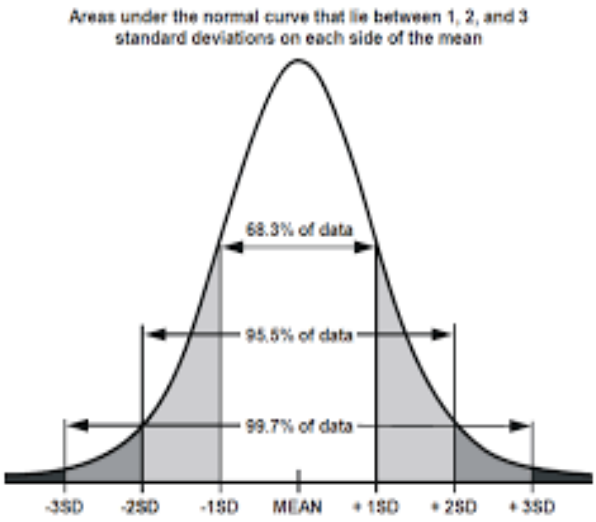
Statistics – Subjective part

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

Ans- The normal distribution, also known as the Gaussian distribution, is the most important probability distribution in statistics for independent, random variables. Most people recognize its familiar bell-shaped curve in statistical reports. For example, heights, blood pressure, measurement error, and IQ scores follow the normal distribution.



Q.11 How do we handle missing data? What imputation technique do you recommend?

Ans- A number of alternative ways of handling the missing data has been developed.

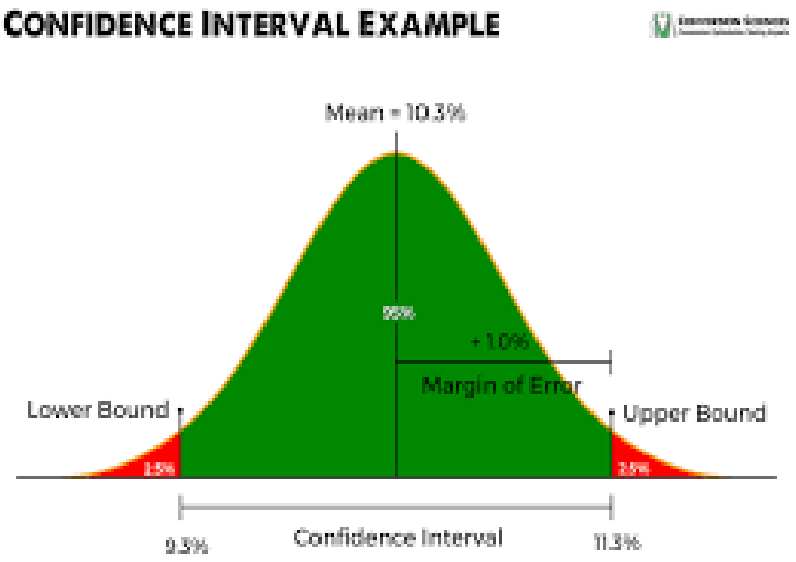
1. Listwise or case deletion
2. Pairwise deletion
3. Mean substitution
4. Regression imputation
5. Last observation carried forward
6. Maximum likelihood
7. Expectation-Maximization
8. Multiple imputation

Best techniques to handle missing data:-

- Use deletion methods to eliminate missing data. The deletion methods only work for certain datasets where participants have missing fields.
- Use regression analysis to systematically eliminate data.
- Data scientists can use data imputation techniques.

Q.12 What is A/B Testing?

Ans- A/B testing is basically statistical hypothesis testing, or, in other words, statistical inference. It is an analytical method for making decisions that estimates population parameters based on sample statistics.



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

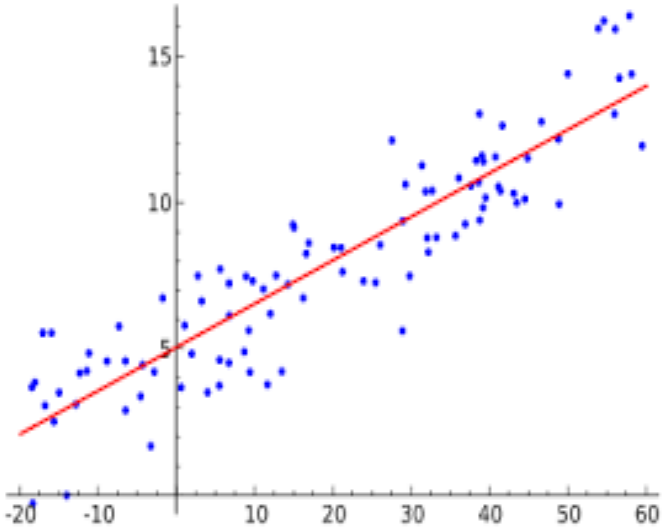
Ans- The process of replacing null values in a data collection with the data's mean is known as mean imputation. Mean imputation is typically considered terrible practice since it ignores feature correlation.

Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

Second, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

Q.14 What is linear Regression in Statistics?

Ans- In statistics, **linear regression** is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables).



Q.15 What are the various branches of Statistics?

Ans- There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.