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

   Ans - true

2) **Ques 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 - Central Limit Theorem

3) **Ques 3- Which of the following is incorrect with respect to use of Poisson distribution?**

   Ans - Modeling bounded count data

4) **Ques 4- Point out the correct statement.**

   Ans - All of the mentioned

5) **Ques 5-** _____ random variables are used to model rates

   Ans – Poisson

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

   Ans – false

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

   Ans – Hypothesis

8) **Ques 8 - . Normalized data are centered at_____and have units equal to standard deviations of the original data.**

   Ans - . 0

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

   Ans -) Outliers cannot conform to the regression relationship

10) **Ques 10 - What do you understand by the term Normal Distribution?**

Ans - It is the most important probability distribution in statistics for independent and random variable. It tell us about how data values are distributed .

The importance of normal distribution:

1) The statistical hypothesis test
2) Both linear and non-linear regression assumes that the **residual** follows the normal distribution.

3) Apart from this most of the **statistical software programs** support some of the probability functions for normal distribution.

There are two main parameter of normal distribution, Mean and std. deviation.

With the help of these two parameter we can we can decide the shape and probabilities of the distribution wrt. our problem statement. As the parameter value changes the shape and distribution changes.

11) **Ques 11- How do you handle missing data? What imputation techniques do you recommend?**

Ans – Missing data appears when no value is available in any variable. Due to Missing data, the statistical power of the analysis can reduce, which can impact the validity of the results.

Detecting missing data:

Missingno is a simple Python library that presents a series of visualizations to recognize the behavior and distribution of missing data inside a pandas data frame. It can be in the form of a barplot, matrix plot, heatmap, or a dendrogram.

The msno.matrix() is a nullity matrix that will help to visualize the location of the null observations.

The imputataion technique to handle the missing data is k_nearest neighbor imputation :

The KNN algorithm helps to impute missing data by finding the closest neighbors using the Euclidean distance metric to the observation with missing data and imputing them based on the non-missing values in the neighbors.

12)   What is A/B testing?

Ans - Split testing is another name of A/B testing and it's a common or general methodology. It's used when one wants to test a new feature or a product.

The main goal here is to design an experiment that gives repeatable results and robust to make an informed decision to launch it or not.

this test includes a comparison of two web pages by representing variants A and B for them, as the number of visitors is similar the conversion rate given by the variant becomes better. Overall, it's an experiment where two or more variations of the same web page are compared against together by showcasing them to real-time visitors, and through that determines which one performs better for a given goal. A/B testing is not only used or limited by web pages only, it can be used in emails, popups, sign-up forms, apps, and more.

13)   Is mean imputation of missing data acceptable practice?

Ans - Mean imputation: So simple. And yet, so dangerous.

Perhaps that's a bit dramatic, but mean imputation (also called mean substitution) really ought to be a last resort.

It's a popular solution to missing data, despite its drawbacks. Mainly because it's easy

Mean imputation does not preserve the relationships among variables.

Mean Imputation Leads to an Underestimate of Standard Errors.

14) What is linear regression in statistics?

    Ans -Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable that you are using to predict the other variable is called independent variable.

This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values. There are simple linear regression calculators that use a "least squares" method to discover the best-fit line for a set of paired data. You then estimate the value of X (dependent variable) from Y (independent variable).

15) What are the various branches of statistics?

    Ans – various branches of statistics are:

   I.   Discriptive statistics:

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

   II.   Inferential ststistics:

    The branch of statistics that analyzes sample data to draw conclusions about a population.