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 theorems 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

3. Which of the following is incorrect with respect to use of Poisson distribution?
b) Modeling bounded count data.

4. Point out the correct statement.
d) All of the mentioned

5. _____ random variables are used to model rates.
c) Poisson

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

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

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.
a) 0

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

- A normal distribution has a bell shaped curve. It is symmetric around it's mean.It is also called as Gaussian distribution. The inference that we get from the distribution is that the points near the mean are frequent in occurrence.The normal distribution is the most common type of distribution assumed in technical stock market analysis and in other types of statistical analyses. The standard normal distribution has two parameters: the mean and the standard deviation.

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

- Missing data can be handled in multiple ways. The domain in which the data lies also matters while handling it.
  One of the common techniques is to replace the missing values with the mean or median of that feature. This technique is helpful if the missing values are less. This imputation technique that makes use of the mean to substitute the missing values is called mean imputation. However, there are other imputation techniques that can be used. Like,
  a)Regression Imputation: Instead of just taking the mean, you're taking the predicted value, based on other variables.

b) Interpolation and Extrapolation.
c)Hot and cold deck Imputation.

12. What is A/B testing?
    - ❖ A/B testing is a shorthand for a simple controlled experiment in which two samples (A and B) of a single vector-variable are compared. These values are similar except for one variation which might affect a user's behavior. A/B tests are widely considered the simplest form of controlled experiment. However, by adding more variants to the test, its complexity grows. A/B tests are useful for understanding user engagement and satisfaction of online features like a new feature or product. Large social media sites like LinkedIn, Facebook, and Instagram use A/B testing to make user experiences more successful and as a way to streamline their services.

13. Is mean imputation of missing data acceptable practice?
    - ❖ Yes. When the size of the data is less and if the distribution is not highly skewed, mean imputation can be used.

14. What is linear regression in statistics?
    - ❖ If we go by the meaning of the words, regression means the strength of the relationship. When the relationship between the independent and dependent variable is linear, we call it Linear Regression. Linear regression is a basic and commonly used type of predictive analysis.
      The relation between the independent and the dependent variable is given by y=mx+c. Where y= dependent variable, x=independent variable , c is the intercept(i.e the value when x is zero), m is the slope which indicates the change in y with unit change in x.
      For multiple independent features the linear function becomes a plane and then a hyperplane.

15. What are the various branches of statistics?

- The two main branches of statistics are :

  i) Descriptive statistics: Descriptive statistics in short is summarizing data. It deals with the central tendency, spread and skewness of data. Measure of central tendency describes the center of the data. This includes mean, median, mode.Measure of spread describes the dispersion of the data around it's mean.This includes variance, standard deviation, range , interquartile range etc. The measure of skewness describes the shape of the data. This includes the coefficient of skewness.

  ii) Inferential statistics: It is about inferring something about the population from it's sample.
  Inferential statistics have two main uses:

  a) Making estimates about populations.

  b) Testing hypotheses to draw conclusions about populations.