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.
   1. True
   2. False

True

1. 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?
   1. Central Limit Theorem
   2. Central Mean Theorem
   3. Centroid Limit Theorem
   4. All of the mentioned Ans=Central limit theorem
2. Which of the following is incorrect with respect to use of Poisson distribution?
   1. Modeling event/time data
   2. Modeling bounded count data
   3. Modeling contingency tables
   4. All of the mentioned

Ans B

1. Point out the correct statement.
   1. The exponent of a normally distributed random variables follows what is called the log- normal distribution
   2. Sums of normally distributed random variables are again normally distributed even if the variables are dependent
   3. The square of a standard normal random variable follows what is called chi-squared distribution
   4. All of the mentioned
2. random variables are used to model rates.
   1. Empirical
   2. Binomial
   3. Poisson
   4. All of the mentioned

Ans poission

1. 10. Usually replacing the standard error by its estimated value does change the CLT.
   1. True
   2. False
2. 1. Which of the following testing is concerned with making decisions using data?
   1. Probability
   2. Hypothesis
   3. Causal
   4. None of the mentioned

Ans hypothesis

1. 4. Normalized data are centered at and have units equal to standard deviations of the original data.
   1. 0
   2. 5
   3. 1
   4. 10

0

1. Which of the following statement is incorrect with respect to outliers?
   1. Outliers can have varying degrees of influence
   2. Outliers can be the result of spurious or real processes
   3. Outliers cannot conform to the regression relationship
   4. None of the mentioned

A

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

1. What do you understand by the term Normal Distribution?
2. How do you handle missing data? What imputation techniques do you recommend?
3. What is A/B testing?
4. Is mean imputation of missing data acceptable practice?
5. What is linear regression in statistics?
6. What are the various branches of statistics?

**Ans 10.** Normal Distribution is also called Bell shaped curve or Gaussian distribution. In perfect normal distribution mean=median=mode. In normal distribution data is symmetrical distributed with no skew.

Total area is 1 under curve. In this distribution its follow 3-sigma rule(Emprical rule /68-95-99.7).

**Ans 11.** Missing Value calculation we can do in EDA process. isnull() is widely been in all pre-processing steps to identify null values in the data. In our example, data.isnull().sum() is used to get the number of missing records in each column.

-We can delete the rows or columns with missing data or we can use imputations techniques.

Mean/median/mode imputation and KNN technique

Ans 12 A/b testing is also called split testing which decide better option

Ans 14. Linear regression is type of supervised learning learning that computes the relation between independent variables.

**Ans 15** Descriptive and Infrential

