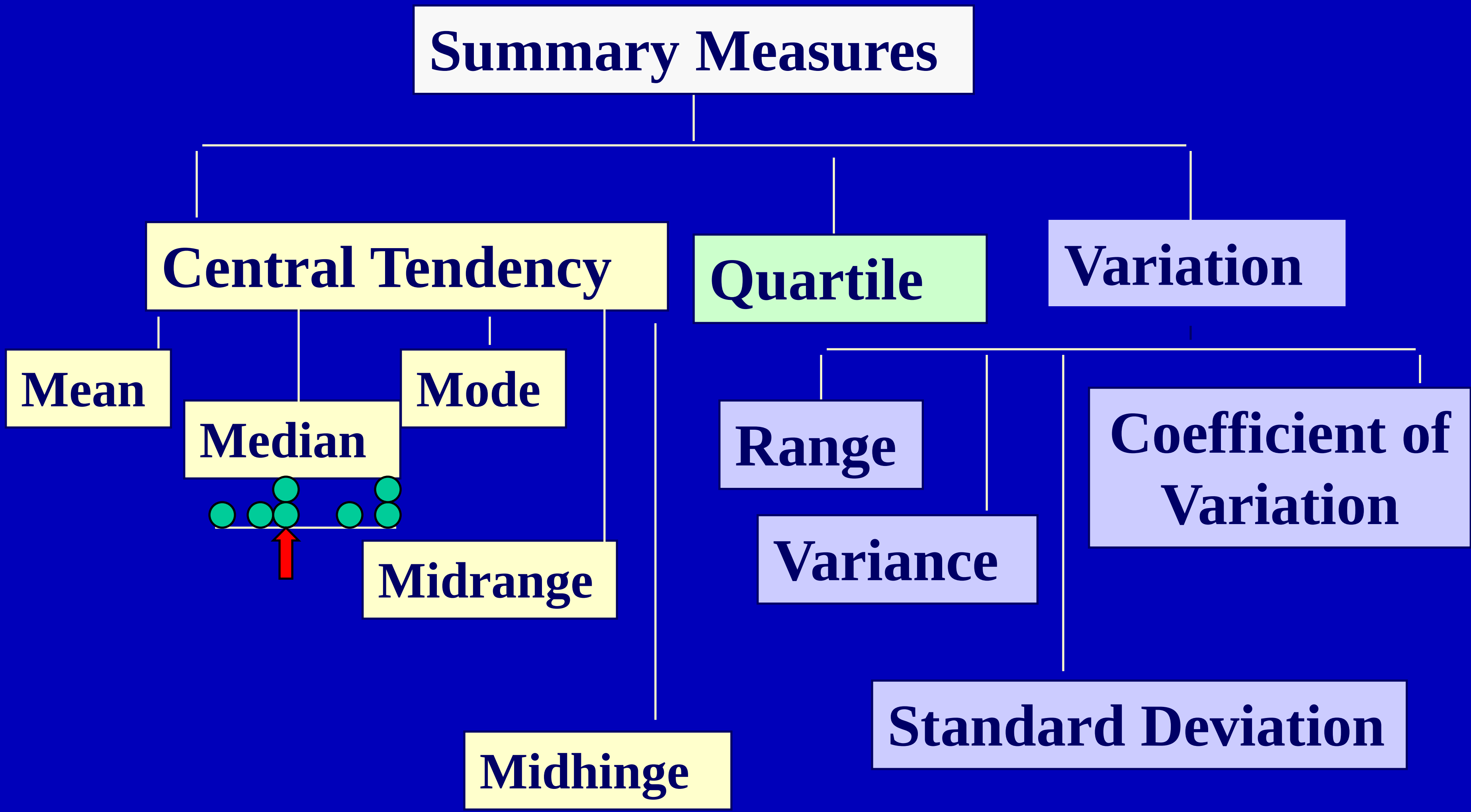
Measures of Central Tendency

Mohammed Zimmo

PhD

Al Azhar University

Summary Measures



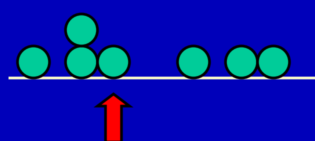
Measures of Central Tendency

Central Tendency

- Mean

$$\sum_{i=1}^{n} x_i \Big/ n$$

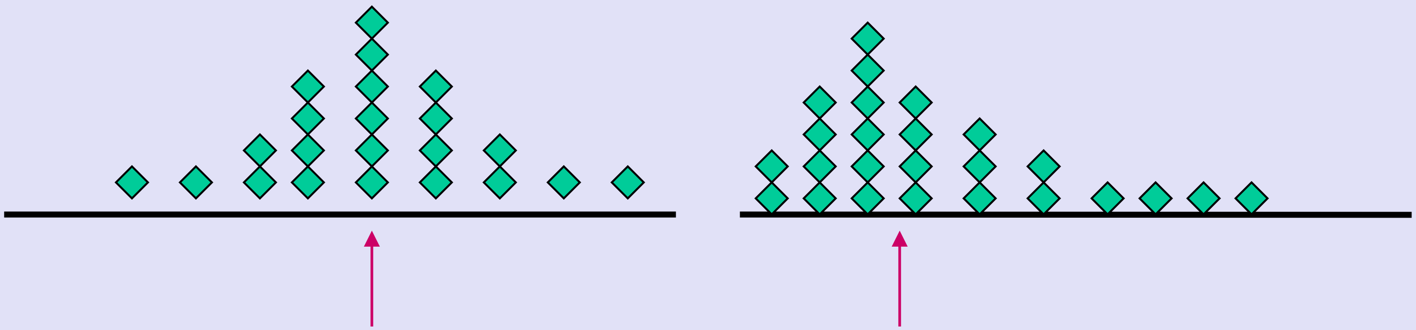
- Median



- Mode
- Midrange
- Midhinge

Measures of Center Tendency

A measure along the horizontal axis of the data distribution that locates the center of the distribution



Some Notations

We are making a series of n (number) observations
The values we observe: as “x-one, x-two, etc”

$$\boxed{x_1, x_2, x_3, \ldots, x_n}$$

Example: Suppose we ask five people how many

hours of they spend on the internet in a week and get

the following numbers: 2, 9, 11, 5, 6

$$\boxed{n = 5, x_1 = 2, x_2 = 9, x_3 = 11, x_4 = 5, x_5 = 6}$$

Mean

The Most Common Measure of Central Tendency

It is the sum of the measurements divided by the total number of measurements

$$\boxed{\bar{x} = \frac{\sum x_i}{n}}$$

$n =$ number of measurements

$\sum x_i =$ sum of all the measurements

Example

Time spend on internet:

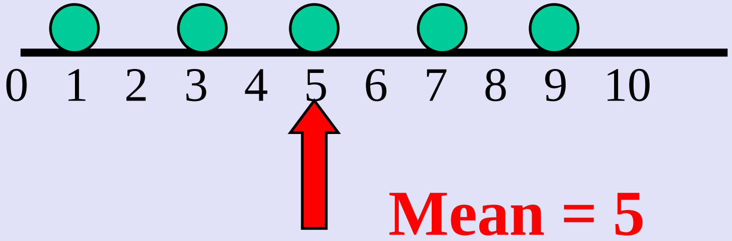
2, 9, 11, 5, 6

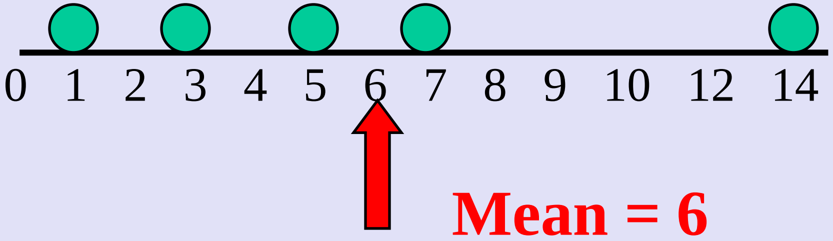
$$\bar{x} = \frac{\sum x_i}{n} = \frac{2+9+11+5+6}{5} = \frac{33}{5} = 6.6$$

If we were able to enumerate the whole population, the **population mean** would be called **μ** (the Greek letter "mu").

Mean

Affected by Extreme Values (Outliers)





Advantages of Mean

- ❑ Crucial for inferential statistics
- ❑ Easy to calculate
- ❑ Well defined mathematical formula
- ❑ Takes into account all the values
- ❑ Widely used
- ❑ Calculated in every case

Median

- The **median** of a set of measurements is the middle measurement when the measurements are ranked from smallest to largest
- The **position of the median** is

$$\textbf{Odd number: } \boldsymbol{Median = \frac{n+1}{2}}$$

$$\textbf{Even number: } \boldsymbol{Median = \frac{n}{2}}\textbf{, and the next one}$$

Example

- The set: 2, 4, 9, 8, 6, 5, 3 $n = 7$
- Sort: 2, 3, 4, 5, 6, 8, 9
- Position:

$$\boldsymbol{Median = \frac{n+1}{2}} = \frac{\mathbf{7+1}}{\mathbf{2}} = 4^{th}$$

Median = 4^{th} largest measurement

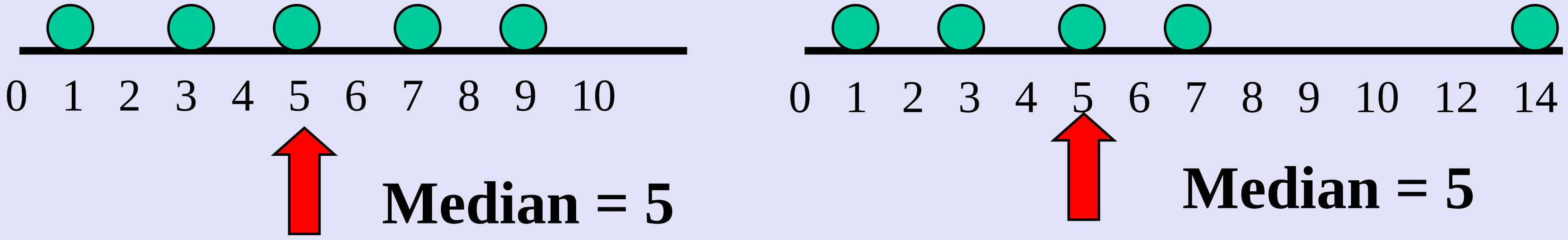
Example

- The set: 2, 4, 9, 8, 6, 5 $n = 6$
- Sort: 2, 4, 5, 6, 8, 9
- Position: $\boldsymbol{Median = \frac{n}{2}} = \frac{6}{2} = 3^{rd}$ and the next one
- $.5(n + 1) = .5(6 + 1) = 3.5^{th}$

Median = (5 + 6)/2 = 5.5 — average of the 3^{rd} and 4^{th} measurements

Median

Not Affected by Extreme Values



Advantages of Median

- Crucial for inferential statistics
- Easy to calculate
- Well defined mathematical formula
- Widely used

Mode

- The **mode** is the measurement which occurs most frequently
- The set: 2, 4, 9, 8, 8, 5, 3
  - The mode is **8**, which occurs twice
- The set: 2, 2, 9, 8, 8, 5, 3
  - There are two modes: **8** and **2** (bimodal)
- The set: 2, 4, 9, 8, 5, 3
  - There is no mode (each value is unique).

Example

The number of quarts of milk purchased by 25 households:

0 0 1 1 1 1 1 2 2 2 2 2 2
2 2 2 3 3 3 3 3 4 4 4 5

Example

The number of quarts of milk purchased by 25 households:

0 0 1 1 1 1 1 2 2 2 2 2 2
2 2 2 3 3 3 3 3 4 4 4 5

- **Mean?**

$$\bar{x} = \frac{\sum x_i}{n} = \frac{55}{25} = 2.2$$

Example

The number of quarts of milk purchased by 25 households:

0 0 1 1 1 1 1 2 2 2 2 2 2
2 2 2 3 3 3 3 3 4 4 4 5

- Mean?

$$\boxed{\bar{x} = \frac{\sum x_i}{n} = \frac{55}{25} = 2.2}$$

- **Median?**

$$\boxed{m = 2}$$

Example

The number of quarts of milk purchased by 25 households:

0 0 1 1 1 1 1 2 2 2 2 2 2
2 2 2 3 3 3 3 3 4 4 4 5

- **Mean?**

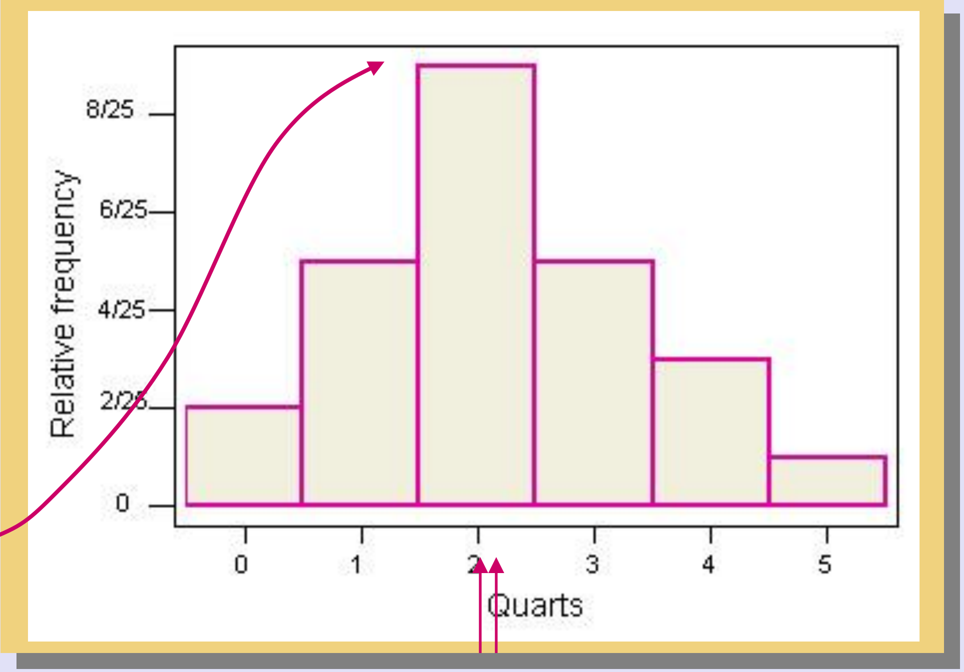
$$\bar{x} = \frac{\sum x_i}{n} = \frac{55}{25} = 2.2$$

- **Median?**

$$m = 2$$

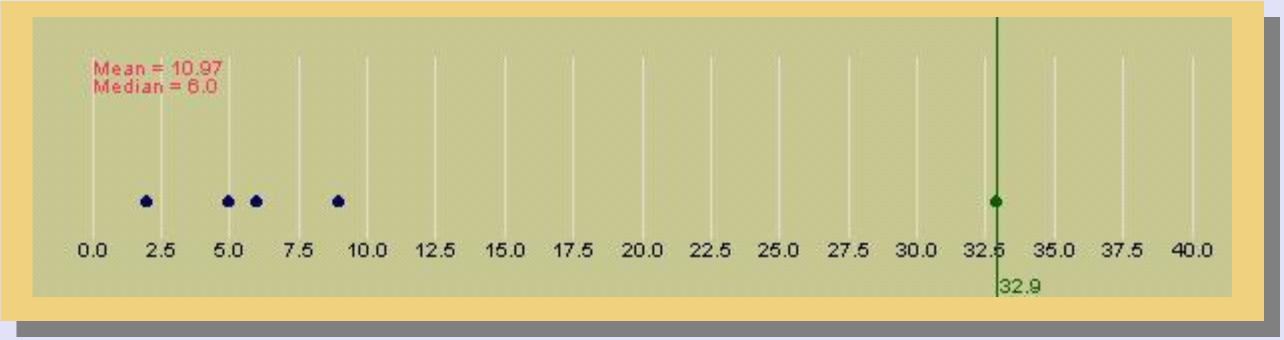
- **Mode? (Highest peak)**

$$\text{mode} = 2$$



Extreme Values

- The **mean** is more easily affected by extremely large or small values than the median



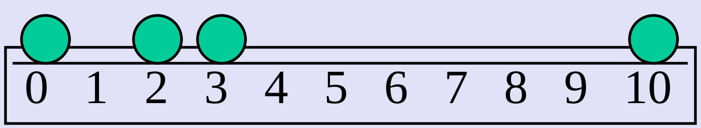
Midrange

- **A Measure of Central Tendency**
- **Average of Smallest and Largest**

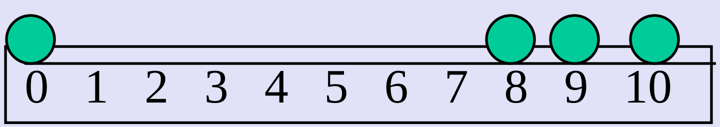
Observation:

$$\text{Midrange} = \frac{x_{largest} + x_{smallest}}{2}$$

- **Affected by Extreme Value**

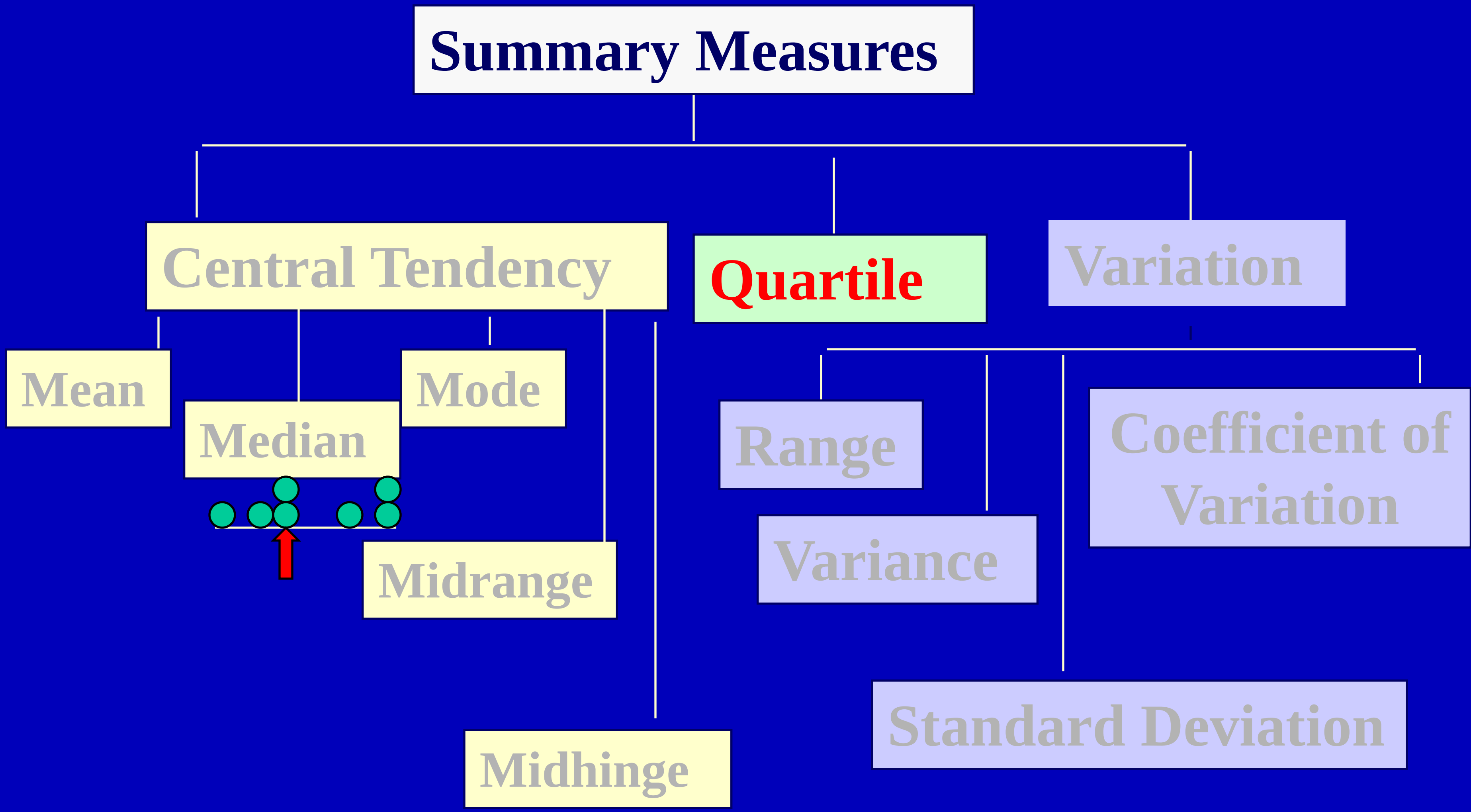


Midrange = 5



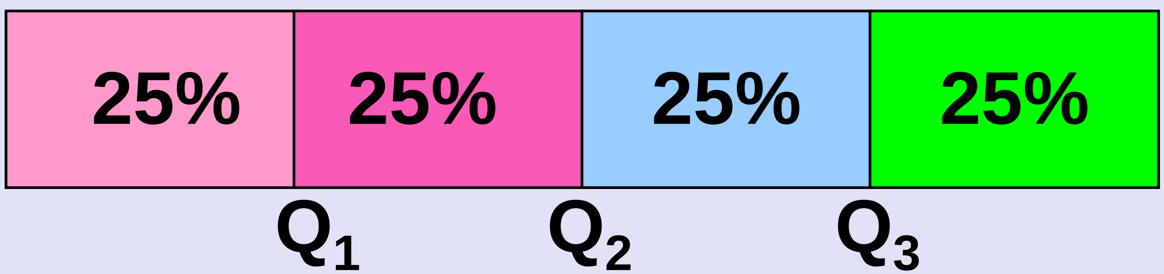
Midrange = 5

Summary Measures



Quartiles

- **Not a Measure of Central Tendency**
- **Split Ordered Data into 4 Quarters**



Quartiles

- **Not a Measure of Central Tendency**
- **Split Ordered Data into 4 Quarters**

25%	25%	25%	25%
	Q_1	Q_2	Q_3

- **Position of i-th Quartile:** **position of point** $Q_i = \frac{i(n+1)}{4}$

Data in Ordered Array: 11 12 13 16 16 17 18 21 22

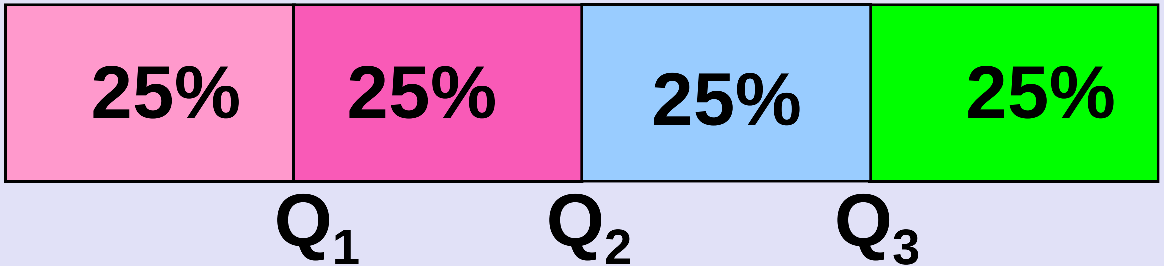
Position of Q_1 $= \frac{1 \cdot (9+1)}{4} = 2.50$ **Q_1 =12.5**

Midhinge

- **A Measure of Central Tendency**
- **The Middle point of 1st and 3rd Quarters**

$$\textbf{Midhinge} = \frac{Q_1 + Q_3}{2}$$

- **Not Affected by Extreme Values**



Midhinge

- **A Measure of Central Tendency**
- **The Middle point of 1st and 3rd Quarters**

