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.
   Answer:- 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?
   Answer:- Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?
   Answer:- 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

5. _____random variables are used to model rates.
   Answer:- Empirical

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   Answer:- False

7. 1. Which of the following testing is concerned with making decisions using data?

   Answer:- Hypothesis

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.
   Answer:- 0

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

Answer:- In Normal distribution data are symmetrically distributed with no skew. The measure of central tendency (Mean, Median. Mode) are exactly same in normal distribution. is an arrangement of a data set in which most values cluster in the middle of the range and the rest taper off symmetrically toward either extreme.

11. How do you handle missing data? What imputation techniques do you recommend?
Answer:-We will use Imputers for handling of missing data.
Techniques used in imputation:-KNN imputers, Iterative imputers, mean imputers, substitution, hot desk imputation, cold desk imputation, regression imputation,

12. What is A/B testing?
Answer:- An AB test is an example of statistical hypothesis testing, a process whereby a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not.

13. Is mean imputation of missing data acceptable practice?
Answer:- Mean imputation is typically considered as terrible practices since it ignores feature correlation.

14. What is linear regression in statistics?
Answer:- Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is an explanatory variable, and the other is considered to be a dependent variable. It is a model that estimates relationship between one independent variable & one dependent variable using straight line. Both variable should be quantitative.

15. What are the various branches of statistics?
Answer:-Two major areas of statistics known as descriptive statistics & Inferential statistics.
Descriptive statistics:- which describes the properties of sample and population data**.** Descriptive statistics include mean (average), variance, skewness, and kurtosis. It measures central tendency (mean. Median, mode).

Inferential statistics:- Inferential statistics use statistical models to help you compare your sample data to other samples or to previous research.