Statistics for Computing

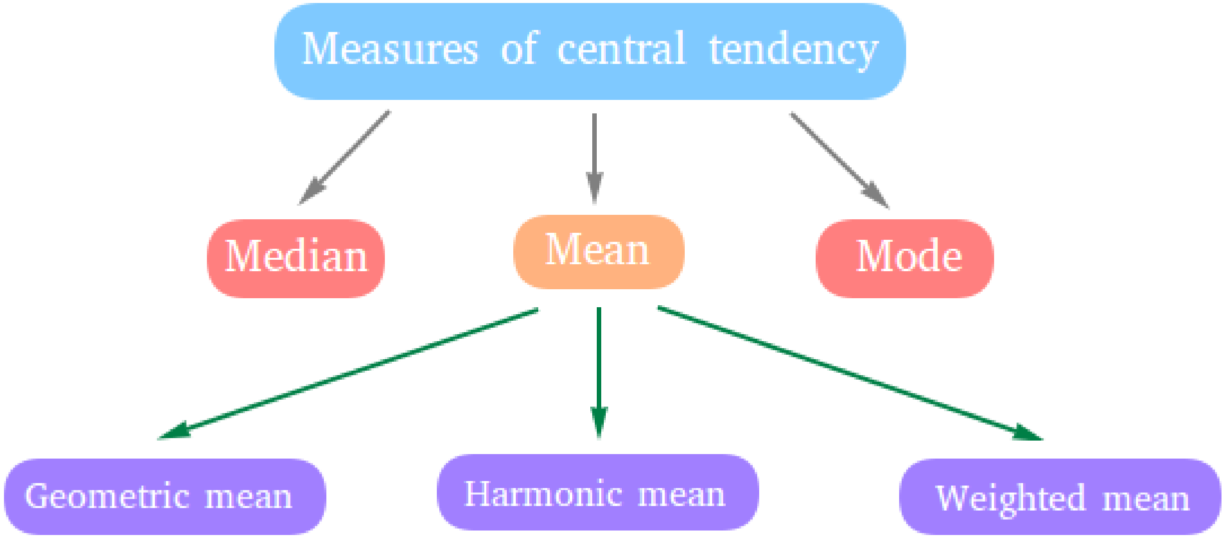
(CSC 502 0.0)

MSc in Computer Science

Prof. (Dr.) R.M. KAPILA RATHNAYAKA

B.Sc. Special (Math. & Stat.) (Ruhuna), M.Sc. (Industrial Mathematics) (USJ),
M.Sc. (Stat.) (WHUT, China),
Ph.D. (Applied Statistics, WHUT)

MEASURES OF CENTRAL TENDENCY



Introduction : Measures of Central Tendency

- A measure of central tendency is a **single value** that attempts to describe a set of data by identifying the central position within that set of data.

- As such, measures of central tendency are sometimes called measures of central location.

- They are also classed as summary statistics. In addition to central tendency, the variability and distribution of your data set is important to understand when performing descriptive statistics.

