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 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

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.

b) Binomial

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?

b) Outliers can be the result of spurious or real processes

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

10. What do you understand by the term Normal Distribution?

Normal distribution is always normally distributed irrespective of sample size. In normal distribution, the mean,median and mode is equal and its 0 for all three. And the standard deviation is 1. It also be called as bell curve.

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

If the variable has 50% of the data is missing then will drop the col rather feeding to model. If less than that, we will fill the missing values using below techniques.

1.For numeric variable: In simple problem we can use mean/median to replace missing value.

2. For categorical variable we can use mode to replace missing value.

3.we can use the forward fill and backfill

4, univariate and multivariate approach.

12. What is A/B testing?

It a way to compare the performance of two variables in the controlled environment. And it's a hypothesis method to make decision that estimate the population parameters based on sample statistics.

13. Is mean imputation of missing data acceptable practice?

Yes.

14. What is linear regression in statistics?

Linear regression is to find the relationship between X and Y. it will tell us that, if x unit changes and how much y unit will change. Which means we will predict the Y variable outcome based the X values as input.

15. What are the various branches of statistics.

There are 2 branches of statistics

1. Descriptive statistics
2. Inferential statistic.