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

Ans - 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?*

Ans - Centroid Limit Theorem

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

Ans - Modeling bounded count data

*4. Point out the correct statement.*

Ans - All of the mentioned

*5. \_\_\_\_\_\_ random variables are used to model rates.*

Ans - Poisson

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

Ans - False

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

Ans - Hypothesis

*8. Normalized data are centered at\_\_\_\_\_\_and have units equal to standard deviations of the*

*original data.*

Ans - 0

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

Ans - Outliers cannot conform to the regression relationship

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

A normal distribution is a data set structure in which the majority of values group in the centre of the distribution and the remaining cut symmetrically to either end. A basic example of a normal distribution pattern is height.

A graphical representation of a normal distribution is often referred to as a bell curve. The exact form depends on the population distribution, but the peak is always in the centre, and the curve is always symmetrical. The mean, mode, and median of a normal distribution are all the same.

Formula

μ = Mean

σ = Standard Deviation

π≈3.14159

e≈2.71828

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

Getting rid of the findings with missing data is one solution to this issue. However, you run the risk of missing data points containing critical information. Imputing the missed values would be a safer strategy. To put it another way, we must conclude certain missing values from the current records.

**Imputation Using (Mean/Median) Values -** This functions by measuring the mean/median of the non-missing values in a column and then replacing the missing values in each column individually and independently. It's just useful for numeric results.

**Imputation Using (Most Frequent) or (Zero/Constant) Values -** It substitutes missing data with the most common values within each column when dealing with categorical features (strings or numerical representations).

**Imputation Using k-NN -** To estimate the values of some new data points, the algorithm employs a methodology known as "function similarity." This means that a value is given to the new point depending on how comparable it is to the points in the training set.

*12. What is A/B testing?*

A simple randomised control experiment is A/B testing. It's a method of comparing two iterations of a variable in a managed environment to see which works better. You may either do random trials or use scientific and mathematical techniques. One of the most well-known and commonly used mathematical methods is A/B testing.

*13. Is mean imputation of missing data acceptable practice?*

Yes, we can use mean imputation to replace missing data with the most frequent values within each column but

1. It also doesn’t factor the correlations between features
2. It can introduce bias in the data.

*14. What is linear regression in statistics?*

Through applying a linear equation to observed data, linear regression demonstrates the relationship between two variables. One variable is supposed to be independent, and the other is supposed to be dependent. For example, a person's weight is proportional to his height.

A linear regression line equation is written in the form of:

**Y = a + bX**

*15. What are the various branches of statistics*

Two branches, descriptive statistics and inferential statistics,

**Descriptive Statistics** - The branch of statistics that focuses on collecting, summarizing, and presenting a set of data.

**Inferential Statistics** - The branch of statistics that analyses sample data to draw conclusions about a population.