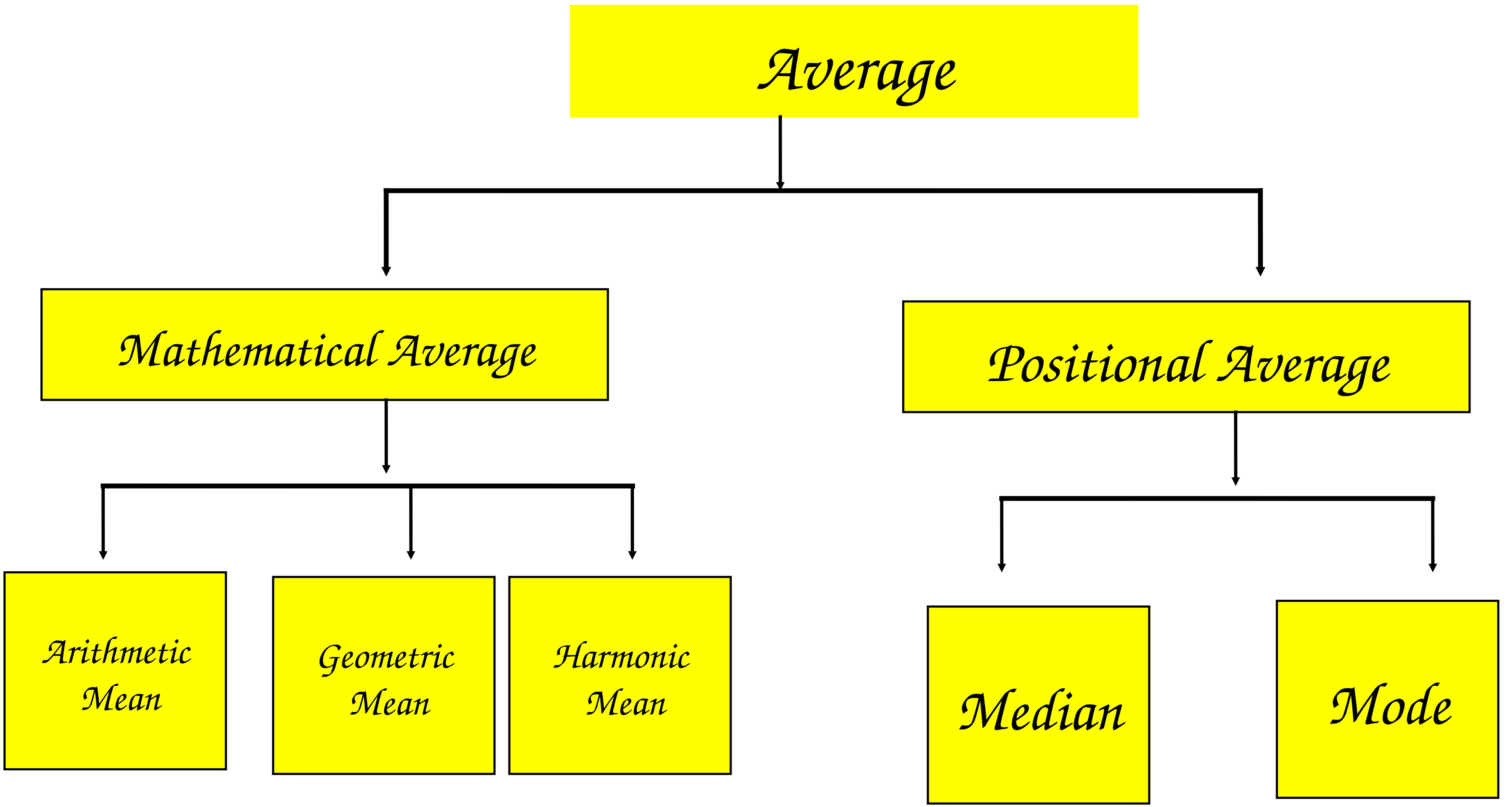
Measure of Central Tendency

- In statistics, a **central tendency** (or, more commonly, a **measure of central tendency**) is a central or typical value for a probability distribution.

- It may also be called a **center** or **location** of the distribution.
  It helps you find the middle, or the average, of a data set.

- The most common measures of central tendency are the
  - Arithmetic mean,
  - Median
  - Mode
- A central tendency can be calculated for either a finite set of values or for a theoretical distribution, such as the normal distribution.

MEASURES OF CENTRAL TENDENCY



The Arithmetic Mean (Mean/ Average)

- This is the most widely used measure of central tendency of any group of data.

$$\boxed{Sample\ \ Mean = \frac{Sum\quad of\quad observations}{Total\quad number\quad of\quad observations}}$$

- If $x_1, x_2, \ldots\ldots\ldots, x_n$ are sample observations for a variable X, then the sample mean (arithmetic mean) is usually denoted by;

$$\boxed{\bar{x} = \frac{x_1 + x_2 + \ldots\ldots\ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

Example:

- Consider the following set of data and find the mean

5, 6, 2, 4, 7, 8, 3, 5, 6, 6

Example:

- Find the mean

5, 6, 2, 4, 7, 8, 3, 5, 6, 6

$$\frac{5+6+2+4+7+8+3+5+6+6}{10}$$

$$= \frac{52}{10}$$

$$= 5.2$$

Example :

let us consider the monthly salary ($) of 10 employees of a firm.

2500, 2700, 2400, 2300, 2550, 2650, 2750, 2450, 2600, 2400

Calculate the arithmetic mean.

Example:

let us consider the monthly salary ($) of 10 employees of a firm.