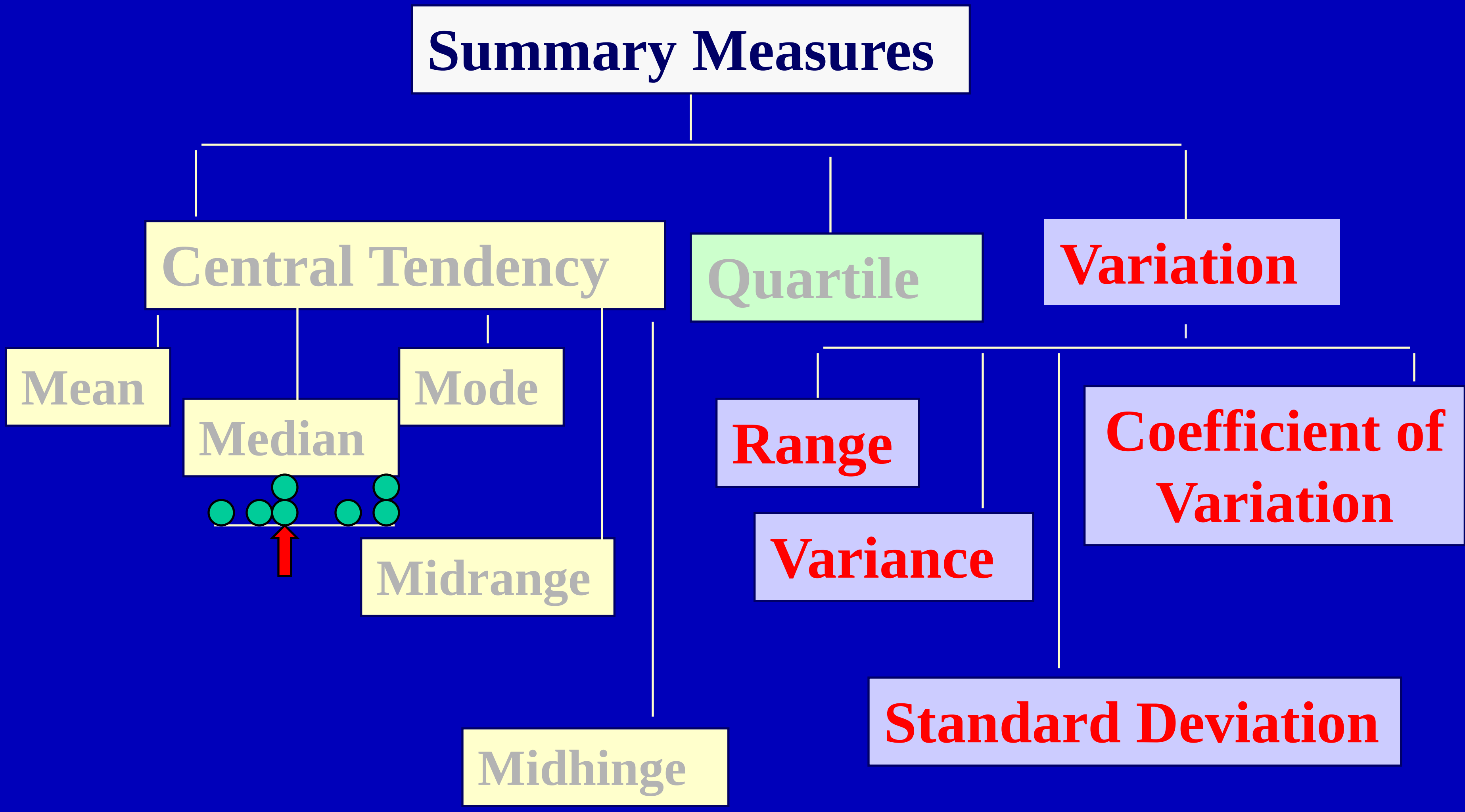
Midhinge = $$\frac{Q_1 + Q_3}{2}$$

- **Not Affected by Extreme Values**

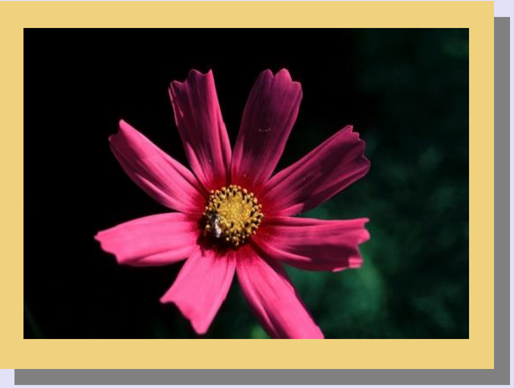
Data in Ordered Array: 11 12 13 16 16 17 18 21 22

Midhinge = $$\frac{Q_1 + Q_3}{2} = \frac{12.5 + 19.5}{2} = 16$$

Summary Measures



The Range



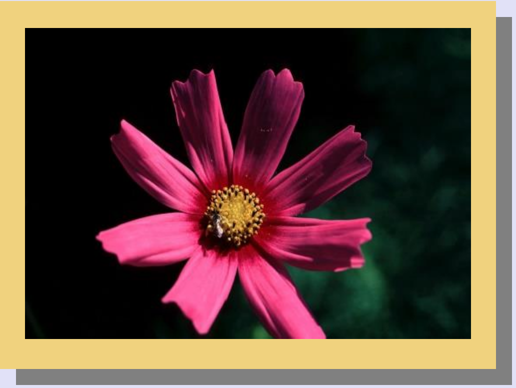
- The **range (R)**
  - Is the difference between the largest and smallest measurements
- **Example:** A botanist records the number of petals on 5 flowers:

5, 12, 6, 8, 14

- The range is **R = 14 – 5 = 9.**

Quick and easy, but only uses 2 of the 5 measurements.

Interquartile Range



- Measure of Variation
- Also Known as **Midspread**

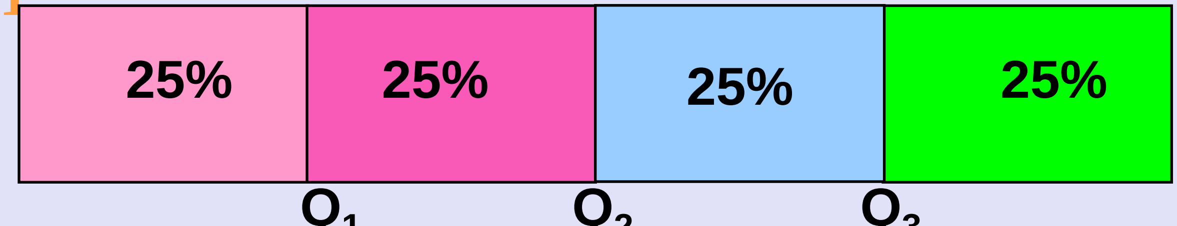
Spread in the Middle 50%

- Difference Between Third & First

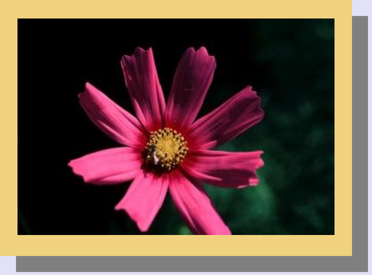
Quartiles: Interquartile Range = $Q_3 - Q_1$

