**Data Analytics using Python**

# **Week1**

## Introduction to Data Analytics

* Variables:- An entity which is capable of taking different type of values.
* Measurements:- The process which is used to assign values to the variables
* Data:- Recorded measurements

Data is important because to make better decisions, evaluate performance and to understand the consumer and the market.

Data Analysis:- It is a process of transforming, arranging and examining the raw data into a specific format to fetch valuable information from it.

Data Analysis: PAST

Data analysis is the process of examining, transforming, and

arranging raw data in a specific way to generate useful

information from it

Data Analytics: FUTURE

Data analysis is the process of examining, transforming, and

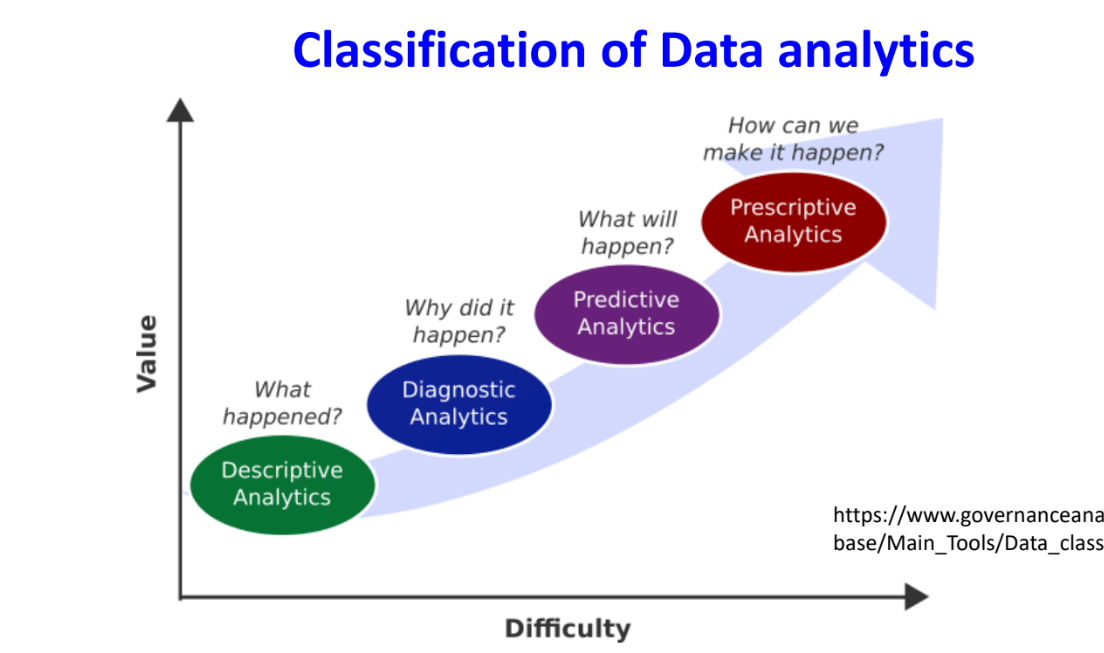
arranging raw data in a specific way to generate useful

information from it

Analysis =/ Analytics

### Classification of Data analytics

1. Descriptive Analytics
2. Diagnostic Analytics
3. Predictive Analytics
4. Prescriptive Analytics



A patient’s healthcare journey follows different analytics types:

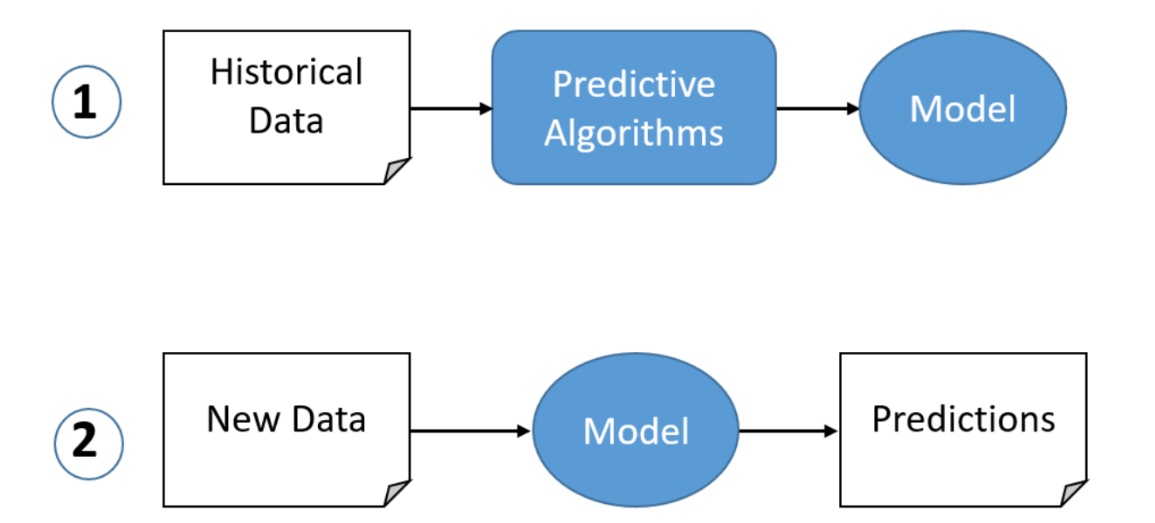
* **Descriptive Analytics**: Summarizes past health records (e.g., a report showing the patient’s past blood sugar levels).

Data Quires, Descriptive statistics, Data Visualization

* **Diagnostic Analytics**: Identifies causes of illness (e.g., analyzing test results to determine why the sugar levels are high).-> Why did it happen?

Data recovery, Data Mining, Correlations

* **Predictive Analytics**: Forecasts future health risks (e.g., predicting the likelihood of diabetes based on lifestyle and genetics).

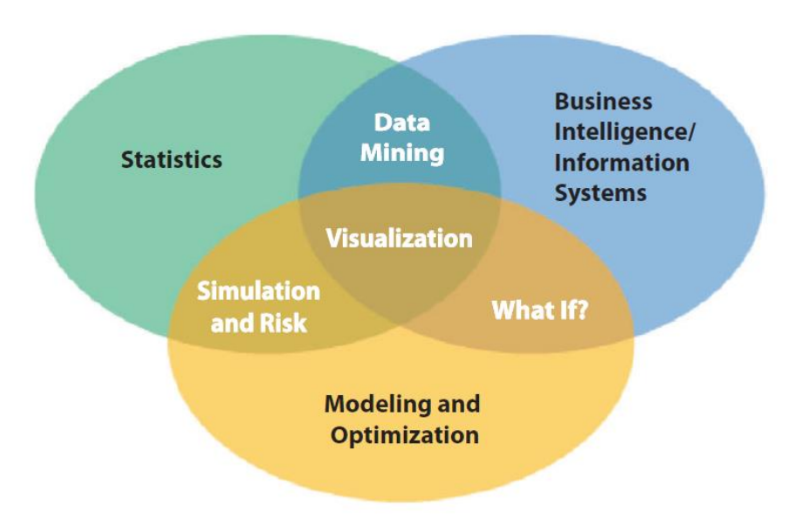


Linear regression, Time-Series Anlayisis,

* **Prescriptive Analytics**: Suggests the best course of action (e.g., recommending a personalized diet and exercise plan to control sugar levels).