2500, 2700, 2400, 2300, 2550, 2650, 2750, 2450, 2600, 2400

- The arithmetic mean is

$$\overline{x} = \frac{2500+2700+2400+2300+2550+2650+2750+2450+2600+2400}{10} = 2530.$$

Example (04) :

The arithmetic mean of the numbers 8,3,5,112,10 is, The arithmetic mean of the numbers 8,3,5,112,10 is,

$$\boxed{\bar{x} = \frac{8+3+5+112+10}{5} = \frac{138}{5} = 27.6}$$

Note :

- <u>Sample mean does not provide a good measure when the sample contains a few extreme values</u>.

Sample Median

- When the items of a series are arranged in ascending or descending order of magnitude, the value of the middle item in the series is known as median in the case of individual observations.

- If the total number of items in a series are odd then the value of the $\left(\frac{n+1}{2}\right)^{th}$ item gives the median.

- On the other hand, if the total number of items in a series are even then the value of the $\frac{\left(\frac{n}{2}\right)^{th}+\left(\frac{n}{2}+1\right)^{th}}{2}$ item gives the median.

Example

1. Find the median of the following items.

8, 4, 8, 3, 4, 8, 6, 5, 10

2. Find out the value of the median of the following items.

15, 12, 5, 7, 9, 5, 11, 28

Example

1. Find the median of the following items.

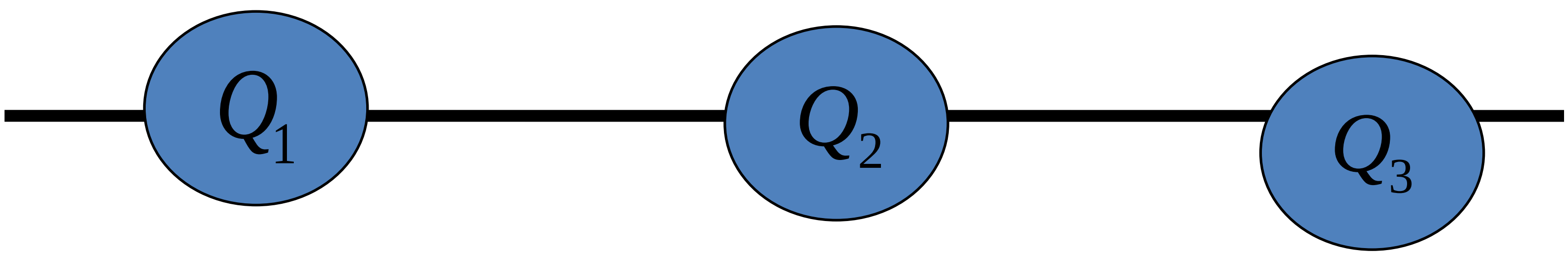
8, 4, 8, 3, 4, 8, 6, 5, 10

2. Find out the value of the median of the following items.

15, 12, 5, 7, 9, 5, 11, 28

Quartiles (1st method)

They are the points that divide the ordered *(ascending or descending)* *data set in to three equal points. They denoted by,*



$$Q_1 = \left(\frac{n+1}{4}\right)^{th} \textit{item}$$

$$Q_3 = \left(\frac{3[n+1]}{4}\right)^{th} \textit{item}$$

$$Q_2 = \left(\frac{n+1}{2}\right)^{th} \textit{item}$$

Example:

