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, properlynormalized, 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 Poissondistribution?
   a) 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- normaldistribution
   b) Sums of normally distributed random variables are again normally distributed even if the variablesare dependent
   c) The square of a standard normal random variable follows what is called chi-squareddistribution
   d) All of the mentioned

5. _____random variables are used to model rates.
   a) Poisson

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

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 theoriginal data.
   a) 0

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

   a) Outliers cannot conform to the regression

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

   ANS-The Normal distribution is a continuous probability distribution that is symmetrical around itsmean and we can say that it appear as a bell curve equal from both sides this is what I understand.

2. How do you handle missing data? What imputation techniques do you recommend?
   ANS- To handle missing data in machine learning we use different methods such as MCAR,MAR,NMAR in this method based on data at what pattern it is missed and by following itspattern we can use different methods to recover it i.e imputation technique .

3. What is A/B testing?

   ANS- A/B testing is a basic randomized controlled experiment. It is a way to compare the two versionsof a variable to find out which perform better in a controlled environment. A refers to 'control' or the original testing variable and B refers to "variation" or a new version of original testing variable.

4. Is mean imputation of missing data acceptable practice?
   ANS- Yes its an acceptable practice. Because we use different methods to find outmissing data .

5. What is linear regression in statistics?
   ANS- In statistics, linear regression is a linear approach for modelling the relationship between a scalarresponse and one or more explanatory variables i.e independent and dependent variables.
   In other way it is termed as relationship between input variable and single output variable.

6. What are the various branches of statistics?
   ANS- The two main branches of statistics are descriptive and inferential statistics both of this areemployed in scientific analysis of data and both are equally important.Descriptive statistics deals with the presentation and collection of data whereas infertial statistics makes prediction in simple orI can say it use to define estimate parameters and hypothesis test.