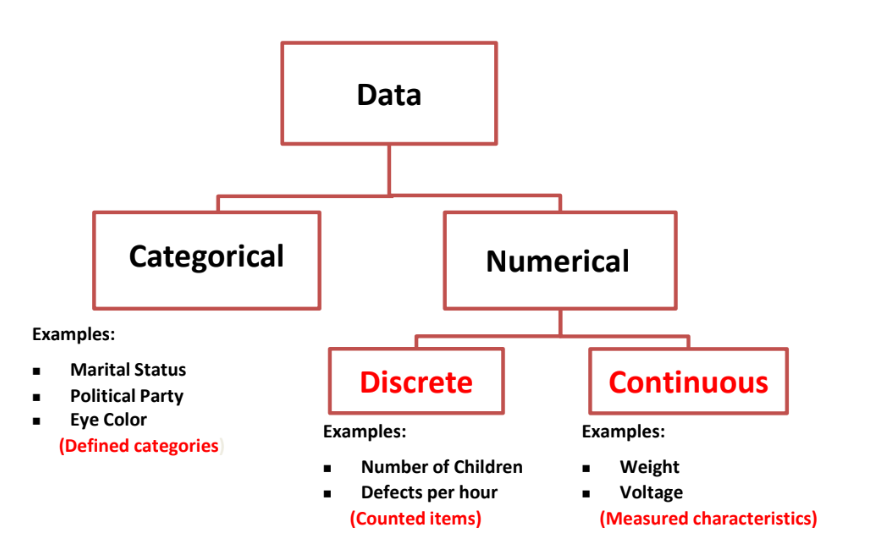
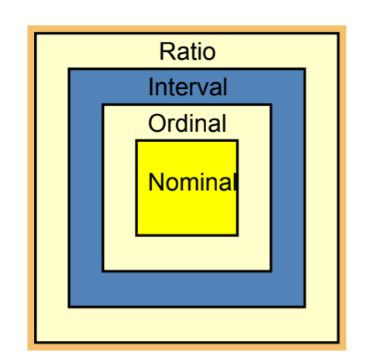
Optimization model, decision analysis

### Elements of Data analytics



The requisite skill set of a data science role is : Mathematics, Technology, Business knowledge.

### Types of Variables



### Levels of Data Measurement

* **Nominal**: Categories without order (e.g., blood type: A, B, AB, O).
* **Ordinal**: Categories with order but no fixed difference (e.g., pain level: mild, moderate, severe).
* **Interval**: Ordered values with equal differences but no true zero (e.g., temperature in Celsius or Fahrenheit).
* **Ratio**: Ordered values with equal differences and a true zero (e.g., heart rate in beats per minute).

### Python Fundamentals-> Week1 Jupyter Notebook.

### Python Fundamentals-> Week1 Jupyter Notebook

## Central tendency and Dispersion and Shape

### Measures of central tendency

A single value describes the characteristics of set of data

* Arithmetic mean
* Weighted mean
* Median
* Percentile

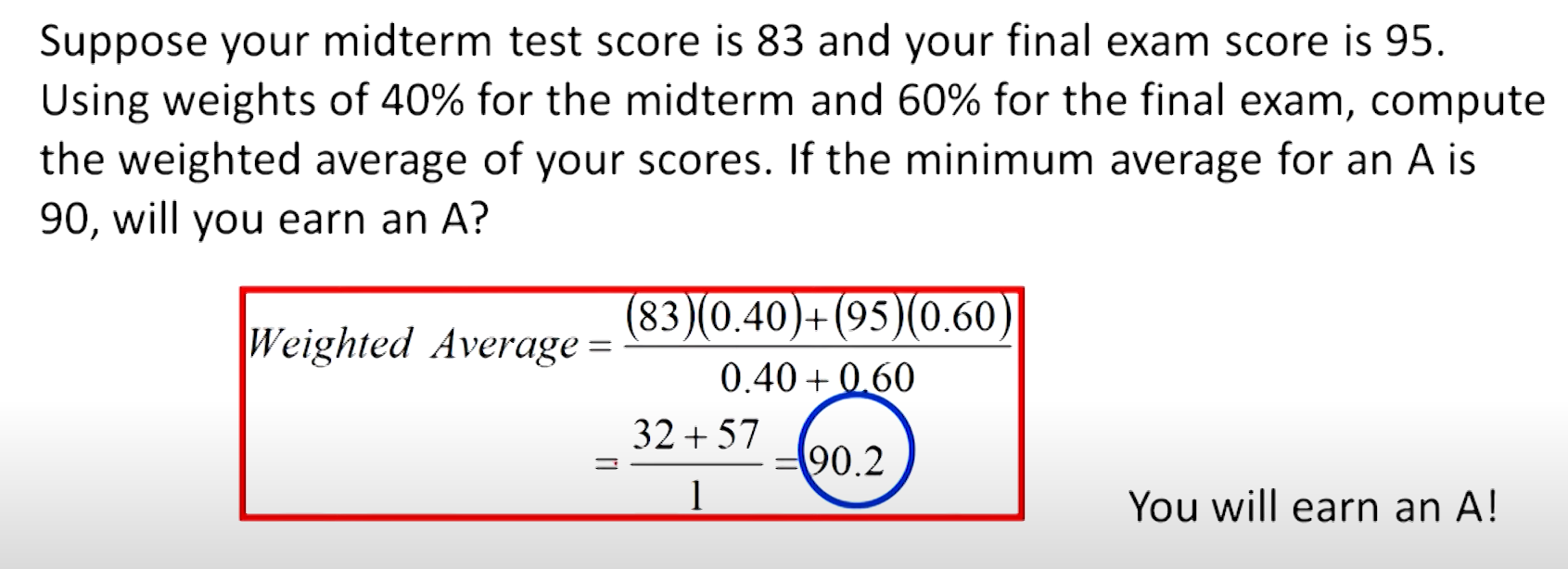
1, **Arithmetic mean**: average of a group of numbers, applicable for interval and ratio data, not applicable for nominal and ordinal data

Calculation of grouped mean formula is summation FM divided by summation f

2, **Weighted Average**: sometimes we want to give more importance to some numbers in the set

