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

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

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

7. Which of the following testing is concerned with making decisions using data?
   a) Hypothesis

8. 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 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

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

A normal distribution is **a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme**

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

12. What is A/B testing?
    A/B testing , also known as split testing, refers to a randomizes experimentation process.

13. Is mean imputation of missing data acceptable practice?
    Mean imputation is **typically considered terrible practice** since it ignores feature correlation.

14. What is linear regression in statistics?
    Linear regression is the most basic and commonly used predictive analysis. Regression estimates are used to describe data and to explain the relationship.

15. What are the various branches of statistics?
    There are three real branches of statistics: **data collection, descriptive statistics and inferential statistics**.