing values can be imputed by mean/median/mode.

Mode:- If the missing values belongs to discrete value column

Mean:-If the missing values have low standard deviation

Median:-If the missing value have high standard deviation then median is more appropriate

Backward/Forward Fill:- you can fill the last known or the value after the missing value in this technique.

Domain Knowledge is necessary to find out which technique is suitable for which dataset.

12. What is A/B testing?

Answer:-

1) A/B testing is used to compare two or more versions of a particular element (such as a webpage, advertisement, or user interface) to determine which version performs better in terms of user engagement or conversion rates.

2) A/B testing is commonly used in marketing and web development to make informed decisions about which design or content variations are more effective.

3) It involves splitting users or participants into different groups, each exposed to a different version, and then analyzing the performance metrics to identify the version that yields better results.

13. Is mean imputation of missing data acceptable practice?

Answer:-

It depends on the dataset we can get the idea from standard deviation if standard deviation is high then mean imputation should be avoided because the variance in dataset is high.Apart from standard deviation we should avoid mean imputation if we have discrete data.

14.What is Linear Regression In Statistics?

Answer:-

Linear regression is a statistical method used to model the relationship between a dependent variable (also called the response variable) and one or more independent variables (also called predictor variables or features). It assumes a linear relationship between the variables, meaning that changes in the independent variables are linearly related to changes in the dependent variable.

The goal of linear regression is to find the best-fitting linear equation that describes the relationship between the variables. This equation is often represented as:

$y = mx + b$

Where y = dependent variable
M = slope,x = x is independent variable and b is y intercept

15. What are the Various Branches of statistics?

Answer:-

Statistics is a broad field that encompasses various branches, each focusing on different aspects of data analysis and interpretation. Some of the main branches of statistics include:

1. **Descriptive Statistics**: This branch deals with summarizing and presenting data in a meaningful way. It includes measures like mean, median, mode, standard deviation, and graphical representations like histograms and box plots.

2. **Inferential Statistics**: Inferential statistics involves making predictions or inferences about a population based on a sample of data. It includes techniques like hypothesis testing, confidence intervals, and regression analysis.