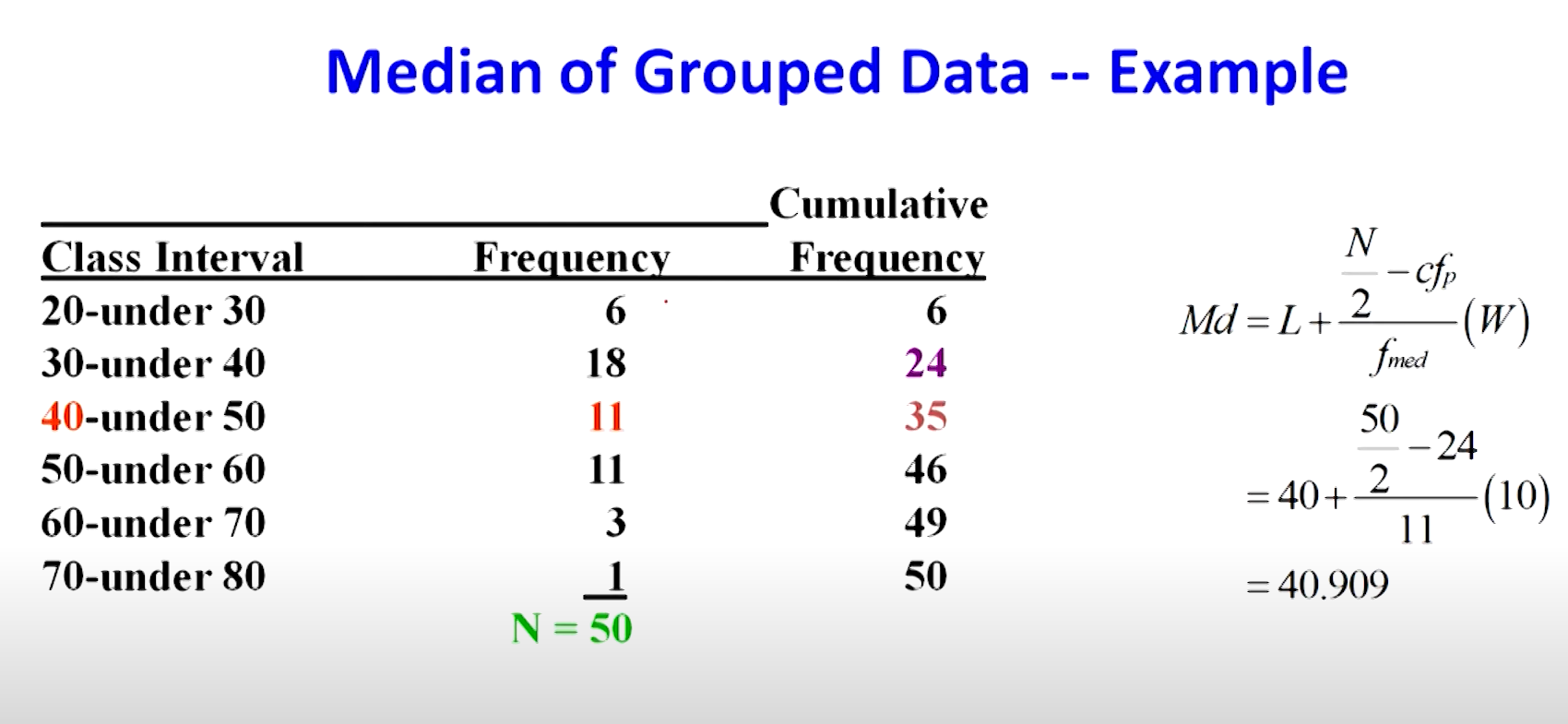


Where x is he data value and w is the associate weight or preference.

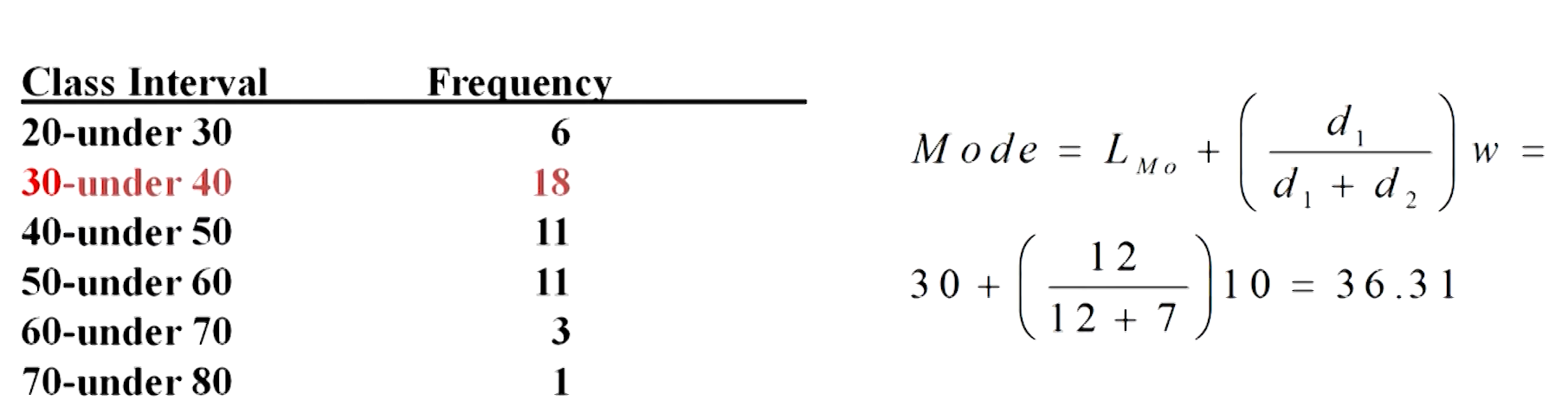


3, **Median**: middle value in a ordered set of numbers, applicable for ordinal, interval and ration data not applicable for nominal data.

