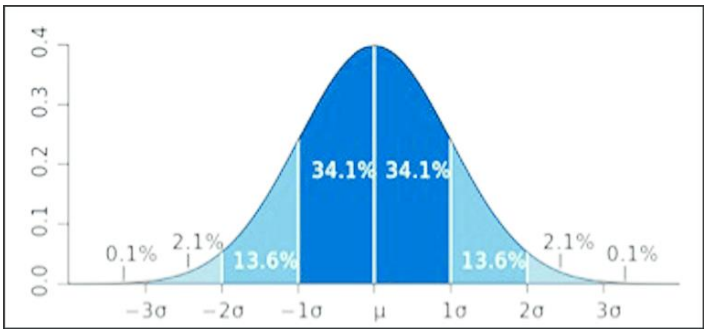
STATISTICS

1. A
2. A
3. B
4. D
5. C
6. B
7. B
8. A
9. C
10. What do you understand by the term Normal Distribution?

    In probability and statistics, the most important continuous probability distribution is the normal distribution, also known as the Gaussian distribution. Also referred to as the bell curve. A normal distribution represents an approximate or exact range of variance in all physical and economic research.

    Any positive standard deviation belongs to the normal distribution. We are aware that the mean aids in establishing the symmetry line of a graph and that the standard deviation aids in determining how widely distributed the data are. The graph gets narrower and the data are closer together if the standard deviation is less. The data are more evenly distributed and the graph becomes wider as the standard deviation increases. The area under the normal curve is divided using the standard deviations. The percentage of data that falls within each split section's respective graph region is indicated.



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

    There are number of techniques which are used to handle missing data. Few of them are

    1. Eliminating Values from data: This is very simple technique to incorporate to handle missing data. But due to using this techniques may possible to lose important information. This technique can be used if only 2-5% data is missing. Percentage of missing value can be very easily calculated using pandas library
    2. Imputation Techniques: This is very powerful technique used in the industry to replace missing of NaN values. Depending on the nature of the data and the issue it

has, there are numerous techniques to impute the missing values. Imputation techniques can be roughly categorized into the following categories:

i. Imputation with constant value: Programmer can directly replace NaN value by any constant value

ii. Imputation using Statistics: Imputation can be done using statistical approach. Simple Imputer is basically used for the same. Syntax is remains same like constant value but instead of the constant value we can use Mean, Median and Most_Frequent values from the data. Mean value is preferable when data is numerical but not skewed. Median value is preferred when data is numerical as well as skewed. Most Frequent value is preferred when data is categorical or numerical data.

For dealing with missing data I prefer the Imputation technique for handling.

12. What is A/B testing?

A/B testing is also called as Split or Bucket Testing is a method of comparing two versions of a webpage or app against each other to determine which one performs better. A/B testing is an experiment in which consumers are randomly presented two or more variations of a website, and statistical analysis is utilized to ascertain which variation performs better for a specific conversion objective.

13. Is mean imputation of missing data acceptable practice?

No, Mean Imputation of missing data is not acceptable in the industries. They lead to an underestimation of standard errors and, thus, overestimation of test statistics. The main reason is that the imputed values are completely determined by a model applied to the observed data, in other words, they contain no error

14. What is linear regression in statistics?

Linear regression analysis is used to predict the value of one variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you use to predict the value of another variable is called the independent variable.

15. What are the various branches of statistics?

Statistics is a study of presentation, analysis, collection, interpretation and organization of data
There are two main branches of statistics Inferential Statistic Descriptive Statistic.

i. Inferential Statistics:

Inferential statistics used to make inference and describe about the population. These stats are more useful when it's not easy or possible to examine each member of the population.

ii. Descriptive Statistics:

Descriptive statistics are used to get a brief summary of data. You can have the summary of data in numerical or graphical form.