Find queries,

(i) 900 72 82 98 70 65 86

(ii) 65 70 82 98 83

(iii)65 70 82 98 83 72

900 72 82 98 70 65 86

Ascending order →

65, 70, 72, 82, 83, 98, 900

$$Q_1 = \left(\frac{n+1}{4} \right)^{th} item$$

$$= \left(\frac{7+1}{4} \right)^{th} item = 2^{nd}\ item = 70$$

$$Q_2 = \left(\frac{n+1}{2} \right)^{th} item$$

$$= \left(\frac{7+1}{2} \right)^{th} item = 4^{th}\ item = 82$$

$$Q_3 = \left(\frac{3[n+1]}{4} \right)^{th} item$$

$$= \left(\frac{3[7+1]}{4} \right)^{th} item = 6^{th}\ item = 98$$

65 70 82 98 83

Ascending order

65 70 72 82 83 98

$$Q_1 = \left(\frac{n+1}{4}\right)^{th} item$$

$$= \left(\frac{6+1}{4}\right)^{th} item = 1.75^{th} \ item$$

$$= 65 + 5(Difference)0.75 = 68.75$$

$$Q_2 = \left(\frac{n+1}{2}\right)^{th} item$$

$$= \left(\frac{6+1}{2}\right)^{th} item = 3.5^{th} item$$

$$= 72 + 10 * 0.5 = 77$$

$$Q_3 = \left(\frac{3[n+1]}{4}\right)^{th} item$$

$$= \left(\frac{3[6+1]}{4}\right)^{th} item = 5.25^{th} \ item$$

$$= 83 + 15 * 0.25 = 86.75$$

Mode

The mode of the set of number is that value which occurs most frequently.

If each observation occurs the same number of times, then there is no mode.

Example :

(i) 14,19,16,21,18,19,24,15,19

(ii) 6,7,7,3,8,5,3,9,3,3,5,6,7,7,8,9

(iii) 6,7,8,9

Mode

The mode of the set of number is that value which occurs most frequently.

If each observation occurs the same number of times, then there is no mode.

Example :