Data in Ordered Array: 11 12 13 16 16 17 17 18 21

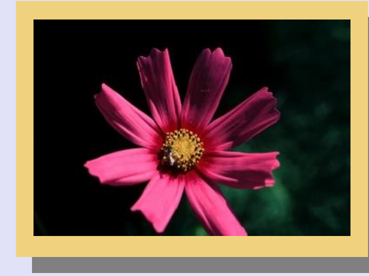
$Q_3 - Q_1$ = 17.5 - 12.5 = 5



The Variance



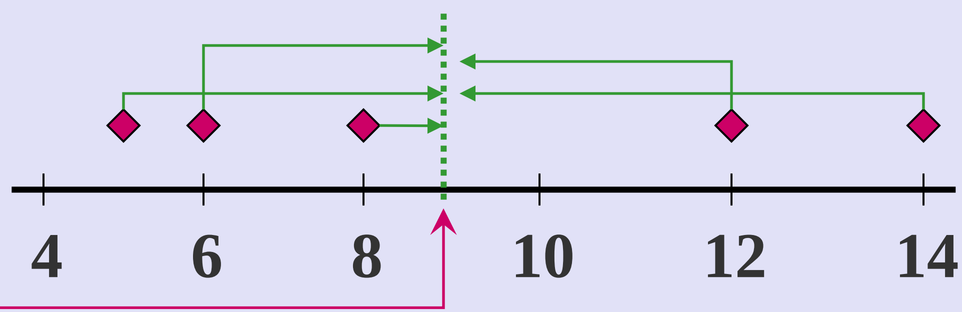
- The **variance**
  - Is measure of variability that uses all the measurements
  - It measures the average deviation of the measurements about their mean



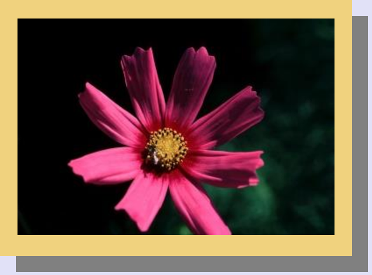
The Variance

- The **variance**
  - Is measure of variability that uses all the measurements
  - It measures the average deviation of the measurements about their mean.
- **Flower petals:** 5, 12, 6, 8, 14

$$\bar{x} = \frac{45}{5} = 9$$



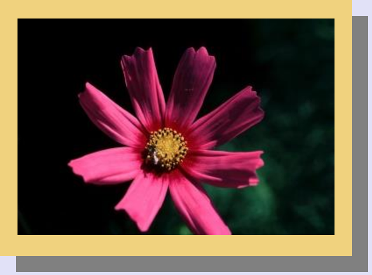
The Variance



- The **variance of a population**
  - Is the average of the squared deviations of the measurements about their mean μ.

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

The Variance



- The **variance of a sample**
  - Is the sum of the squared deviations of the measurements about their mean, divided by $(n-1)$.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

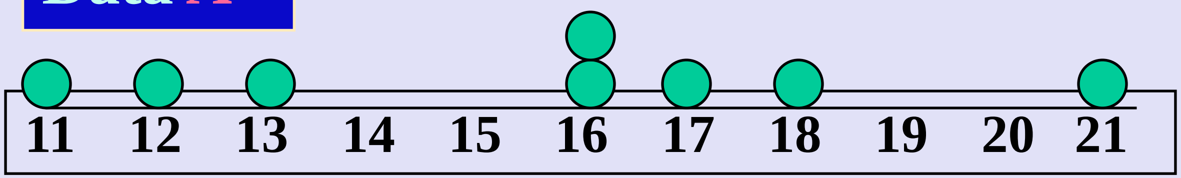
Standard Deviation (SD)

Example: Calculate SD for this observation set: **(7,3,4,6)**

Value (X_i)	**Deviation From mean (X_i - X)**	**$(Dev.)^2$ $(X_i - X)^2$**
7	2	4
3	-2	4
4	-1	1
6	1	1
Total = 20		10
Mean (X) = 5		Mean of $(Dev.)^2$ = 2.5 SD = $\sqrt{2.5}$=1.6

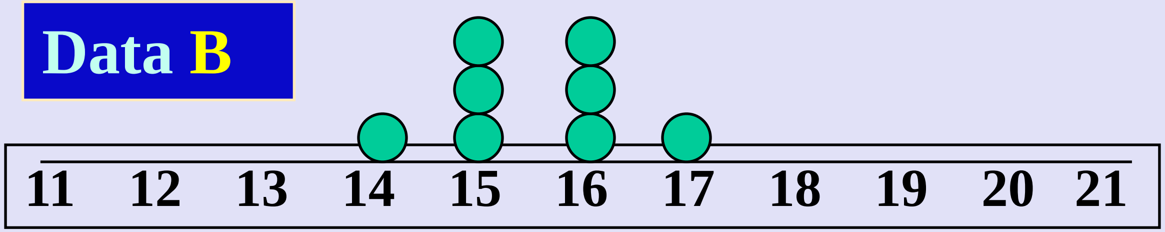
Comparing Standard Deviations

Data A



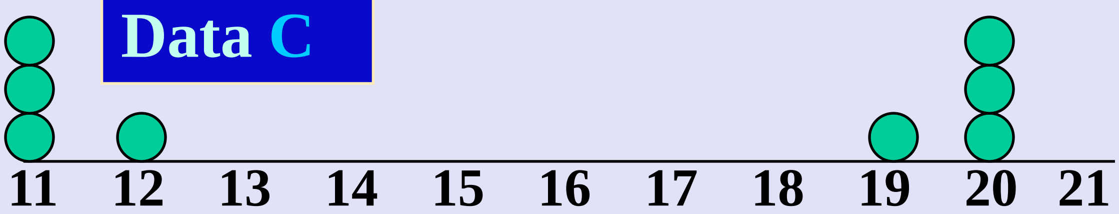
Mean = 15.5
SD = 3.338

Data B



Mean = 15.5
SD = .9258

Data C



Mean = 15.5
SD = 4.57

Standard Error (SE)

It is a measure of the extent to which the sample means deviate from the true population mean

The smaller it is the closer to the true population mean

When sample size is larger the SE is smaller
SE is calculated from the formula:

