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)

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)

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)

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)

5.Random variables are used to model rates.

a)Empirical

b)Binomial

c)Poisson

d)All of the mentioned

Answer:- (C)

6.Usually replacing the standard error by its estimated value does change the CLT.

a)True

b)False

Answere:- (B)

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

a)Probability

b)Hypothesis

c)Causal

d)None of the mentioned

Answere:- (B)

8.Normalized data are centered at and have units equal to standard deviations of the original data.

a)0

b)5

c)1

d)10

Answere:- (A)

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

Answere:- (C)

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

Answerer:- The normal distribution, also known as the Gaussian distribution or bell curve, is a probability distribution that describes the distribution of a continuous variable. It is defined by its mean (μ) and standard deviation (σ). The shape of a normal distribution is symmetric and bell-shaped, with the majority of the observations clustering around the mean and fewer observations as the distance from the mean increases.

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

Answere:- Deletion: Remove observations that contain missing values. This method can be used when the percentage of missing values is small, but it can lead to loss of information if a large number of observations are deleted.

12.What is A/B testing?

13.Is mean imputation of missing data acceptable practice?

Answere:- Mean imputation of missing data is a simple method that can be used for both continuous and categorical variables, Mean/Mode/Median. Imputation Replace missing values with the mean, mode, or median of the variable. This is a simple method and it can be used for both continuous and categorical variables.

Mean imputation can change the distribution of the variable. For example, if the variable is skewed, the mean imputed value may not be representative of the majority of the observations

14.What is linear regression in statistics?

Answere:- Linear regression is a statistical method that is used to analyze the relationship between a dependent variable (Y) and one or more independent variables (X). The goal of linear regression is to find the line of best fit that describes the relationship between the variables

15.What are the various branches of statistics?

Answere:- Descriptive statistics

Measures of central tendency:- mean,median,mode.

Measures of variability: - The measure of variability help statisticians to analyze the distribution spread out of a given set of data. Some of the examples of measures of variability include quartiles, range, variance and standard deviation

Inferential Statistics

Regression analysis

Analysis of variance (ANOVA) Analysis of covariance (ANCOVA) Statistical significance (t-test) Correlation analysis