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

   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

   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

   d) All of the mentioned

5. _____random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned

   c) Poisson

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

   Answer: 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

   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: Normal distribution is also knows as bell curve (the curve observed is in bell shaped) is a distribution that occurs naturally, it is assumed that during any measurement, values will follow a normal distribution with an equal number of measurements above and below the mean value.

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

    Answer: Missing data can be handled by using one of following approaches:
    a. Substituting the missing value using mean/median
    b. If the percentage of missing values is small, then it can be ignored as well
    b. Predicting missing values using Regression Imputation technique

    Imputation technique that I would recommend would be Predicting missing values using Regression Imputation technique

12. What is A/B testing?

    Answer: A/B testing also known as split testing is a statistical way of comparing 2 or more versions A/B, to not only understand which version is better but also to determine which is statistically significant

13. Is mean imputation of missing data acceptable practice?

    Answer: Ideally Mean imputation comes with problems of its own, it doesn't preserve relationship among variables or in other words it ignores co-relation. Thus it is not acceptable when variables in data have strong co-relation.

14. What is linear regression in statistics?

    Answer: Linear regression is regression model that determines relationship between independent and dependent variable. It is used to perform predictive analysis

15. What are the various branches of statistics?

    Answer: There are two main branches of statistics
    a. Descriptive Statistic : Which is used to get brief summary of data in numerical or graphical form
    b. Inferential Statistics: It is process of making inferences about a population using sample data.