STATISTICS WORKSHEET-1

 # Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

Q1. Bernoulli random variables take (only) the values 1 and 0.

Ans:- a) True

Q2. 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?

Ans:- a) Central Limit Theorem

Q3. Which of the following is incorrect with respect to use of Poisson distribution?

Ans:- b) Modeling bounded count data

Q4. Point out the correct statement.

Ans:-d) All of the mentioned

Q5. _____ random variables are used to model rates.

Ans:-c) Poisson

Q6. Usually replacing the standard error by its estimated value does change the CLT.

Ans:- b) False

Q7. Which of the following testing is concerned with making decisions using data?

Ans:-b) Hypothesis

Q8. Normalized data are at_____and centered have units equal to standard deviations of the original data.

Ans:-a) 0

Q9. Which of the following statement is incorrect with respect to outliers?

Ans:-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?

Ans:- The normal distribution is a probability distribution that (roughly) describes many common datasets in the real world. It is the most common type of distribution, and it arises naturally in statistics through random sampling techniques. It has symmetric **bell-shape and** mean, median are both equal and located at the center of the distribution.

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

Ans:- Missing data can be handle with a variety of ways. I believe the most common reaction is to ignore it. Choosing to make no decision, on the other had, indicates that your statistical programmed will make the decision for you. Another common strategy is imputation. Imputation is the process of substituting an estimate for missing values and analyzing the entire data set as if the imputed values were the true observed values.

I will recommend Hot deck imputation because in this technique we are limited to just feasible values. In other words, if age is only allowed to be between 5 and 10 in your research, you will always obtain a value between 5 and 10. Another factor is the random element, which introduces some variation. For exact standard errors, this is crucial.

12. What is A/B testing?

Ans:-**"A/B testing"** is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

13. Is mean imputation of missing data acceptable practice?

Ans:- Mean imputation is typically considered terrible practice since it ignores feature correlation. It preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased

14. What is linear regression in statistics?

Ans:- Linear regression strives to show the relationship between two variables by applying a linear equation to observed data. It is supposed one variable to be an independent variable, and the other is to be a dependent variable. For example, the **salary** of the person is linearly related to his **experience.**

15. What are the various branches of statistics?

Ans:- There are three branches of statistics:

1.Data collection

2. Descriptive statistics

3 Inferential statistics.