**Department of Computer Engineering**

**Academic Year:** 2022-2023 **Semester:** VIII

**Subject:** Applied Data Science **Class / Division:** BE/CMPN/B

**Name:** Domingos J Pinto **Roll Number:** 007

**Date: Seat Number:** BE/CMPN/B

**Experiment No.: 1**

**Explore the descriptive statistics on the given dataset.**

**Aim :** Explore the descriptive statistics on the given dataset.

**I-THEORY**

* **Statistics** is the science of collecting, analyzing, presenting, and interpreting data, as well as of making decisions based on such analyses. Statistics is at the heart of data analytics. Statistics is a branch that deals with every aspect of the data. The main purpose of using statistics is to plan the collected data in terms of experimental designs and statistical surveys. Statistics is considered a mathematical science that works with numerical data. Statistical knowledge helps to choose the proper method of collecting the data and employ those samples in the correct analysis process in order to effectively produce the results. In short, statistics is a crucial process which helps to make the decision based on the data.

**TYPES OF STATISTICS**

* ***Inferential Statistics*** consists of methods that use sample results to help make decisions or predictions about a population.
* ***Descriptive Statistics*** consists of methods for organizing, displaying, and describing data by using tables, graphs, and summary measures**.**

**Types of variables**

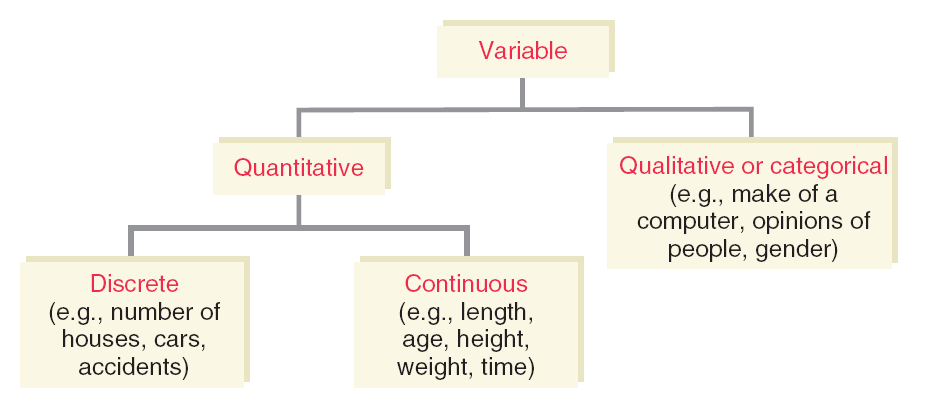
* **Quantitative Variables**
  + **Discrete Variables**
  + **Continuous Variables**
* **Qualitative or Categorical Variables**
* **Quantitative Variables:** A variable that can be measured numerically is called a *quantitative variable*. The data collected on a quantitative variable are called *quantitative data*.
* **Quantitative Variables: Discrete**
* A variable whose values are countable is called a *discrete variable*.
* In other words, a discrete variable can assume only certain values with no intermediate values.

**Quantitative Variables: Continuous**

* A variable that can assume any numerical value over a certain interval or intervals is called a *continuous variable*.

**Qualitative or Categorical Variables**

* A variable that cannot assume a numerical value but can be classified into two or more nonnumeric categories is called a *qualitative* or *categorical variable*. The data collected on such a variable are called *qualitative data*.

****

**Figure 1 :- Types of Variable**

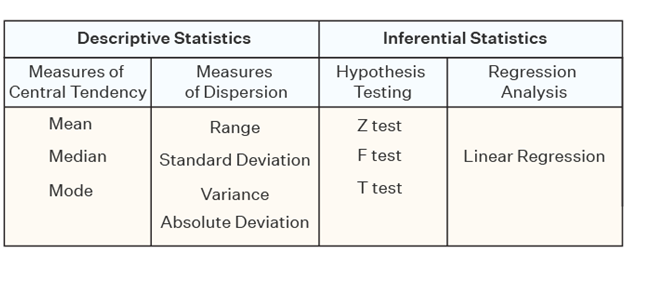
# **Descriptive and Inferential Statistics**

Descriptive and inferential statistics are two fields of statistics. Descriptive statistics is used to describe data and inferential statistics is used to make predictions. Descriptive and inferential statistics have different tools that can be used to draw conclusions about the data.

In descriptive and inferential statistics, the former uses tools such as central tendency, and dispersion while the latter makes use of hypothesis testing, regression analysis, and confidence intervals.

The purpose of descriptive and inferential statistics is to analyze different types of data using different tools. Descriptive statistics helps to describe and organize known data using charts, bar graphs, etc., while inferential statistics aims at making inferences and generalizations about the population data.

