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

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

3. Which of the following is incorrect with respect to use of Poisson distribution?
   a) Modeling event/time data

4. Point out the correct statement.
   a) The square of a standard normal random variable follows what is called chi-squared distribution

5. random variables are used to model rates.
   a) Binomial

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True

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

   a) Hypothesis

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.
   a) 0

9. Which of the following statement is incorrect with respect to outliers?
   a) 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?

The normal distribution is a continuous probability distribution that is symmetrical around its mean, most of the observations cluster around the central peak, and the probabilities for values further away from the mean taper off equally in both directions. It also known as the Gaussian distribution.

11. How do you handle missing data? What imputation techniques do you recommend?
Deleting Rows with missing values or using imputation techniques.
Impute missing values with Mean/Median

12. What is A/B testing?
It is a way to compare the two versions of a variable to find out which performs better in a controlled environment. ... It is a hypothetical testing methodology for making decisions that estimate population parameters based on sample statistics.

13. Is mean imputation of missing data acceptable practice?
Yes, if the data are missing completely at random, the estimate of the mean remains unbiased.

14. What is linear regression in statistics?
Linear regression analysis is used to predict the value of a variable based on the value of other variables. The variable you want to predict is called the dependent variable and the other variable is called independent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values.
Yi = B0+B1X1+B2X2…BiXi

15. What are the various branches of statistics?
Descriptive statistics and Inferential statistics

Descriptive statistics

- Measures of Frequency: Count, Percent, Frequency. ...
- Measures of Central Tendency. Mean, Median, and Mode. ...
- Measures of Dispersion or Variation. Range, Variance, Standard Deviation. ...
- Measures of Position. Percentile Ranks, Quartile Ranks.

Inferential statistics

- Hypothesis testing.
- Confidence Interval.
- Contingency Tables and Chi Square Statistic.
- T-test or Anova.
- Bi-variate Regression.
- Multi-variate Regression.