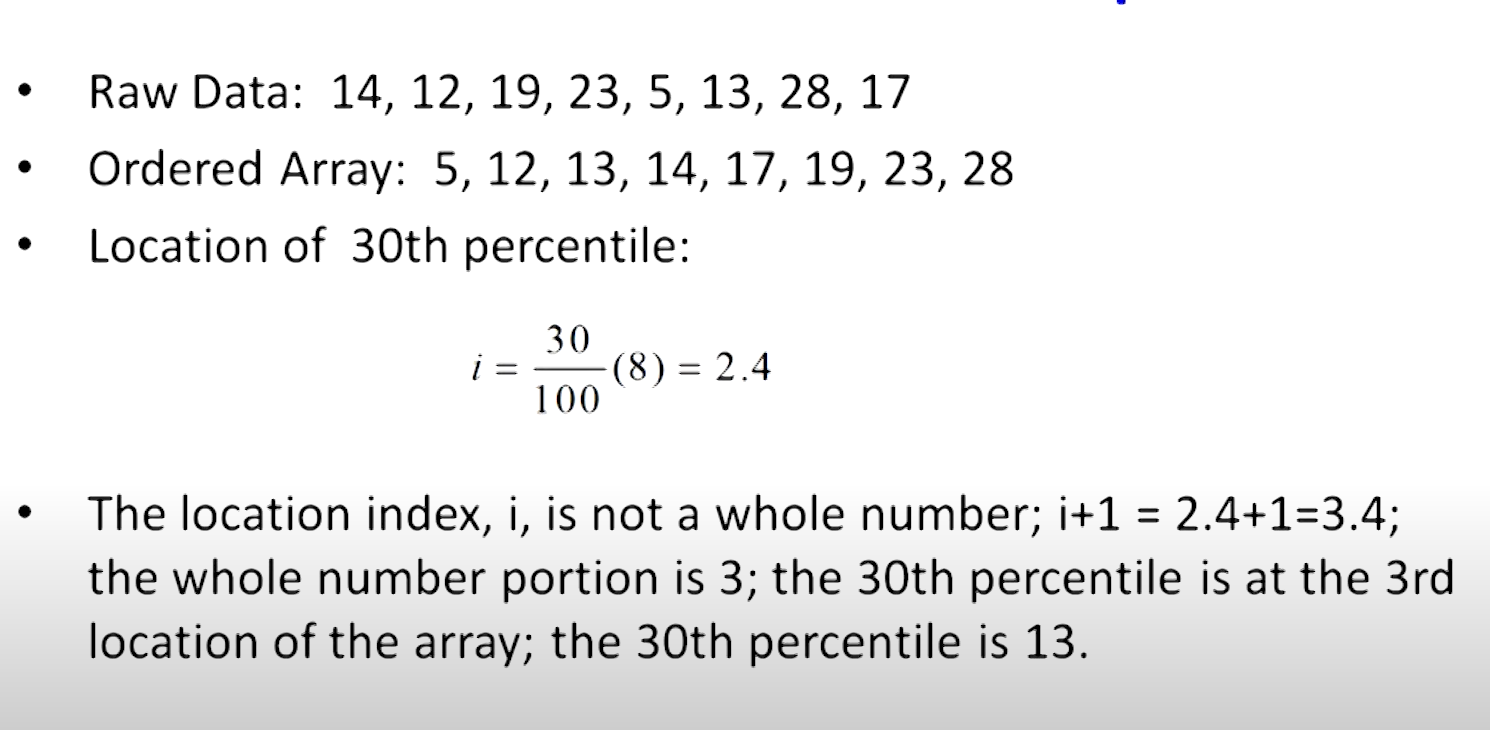
Position of the median is (n+1)/2



4, **Mode:-** The most frequently used value in a dataset is called mode, applicable to all levels of data measurement that is nominal, ordinal, interval and ratio.

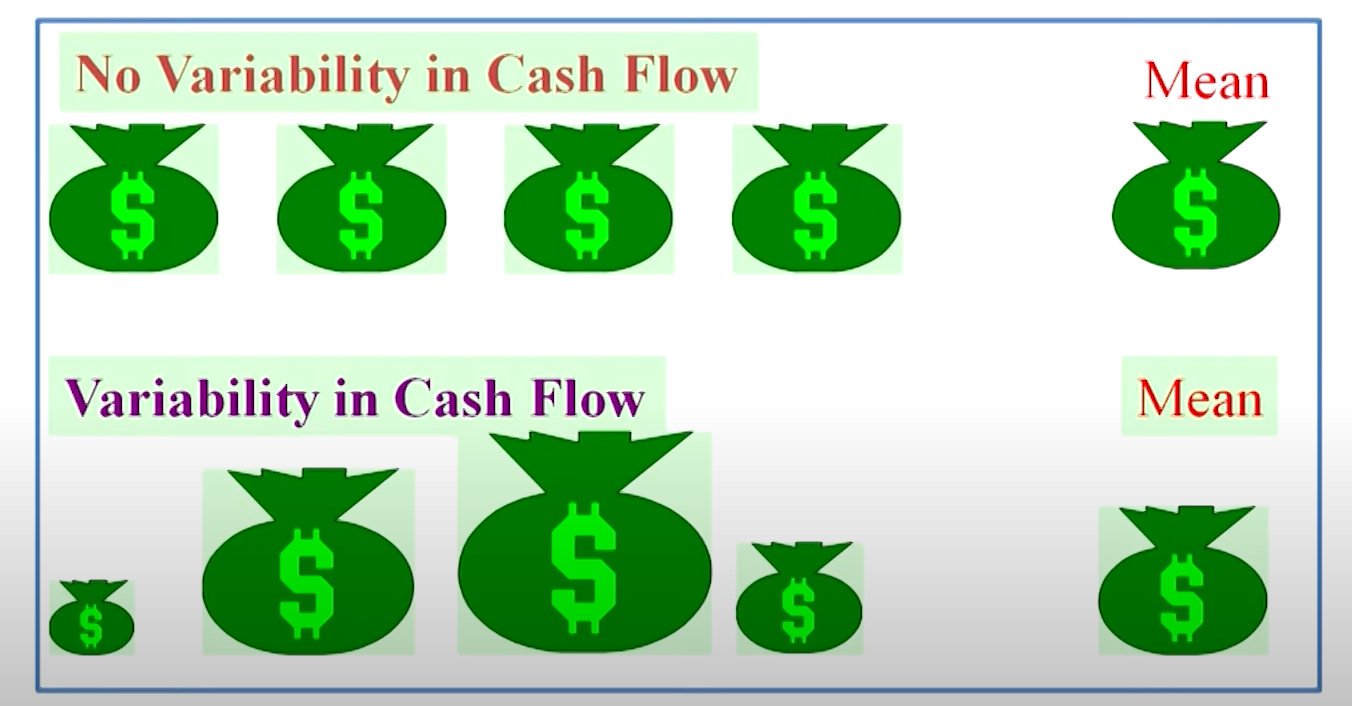


5, **Percentile:-** It divides a group of data into 100 parts, applicable for interval, ratio and ordinal data not applicable for nominal data.



### Measures of Dispersion/Variability

It describes the spread of dispersion of variability.



Common measures of variability

1. **Range**
2. **Inter-quartile range**
3. **Mean Absolute deviation**
4. **Variance**
5. **Standard Deviation**
6. **Z scores**
7. **Coefficient of variation**

1, **Range:** The difference between the largest and the smallest value in the set of data.