(i) 14,19,16,21,18,19,24,15,19

14,15,16,18,19,19,21,24

Mode=19

(ii) 6,7,7,3,8,5,3,9,3,3,5,6,7,7,8,9

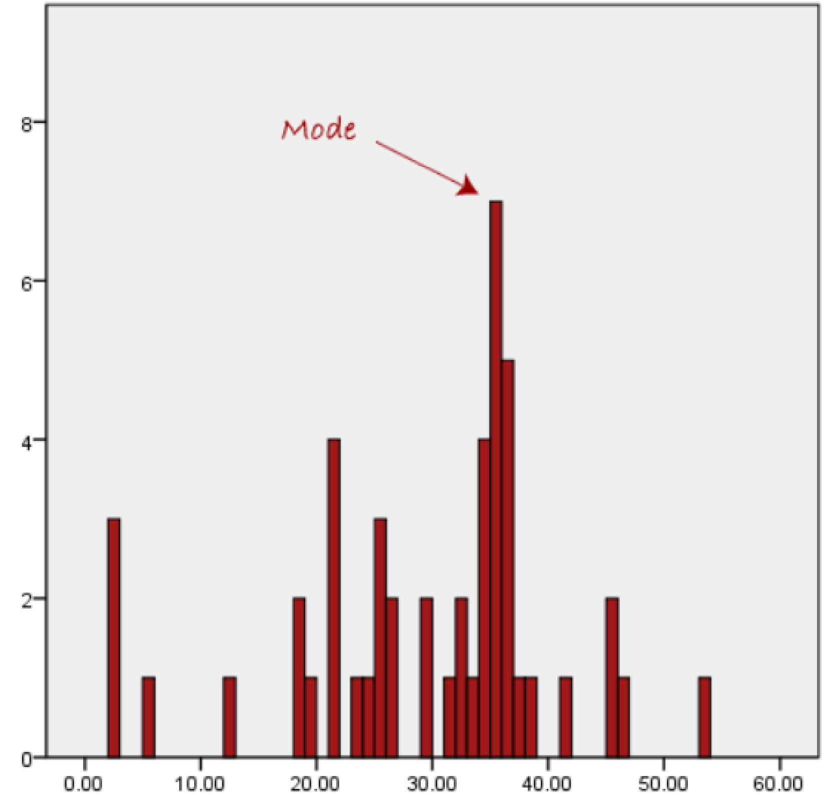
3,3,3,3,5,5,6,6,7,7,7,7,8,8,9,9,

Mode=3 &7

(iii) 6,7,8,9 No Mode

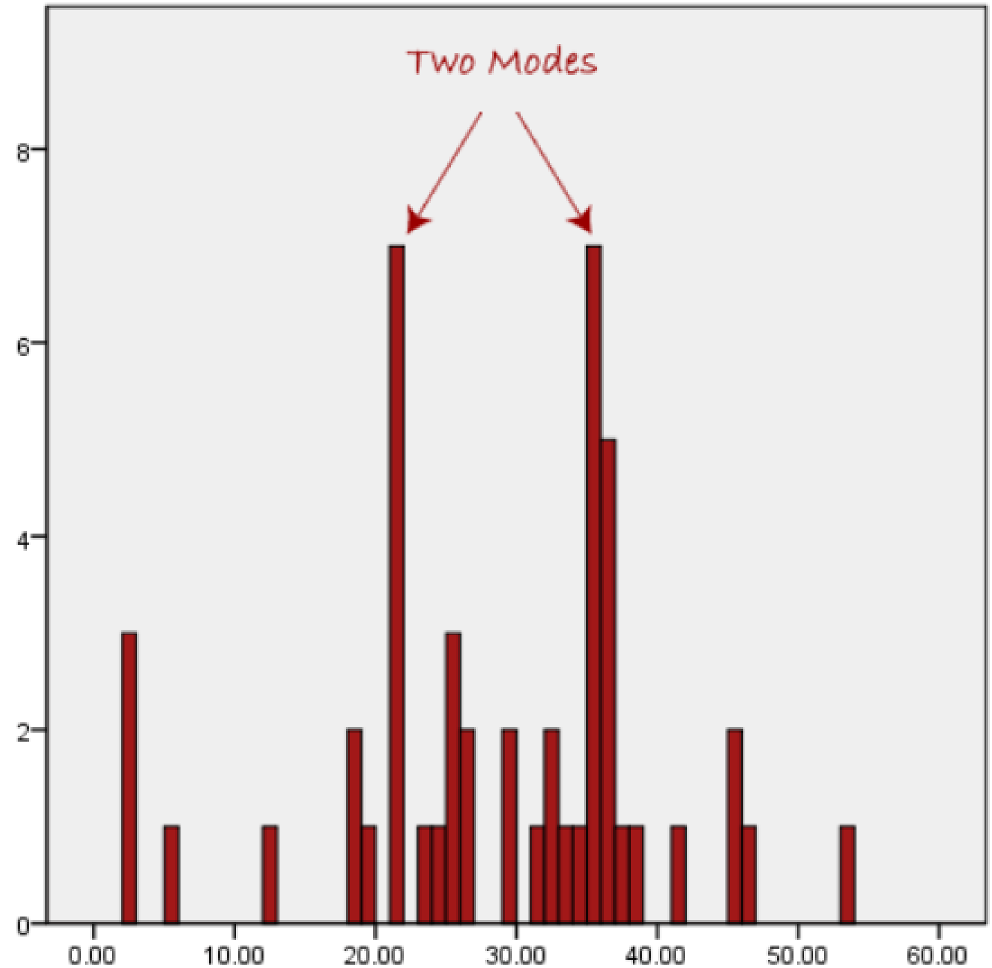
Mode

- The mode is the most frequent score in our data set. On a histogram it represents the highest bar in a bar chart or histogram.
- You can, therefore, sometimes consider the mode as being the most popular option.

