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 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 Distribution

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

b) False (it doesn't change the central limit theorem)

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?

The Normal Distribution is the most widely known and used of all distributions. It is also known as the Gaussian Distribution. It is a bell-shaped curve and can entirely be described by its Mean and Standard Deviation. In Normal Distribution, Mean = 0 and Standard Deviation = 1. Data is symmetrically distributed with no skew, where mean, median and mode are all equal.

MEAN = MODE = MEDIAN

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

Missing data is termed as the data that is not captured for a variable for the observation. Missing data reduces the statistical power of the analysis. There are proven techniques to deal with missing data:

1) Removing Data – when dealing with data that are missing at random, related data can be deleted to reduce the bias. However, removing data may not be the good option if there are no enough observations. If too much data is discarded, it may not be possible to complete a reliable analysis.

2) Imputation – imputing means assigning a unique category to missing values. Imputation method develops reasonable guesses for missing data. It works when the percentage of missing data is low. If missing data is too high, the results lack natural variations which lead to an effective model. Imputation methods can deliver reasonably reliable results.

Here are few recommended imputation techniques:

1. Impute missing values for continuous variable
2. Impute missing values for categorical variable
3. Imputing missing values with mean, median or mode
4. Other imputation methods like Last Observation Carried Forward (LOCF) method.

12. What is A/B testing?

A/B Testing is a statistical hypothesis testing. It is used to compare two or more versions of a variable against each other to determine which one performs better. It is considered as an analytical method for making decisions that estimates population parameters based on sample statistics.

13. Is mean imputation of missing data acceptable practice?

No. Mean imputation does not preserve relationship among variables and the estimate of the mean remains unbiased, if the data remains missing completely at random. Further, it also leads to an underestimate of standard error. Hence, mean imputation is not considered as a good solution and the practice is not acceptable.

14. What is linear regression in statistics?

Linear Regression is a statistical procedure which can be used for calculating the value of a dependent variable from an independent variable. In other words, if we want to use variable X to draw conclusions concerning a variable Y.

Where, Y is dependent variable

X is independent variable.

If the relationship between 2 variables is linear it can be shown by a straight line.

Equation is as follows:

$Y = a + bX$

15. What are the various branches of statistics?

Definition of Statistics

Statistics is a science which deals with the collection, analysis and interpretation of numerical data.

There are 4 branches of statistics:

1. Mathematical Statistics - which helps in forming experimental and statistical distribution.

2. Statistical Method – which helps in collection, tabulation and interpretation of data. It also helps to analyze and return the insights from the data.

3. Descriptive Statistics – deals with the presentation and collection of data, representing data in both classification and diagrammatic way.

4. Inferential Statistics – studies a sample of the same data, helps in making predictions about a population based on the sample drawn from it.