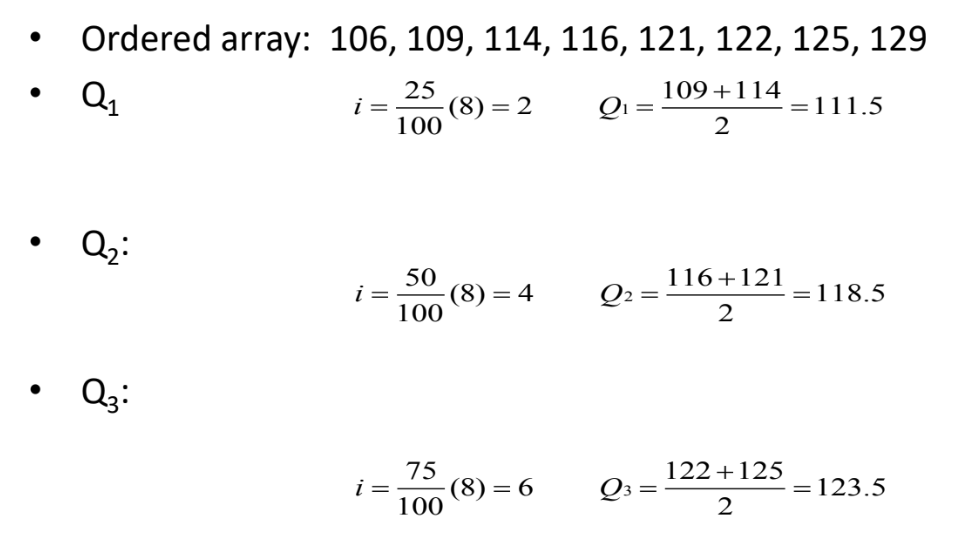
2, **Quartiles:** Measures of central tendency that divide a group of data into four subgroups

Q1-25%

Q2-50%

Q3-75%



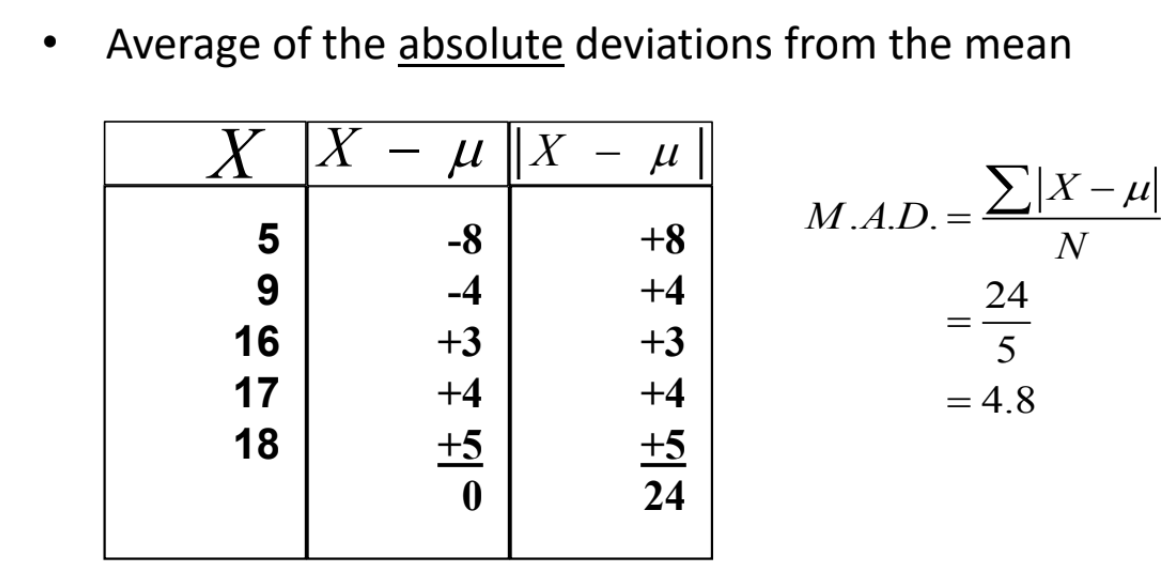


Inter-Quartile Range**:** Range of values between the first and third quartiles

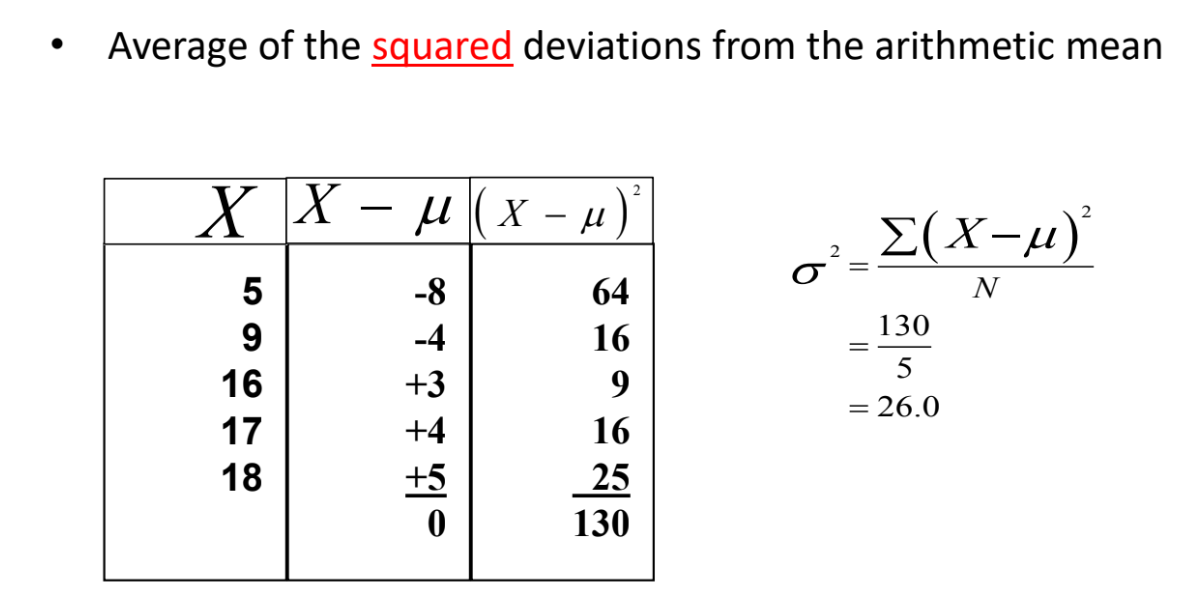
Interquartile Range Q3 – Q1

Deviation from the mean means subtracting the average value in all data values.