$$\mathbf{SE} = \frac{\mathbf{SD}}{\sqrt{\mathbf{n}}}$$

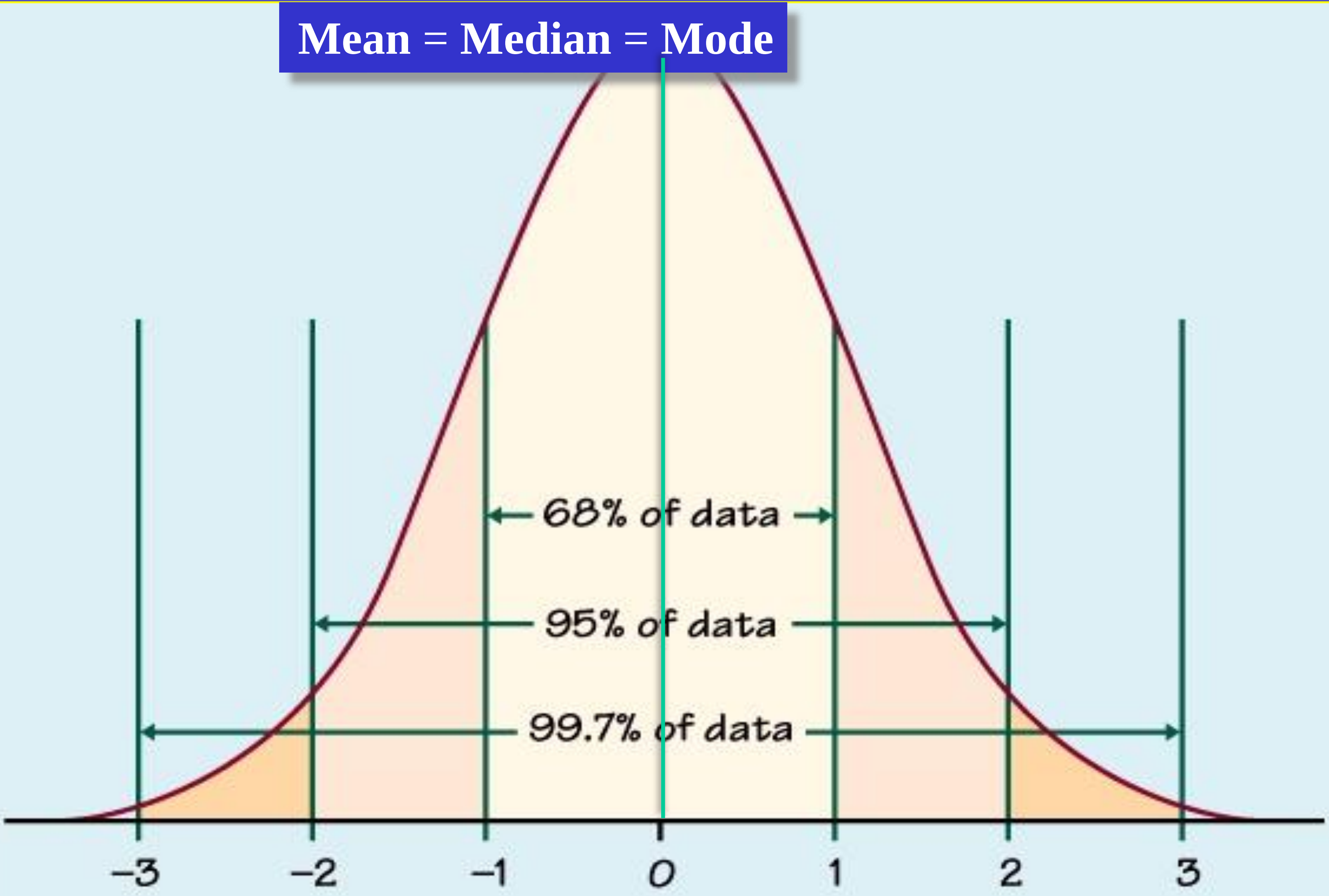
Confidence interval

What is the meaning of 95% confidence interval?

This represents an interval of parameter values consistent with the data

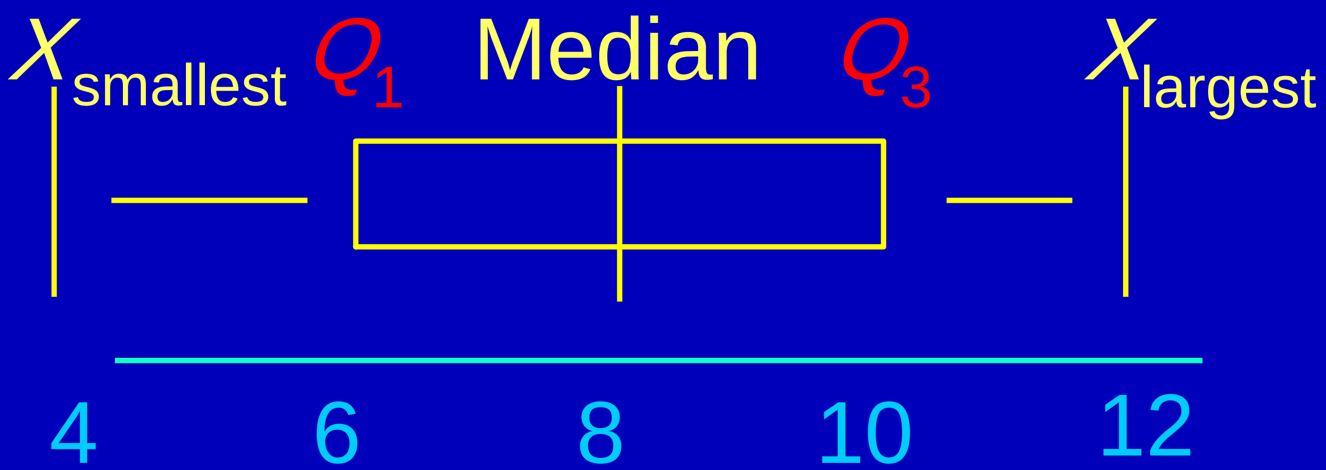
It means that should you repeat an experiment or survey over and over again, 95% percent of the time your results will match the results you get from a population

Normal Distributions



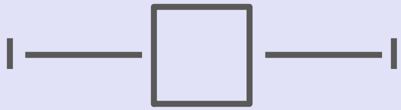
Box-and-Whisker Plot

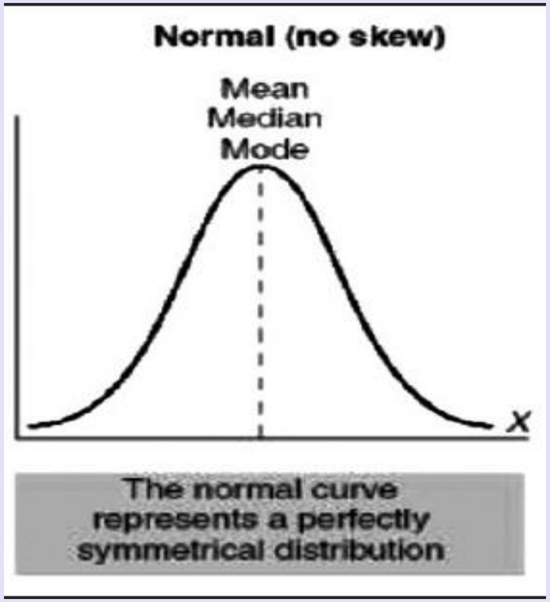
Graphical Display of Data Using 5-Number Summary