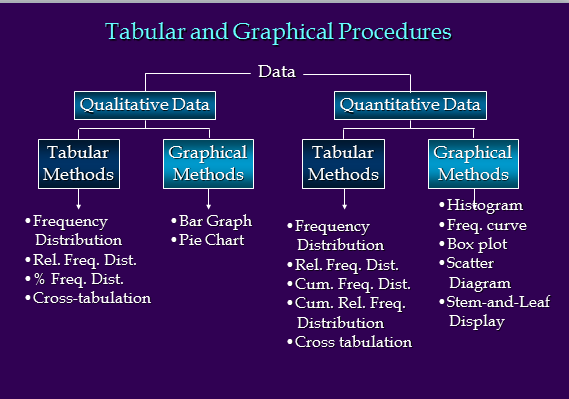
# **Table :-1 Descriptive and Inferential Statistics**



### Descriptive Statistics

Descriptive statistics are a part of statistics that can be used to describe data. It is used to summarize the attributes of a sample in such a way that a pattern can be drawn from the group. It enables researchers to present data in a more meaningful way such that easy interpretations can be made. Descriptive statistics uses two tools to organize and describe data. These are given as follows:

* **Measures of Central Tendency -** These help to describe the central position of the data by using measures such as [mean](https://www.cuemath.com/data/mean/), [median](https://www.cuemath.com/data/median/), and [mode](https://www.cuemath.com/data/mode/)**.**
* **Measures of Dispersion -** These measures help to see how spread out the data is in a distribution with respect to a central point. [Range](https://www.cuemath.com/data/range-in-statistics/), standard deviation, [variance](https://www.cuemath.com/data/variance/), quartiles, and [absolute deviation](https://www.cuemath.com/mean-deviation-formula/) are the measures of dispersion.



**Figure 2 :- Tabular and Graphical procedure**

**II-IMPLEMENTATION - Explore the descriptive statistics on the dataset.**

**Dataset :** Motor Cars

**Description :** It is a dataset of different models of motor cars. The dataset consist of eleven different properties of motor cars. There are 32 rows in this dataset.

**A) Loading the Data**

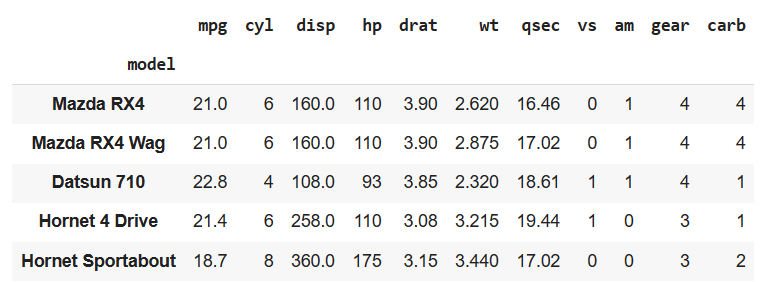
import numpy as np

import pandas as pd

mtcars = pd.read\_csv("mtcars.csv")

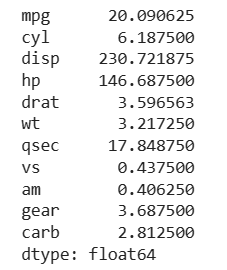
del mtcars['Unnamed: 0']

mtcars.head()

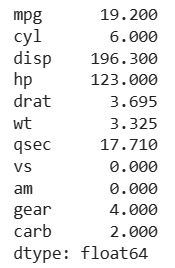


1. **Measure of Central Tendency**

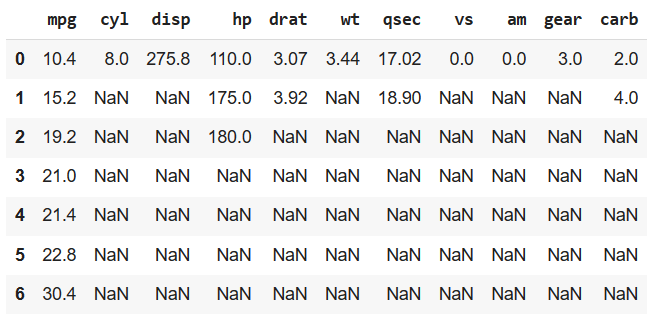
mtcars.mean() # Get the mean of each column



mtcars.median() # Get the median of each column



mtcars.mode() #Get the mode



1. **Measures of Spread / Dispersion**

max(mtcars["mpg"]) - min(mtcars["mpg"]) #Range of mpg



five\_num = [mtcars["mpg"].quantile(0),

mtcars["mpg"].quantile(0.25),

mtcars["mpg"].quantile(0.50),

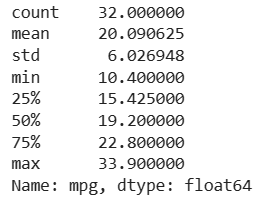
mtcars["mpg"].quantile(0.75),

mtcars["mpg"].quantile(1)]

five\_num



mtcars["mpg"].describe()



mtcars["mpg"].quantile(0.75) - mtcars["mpg"].quantile(0.25) #Interquartile (IQR) range



mtcars["mpg"].var() #variance



mtcars["mpg"].std() #standard deviation



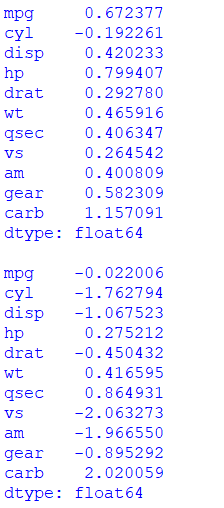
#skewness and kurtosis

print(mtcars.skew())

print()

print(mtcars.kurt())

print()



from numpy import absolute

A = 20

sum = 0 # Initialize sum to 0

# Absolute deviation calculation

for i in range(len(mtcars)):

av = absolute(mtcars["mpg"][i] - A)

# Absolute value of the differences of each data point and A

# Summing all those absolute values

sum = sum + av

# finding the absolute deviation

print(sum / len(mtcars))