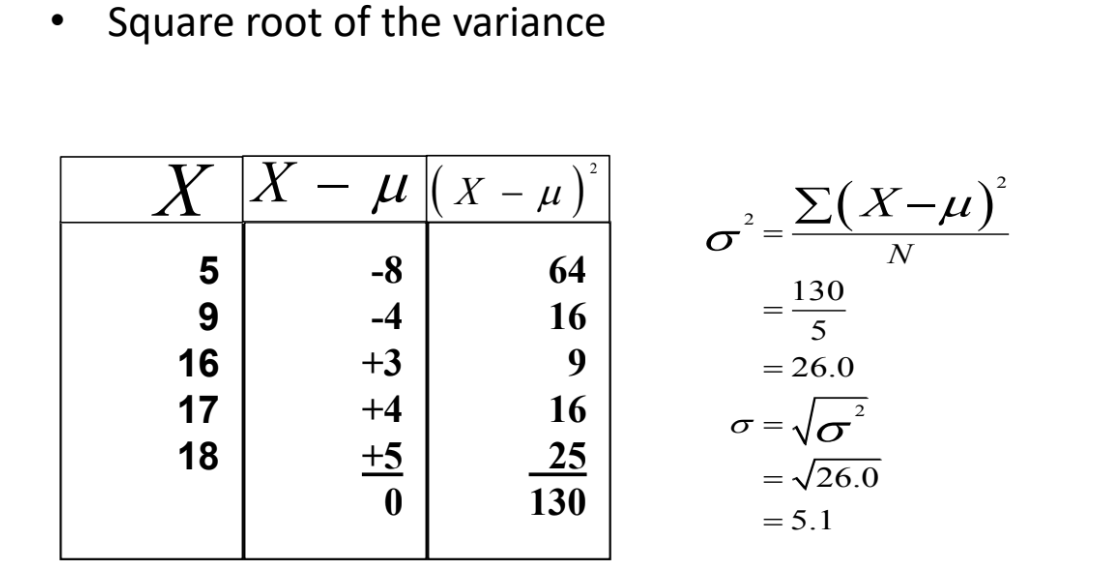
**3, Mean Absolute Deviation:**

****

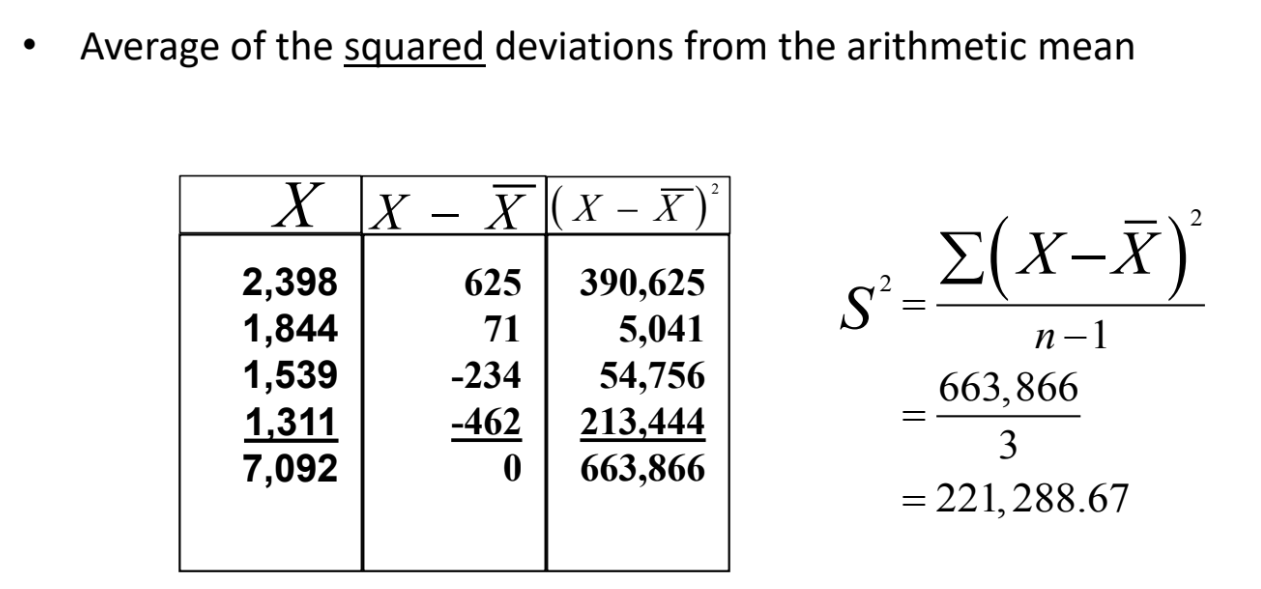
4, **Population variance:**

****

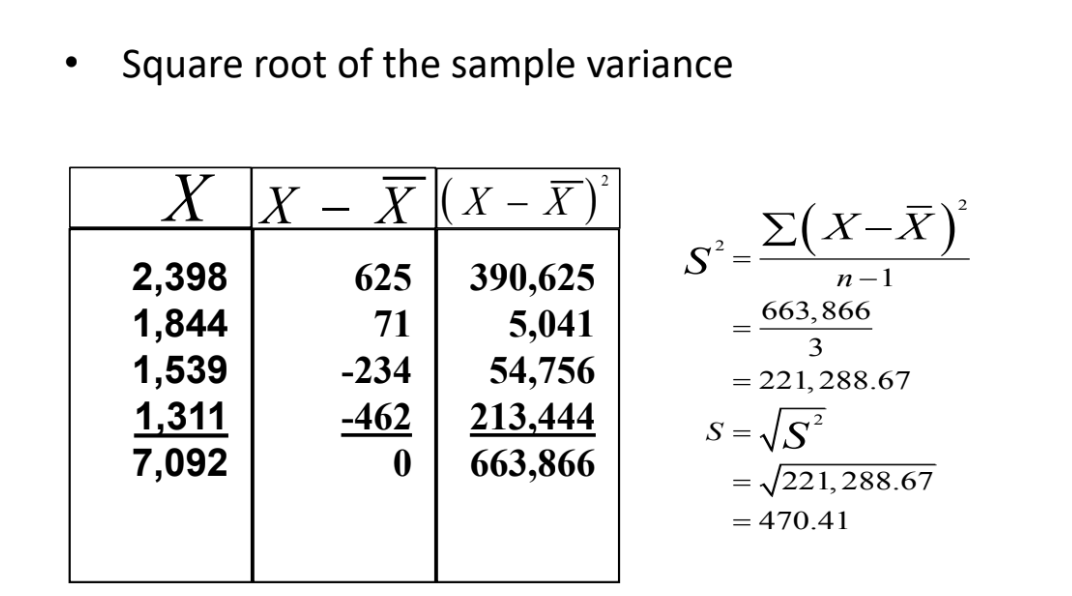
5, **Population Standard Deviation:**



6, **Sample Variance:**

****

**7, Sample Standard deviation:**

****

**Empirical rule:**

The **Empirical Rule**, also known as the **68-95-99.7 Rule**, is a statistical rule that applies to normal (bell-shaped) distributions. It states that:

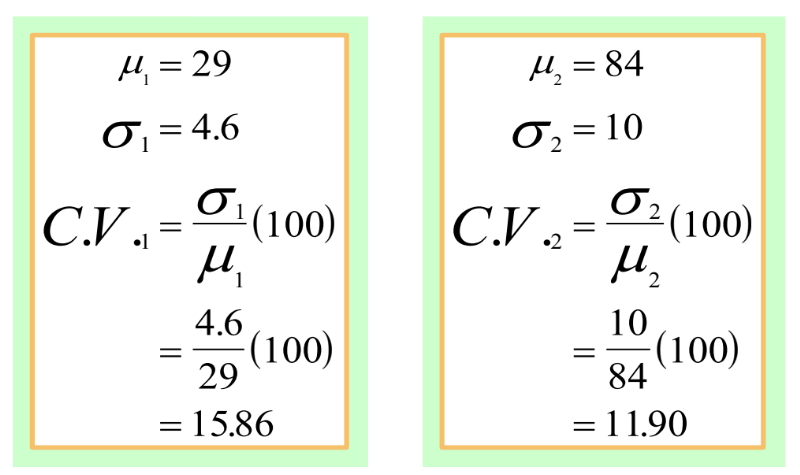
* **68%** of the data falls within **one standard deviation** (σ) of the mean (μ).
* **95%** of the data falls within **two standard deviations** (2σ) of the mean.
* **99.7%** of the data falls within **three standard deviations** (3σ) of the mean.

**Formula Representation**

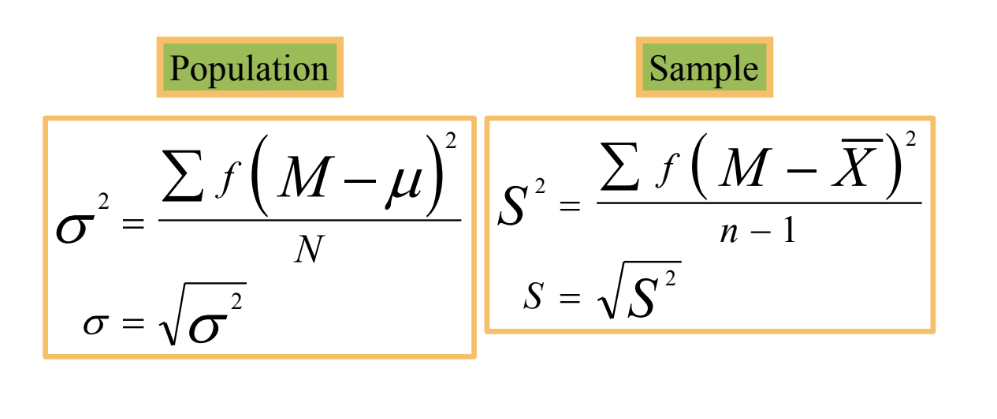
For a dataset that follows a normal distribution:

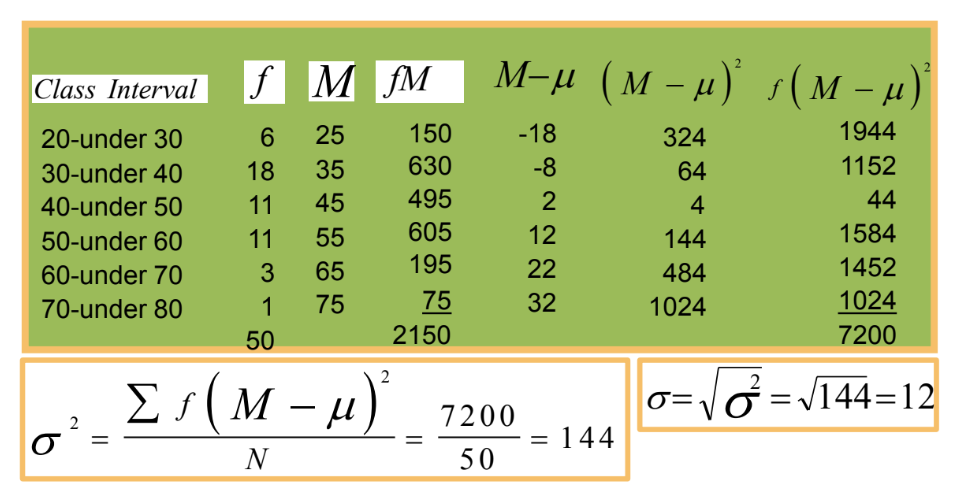
* **μ ± 1σ** → Covers **68%** of the data
* **μ ± 2σ** → Covers **95%** of the data
* **μ ± 3σ** → Covers **99.7%** of the data
* **Chebyshev’s Theorem** (or **Chebyshev’s Inequality**) is a statistical rule that applies to **any** probability distribution, regardless of its shape (not just normal distributions 1/1-k2

7, **Coefficient of variation:** Ratio of the standard deviation to the mean, expressed as a percentage



8, **variation and standard deviation of a grouped data:**





### Measures of Shape

**Skewness:** Absence of symmetry

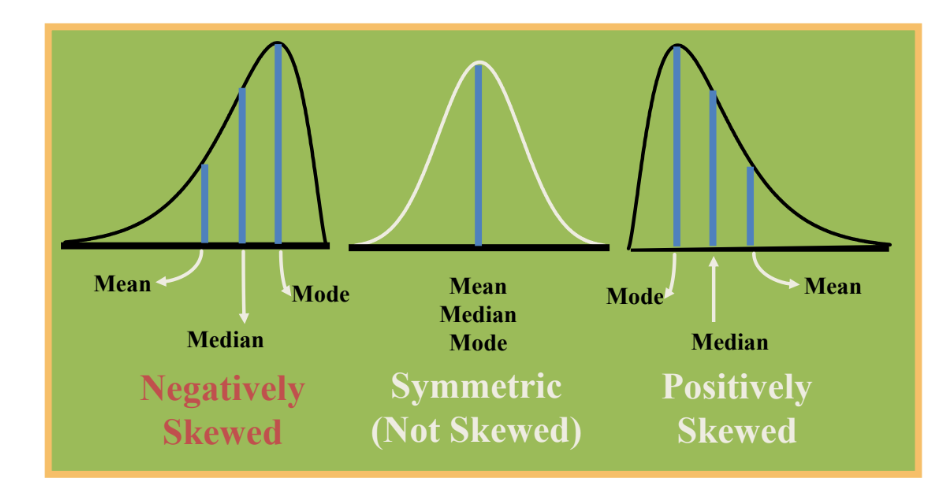
**Kurtosis:** Peakedness of a distribution

