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)

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)

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
   Ans :- (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
   Ans :- (d)

5. _____ random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned
   Ans :- (c)

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

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)

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)

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)

Ql0and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

    Ans :- The normal distribution is fully characterized by its mean and standard deviation. This makes the distribution symmetric and it is depicted as a bell — shaped curve when plotted. A normal distribution is defined by a mean (average) of zero and a standard deviation of +1,-1.

11. How do you handle missing data? What imputation techniques do you recommend?
    Ans :- In the variable or any observation, values are not stored called as missing values / data.
    When data is missing at random, we can use list-wise or pair-wise deletion of the missing observations. However, there can be multiple reasons why this may not be the most feasible option:
    There may not be enough observations with non-missing data to produce a reliable analysis
    In predictive analytics, missing data can prevent the predictions for those observations which have missing data
    External factors may require specific observations to be part of the analysis
    In such cases, we impute values for missing data. A common technique is to use the mean or median of the non-missing observations. This can be useful in cases where the number of missing observations is low. However, for large number of missing values, using mean or median can result in loss of variation in data and it is better to use imputations. Depending upon the nature of the missing data, we use different techniques to impute data that have been described below.

12. What is A/B testing?
    Ans :- A/B testing (also known as split testing or bucket testing) is a method of comparing two versions of a webpage or app against each other to determine which one performs better. A/B testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal.

13. 1s mean imputation of missing data acceptable practice?
    Ans : True, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased. most research studies are interested in the relationship among variables, mean imputation is not a good solution.

14. What is linear regression in statistics?
    Ans :- Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

15. What are the various branches of statistics?

   Ans :- There are two main branches of statistics.

   - Descriptive Statistics :-collecting , summarizing , presenting and analyzing data ꞁ
     Descriptive statistics deals with the collection of data, its presentation in various forms, such as tables, graphs and diagrams and finding averages and other measures which would describe the data.
   - Inferential Statistics :- Using data collected from a small to draw conclusions about a longer group .
     Inferential statistics deals with techniques used for the analysis of data, making estimates and drawing conclusions from limited information obtained through sampling and testing the reliability of the estimates .