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.

   Answer: All of the mentioned

5. _____ random variables are used to model rates.

   Answer: Poisson

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

   Answer: False

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

   Answer: Hypothesis

8. 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 confirm to the regression relationship

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

Answer: Normal distribution means data is distributed normally forming a perfect bell shaped curve with no skewness. Perfect symmetrical shape that can be divided in middle to provide two equal half.

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

Answer: we can use the following imputation techniques

Simple Imputer- To fill missing NAN

KNN imputer- Will try to find the relation with other columns and impute the date according to relations.

Iterative Imputer:- This Method treat other columns which does not have nulls as feature and train on them and treat NULL columns as label.

12. What is A/B testing?

Answer: A/B testing is an experiment on two variants to see which performs better based on a given metric. It is basically a form of statistical and two-sample hypothesis testing.

13. Is mean imputation of missing data acceptable practice?

Answer: Not all the time. It is not a recommended practise as we have other imputation methods. Mean distribution can go well with a perfectly shape (bell curve) data set. But if there is lots of skewness then mean method is not advisable as skewness means outliers and mean will take outliers into consideration and this effects our model building.

14. What is linear regression in statistics?

Answer: Linear regression analysis is used to predict the value of a variable based on the value of another variable. We use regression to predict continuous data.

Building blocks of a linear regression are

1) Discreet/continuous independent variable

2) A best fit regression line

3) Continuous dependent variable

i.e. A linear regression model predicts the dependent variable using a regression line based on the independent variable

Equation of linear regression is Y=mx+c

15. What are the various branches of statistics?

Answer: There are 2 branches of statistics

Descriptive Statistics: When you are able to describe the things it is called descriptive statistics. You will only be able to describe when the population is less. Getting brief summary of a data.

Inferential Statistics: When you have to describe the things but the sample is too large. In order to describe you will use the 'sample' concept.

i.e. Take the same of population to predict/describe the result