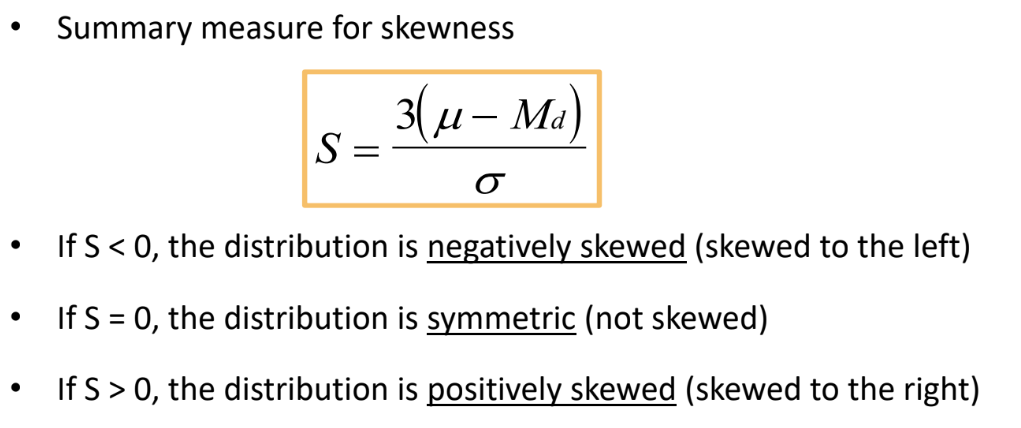
– Leptokurtic: high and thin

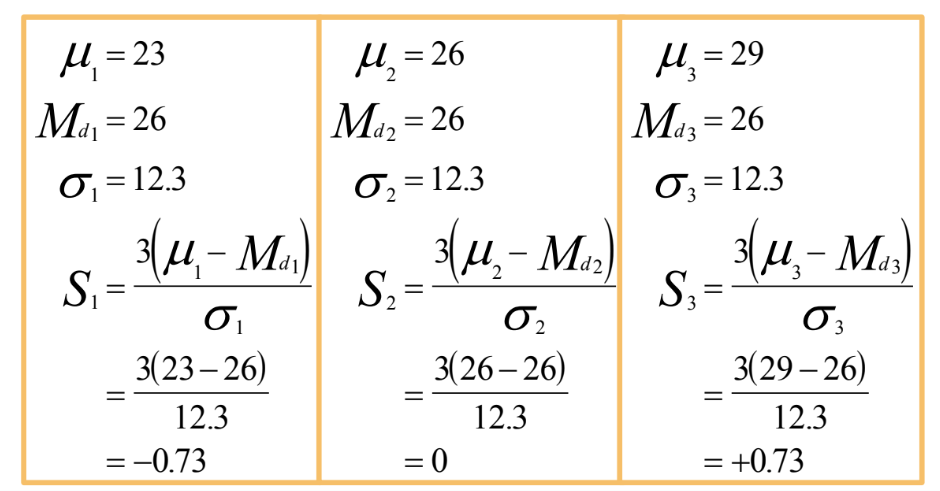
– Mesokurtic: normal shape

– Platykurtic: flat and spread out

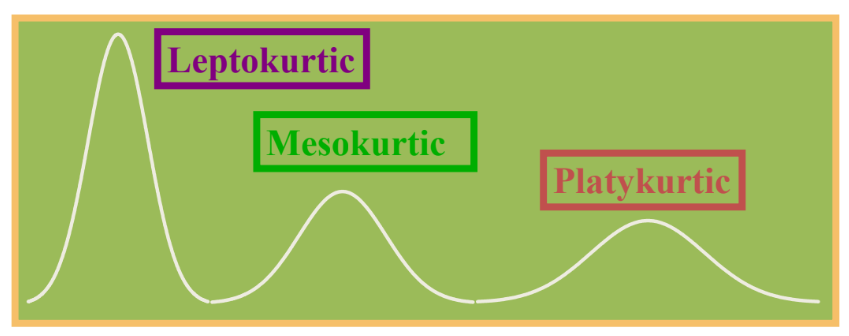
**Box and Whisker Plots:** Graphic display of a distribution Reveals skewness

****

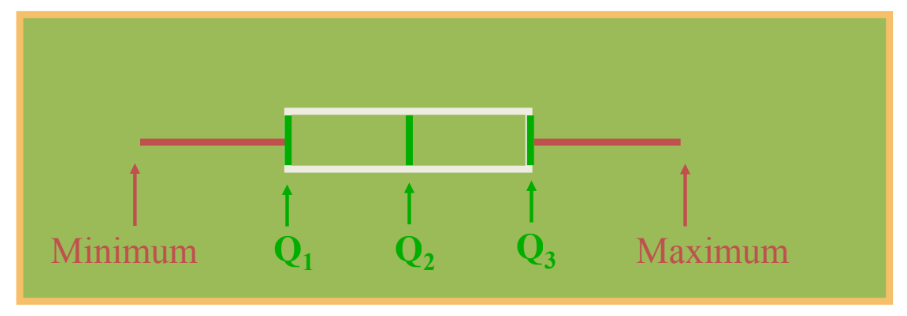
****

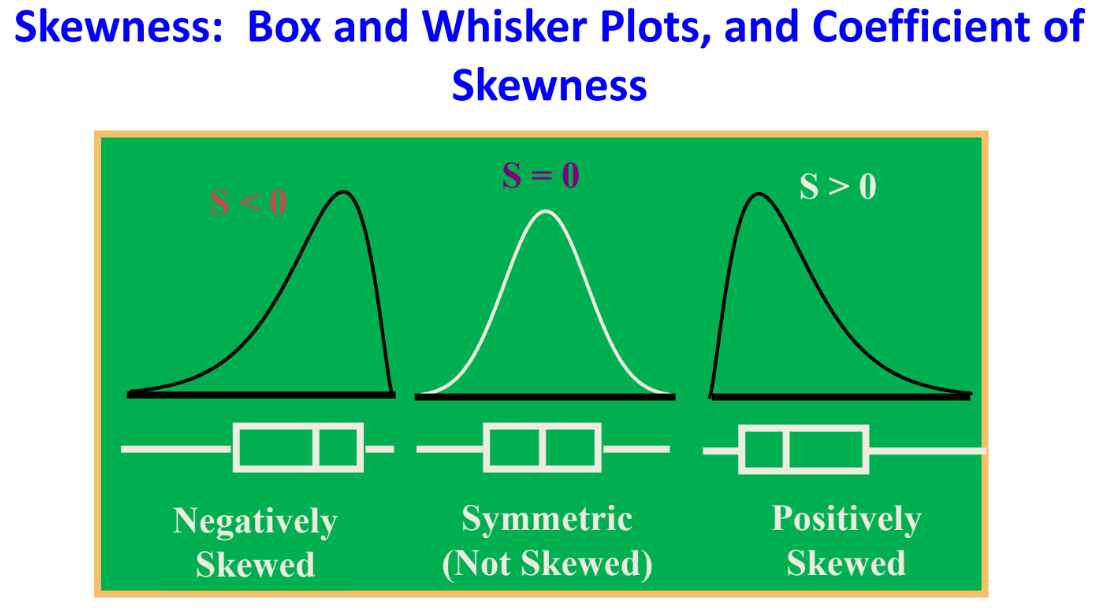
****

**Kurtosis**

****

**Box and Whisker Plots**

****

****