MODULE 1

Summary statistics includes:

- Measures of central tendency are values that represent the center of a dataset.
    - mean, median, mode
- Measures of dispersion are values that represent the spread of a dataset
    - std, variance, interquartile range,
- Measures of position helps to determine the position of a value in relation to other values in the
    - dataset
    - percentile, quartile, boxplot,

Sample is:

- subset of a larger population.
- must be representative of the population
- accurately reflects the population
- data professionals uses samples to make inferences about a population

Parameter

- a characteristic of a population
- must be able to collect data about every member of the population

Sample statistic

- known value from a sample

Sampling is the process of selecting a subset of data from a population. Choosing the right sample size is crucial to get valid test results an avoid statistical errors.

A/B testing is a way to compare two versions of something to find out which version performs better

After a test is complete, you need to determine the **statistical significance** of the result. Statistical significance refers to the claim that the results are not explainable by chance alone.

Hypothesis test helps to quantify whether the result is likely due to chance or it's statistically significant.

Two main types of statistical methods:

- **Descriptive statistics**
    - describe or summarize the main feature of a dataset
    - helps top quickly understand a large amount of data
    - two types:
        - visuals
            - graph, tables, etc.
        - summary stats
            - summarize data the data using single number (mean, std, etc)
- **Inferential statistics**
    - make inferences about a dataset based on a sample of a data

Population is the dataset that the sample is drawn from. Include every possible element that are interested in.

To get a complete picture of a data, it's good to know the measures of the center (mean, median, mode) and the measure of the dispersion (range, std, variance)

Standard deviation is the measure of how spread out the data is from the mean