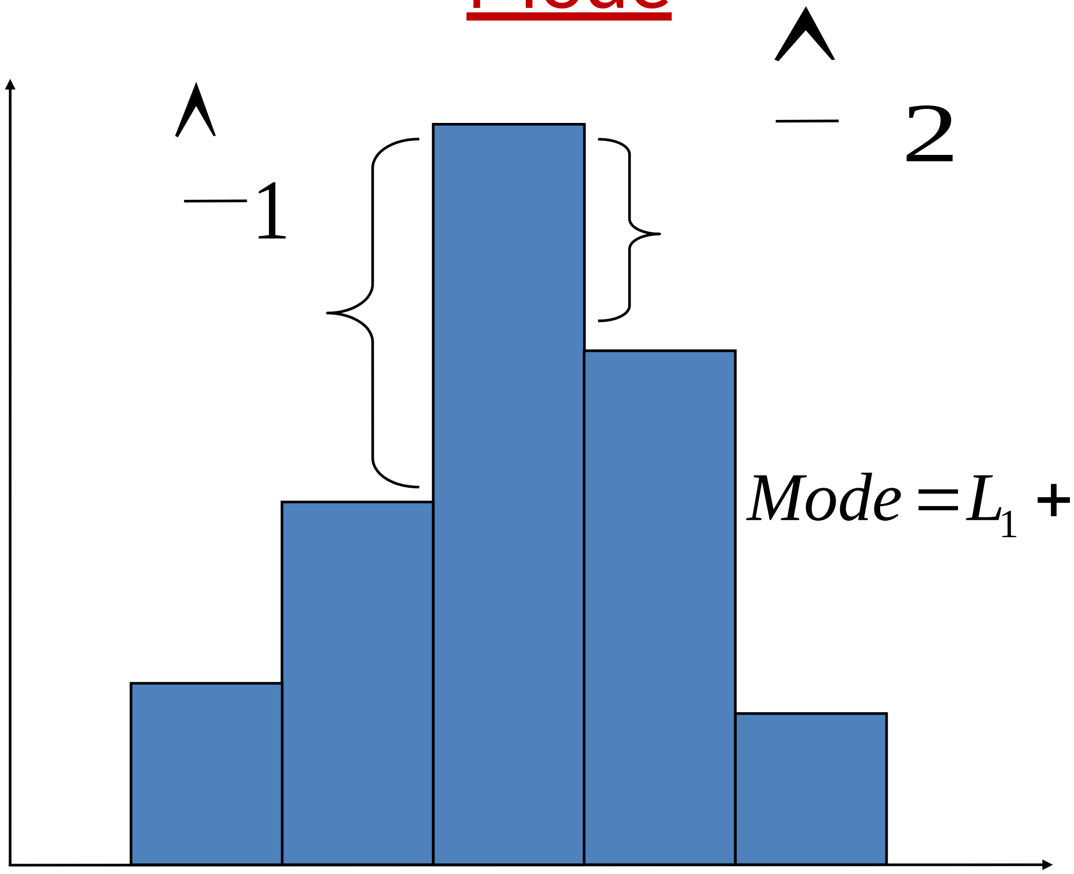


Mode

- However, one of the problems with the mode is that it is not unique, so it leaves us with problems when we have two or more values that share the highest frequency, such as below:



Mode



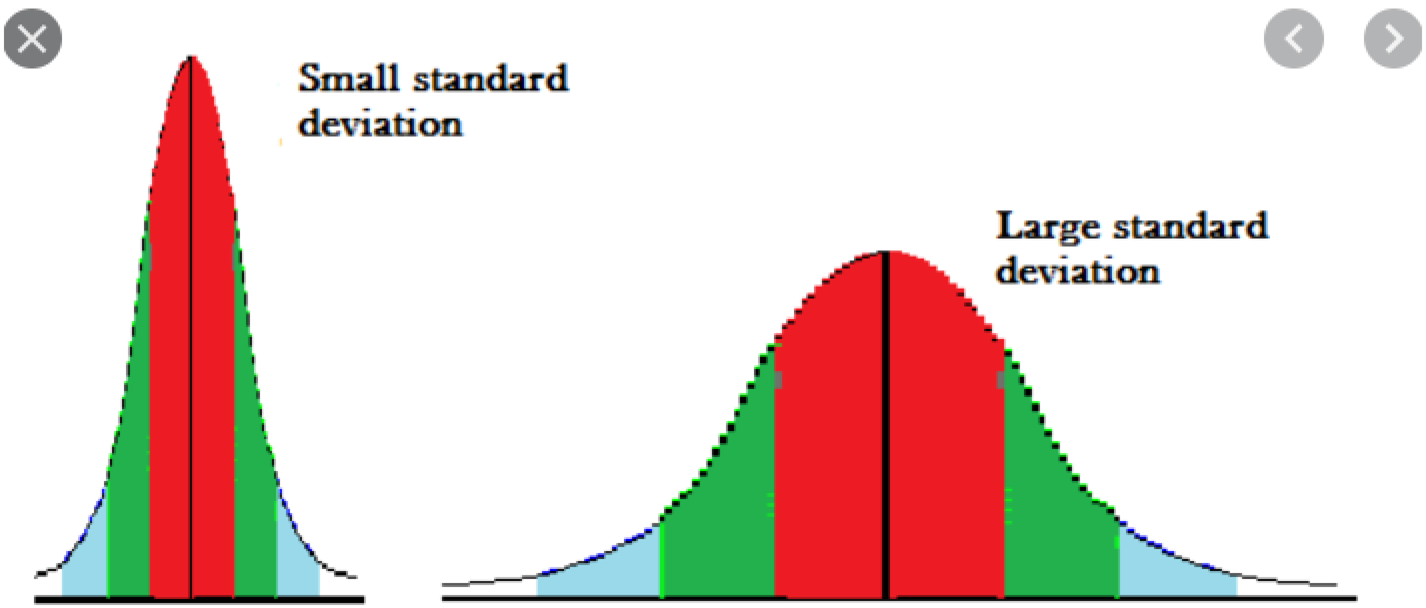
$$Mode = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) C$$

