Statistics Assignment

1. Bernoulli random variables take (only) the values 1 and 0.

   a) True
   b) False

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

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

5. _____ random variables are used to model rates.

   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned

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

    a) True
    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

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

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

Q10-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 probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean, is called as Normal Distribution. It's graphical representation looks like a "bell curve". It is also known as the Gaussian distribution.

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

Ans: While dealing with missing data, data scientists can use two primary methods to solve the error: imputation or data removal. The imputation method substitutes reasonable guesses for missing data. It's most useful when the percentage of missing data is low.

"Mean, Median and Mode" is one of the most common methods of imputing values when dealing with missing data. In cases where there are a small number of missing observations, we can calculate the mean or median of the existing observations and insert them in place of the missing observations.

12. What is A/B testing?

Ans: A/B testing is a form of statistical and two-sample hypothesis testing, which involves determining whether the differences between the two samples are statistically significant or not.

13. Is mean imputation of missing data acceptable practice?

Ans: Mean Imputation of missing data is not an acceptable practice as -

- It does not preserve the relationships among variables
- It leads to an underestimate of standard errors

14. What is linear regression in statistics?

Ans: Linear regression is a data analysis method, used to predict the value of a variable based on the value of another variable. The variable to be predicted is called the dependent variable while, the variable being used to predict the other variable's value is called the independent variable.

15. What are the various branches of statistics?

Ans: There are 2 main branches of statistics: "Descriptive Statistics" and "Inferential Statistics".