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
   ANS: 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
   ANS: d) All of the mentioned
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. 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

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: Normal distribution is a form presenting data by arranging the probability distribution of each value in the data .Most values remain around the mean value making the arrangement symmetric.

11. How do you handle missing data? What imputation techniques do you recommend?
    ANS: There is no one right method to always use to impute missing values.For Example: it is quite possible that deductive imputation could be a much better choice as compared to the regression based methods in cases where there can be logic applied to impute the missing values.
    The recommended imputation techniques are
    - Mean imputation
    - Imputation using k-NN
    - Regression imputation

12. What is A/B testing?
    ANS: A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

13. Is mean imputation of missing data acceptable practice?
    ANS: No, outliers data points will have a significant impact on the mean and hence, in such cases it is not recommended to use the mean values for replacing the missing values. Using mean values for replacing the missing values may not create a great model and hence gets ruled out.

14. What is linear regression in statistics?
    ANS: Linear regression is a linear approach to modelling the relationship between a scalar response and one or more explanatory variables.

15. What are the various branches of statistics?
    ANS: The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are equally important.