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
   Ans) .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
   Ans) 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
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
   Ans) c) Chi-squared distribution.
5. _____random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned
   Ans) c) Poisson
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   b) False
   Ans) 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
   Ans) b) Hypothesis testing
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
   Ans) 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

Ans) 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?

    Ans) It is a continuous distribution or it is also called gaussian distribution distribution. It is a bell-shaped symmetric curve which is defined by the mean and standard deviation. In this distribution most of the values gather around the mean. The shape of the curve is determined by the mean and standard deviation, and about 66% of the data falls in one deviation, 95% of the data fall within 2 standard deviations and 99.7 % of the data falls within third deviation.

11. How do you handle missing data? What imputation techniques do you recommend?
    Ans) Handling missing data involves various methods like imputation by filling the values with mean median or mode of the available data or using machine learning algorithms to predict missing values. By estimating the values based on their previous history called interpolation also a techinique to handle the missing data.
    So, the best imputation technique I would recommend is Machine learning that focus on algorithms and models and make predictions.

12. What is A/B testing?
    Ans) A/B testing is a method is used to compare two versions(A and B), it can be anything such has webpage ,campaign to determine which two versions are better, by applying statistical analysis.

13. Is mean imputation of missing data acceptable practice?
    Ans) Mean imputation is a technique where the missing data are replaced by the mean of the available data but this does not work every time because it can introduce bias, especially when the missing data is not completely random.

14. What is linear regression in statistics?
    Ans) Linear regression is the simplest and most extensively used statistical technique for predictive modelling analysis. It is a way to explain the relationship between a dependent variable (target) and one or more explanatory variables(predictors) using a straight line. There are two types of linear regression - Simple and Multiple**.**

15. What are the various branches of statistics?
    Ans) Mainly it consists of two types one is descriptive statistics and another is inferential statistics
    Descriptive Statistics: This branch involves methods describing and organizing the data. It measures the mean, median, mode,range and standard deviation.
    Inferential Statistics: This branch provides us the useful conclusion or inference about a population based on a sample data. Here the technique used are hypothesis testing, confidence intervals and regression analysis for making predictions.