Distribution Shape & Box-and-Whisker Plots

Symmetric: Mean = Median

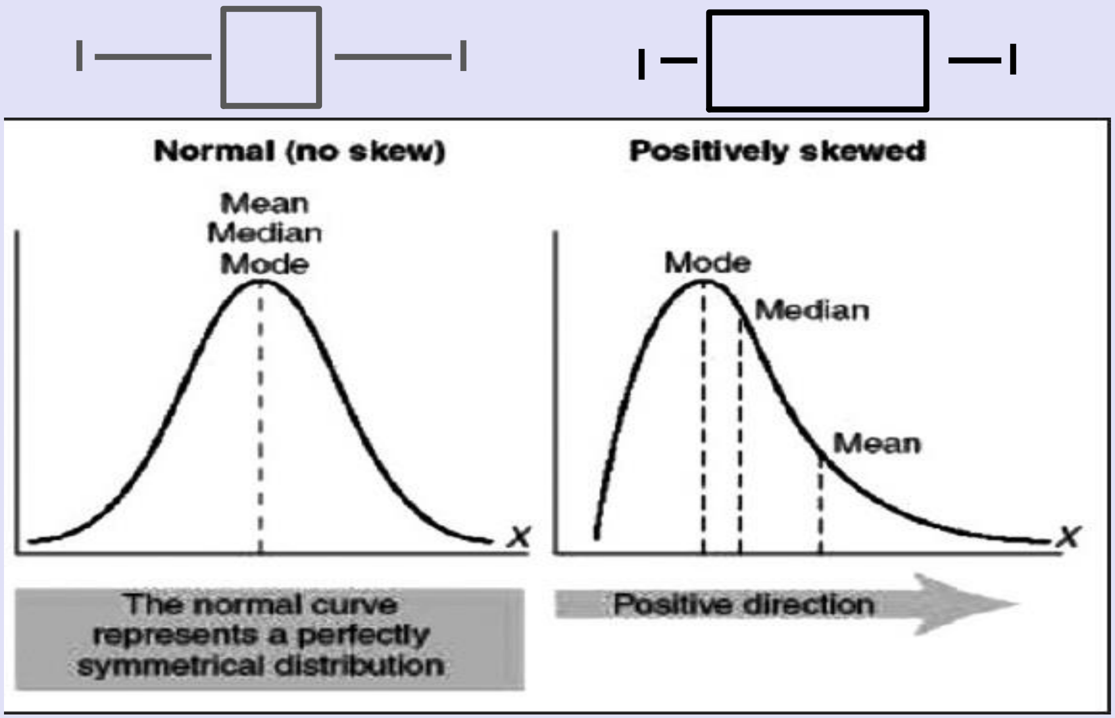




Distribution Shape & Box-and-Whisker Plots

Symmetric: Mean = Median

Skewed right: Mean > Median

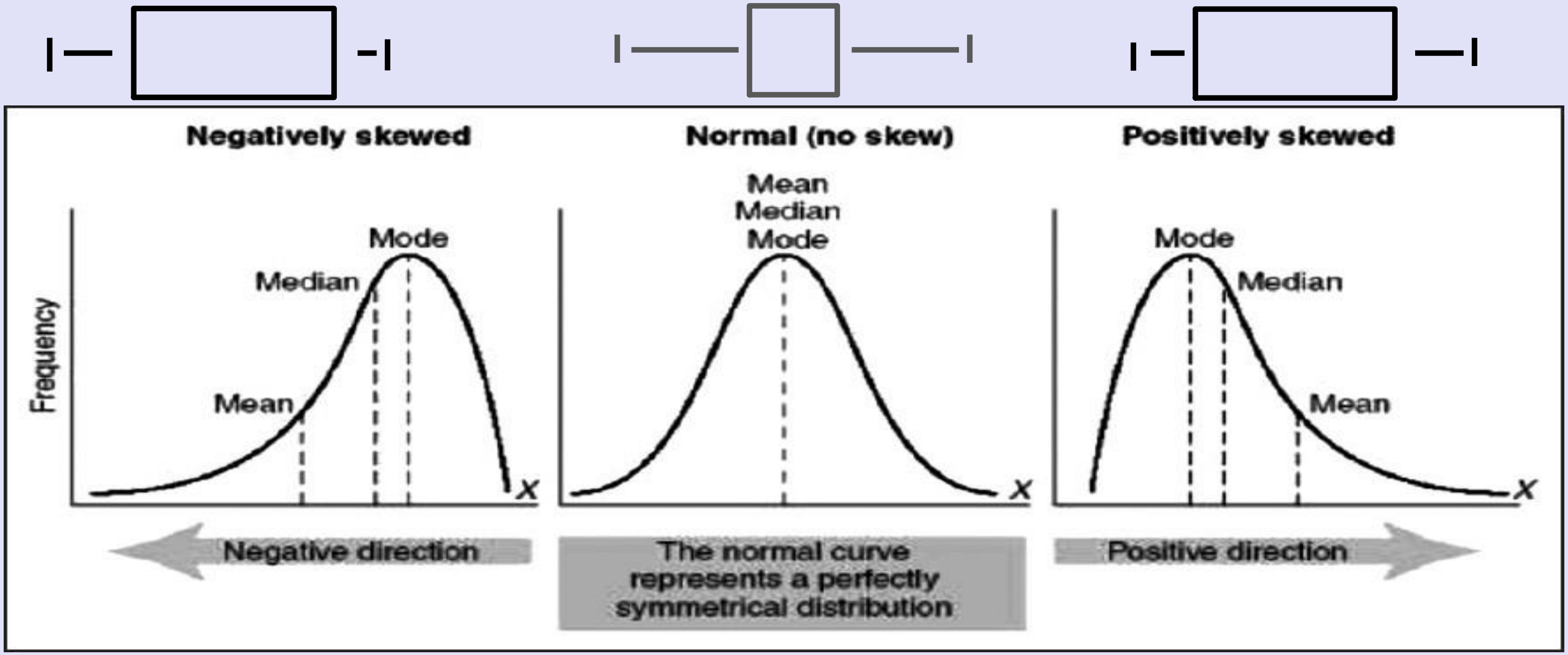


Distribution Shape & Box-and-Whisker Plots

Symmetric: Mean = Median

Skewed right: Mean > Median

Skewed left: Mean < Median




Thank you