L_1 *- Lower class boundary of highest frequency class*

Summary of when to use the mean, median and mode

Type of Variable	Best measure of central tendency
Nominal	Mode
Ordinal	Median
Interval/Ratio (not skewed)	Mean
Interval/Ratio (skewed)	Median

Measures of Variability or Spread or Dispersion of a data set



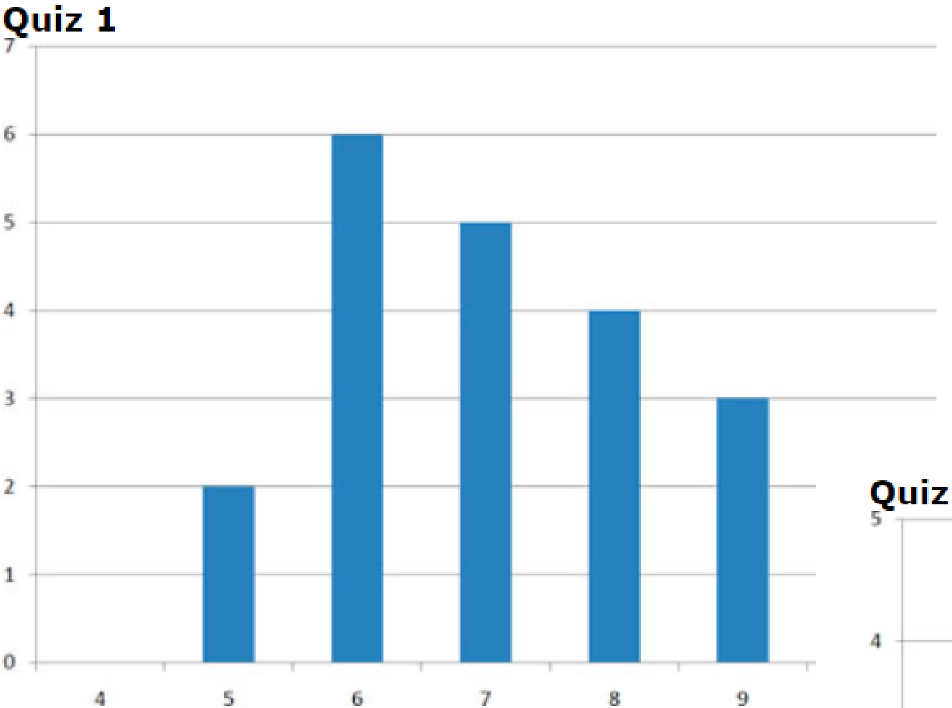
Measures of Variability

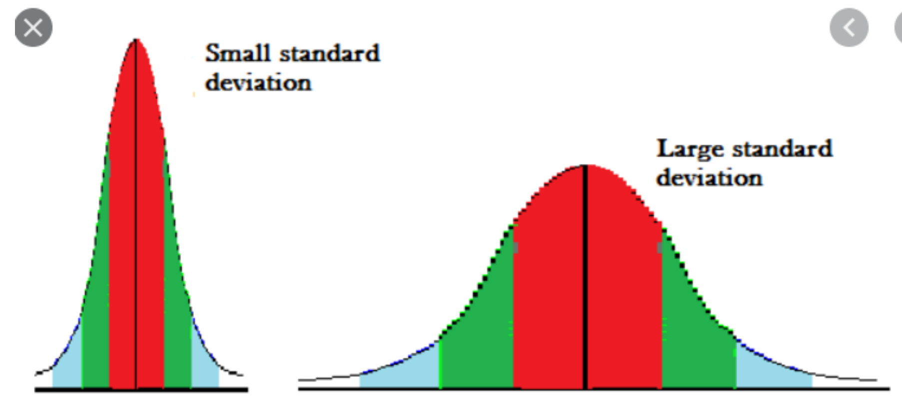
- A measure of variability is a summary statistic that represents the amount of dispersion in a dataset.

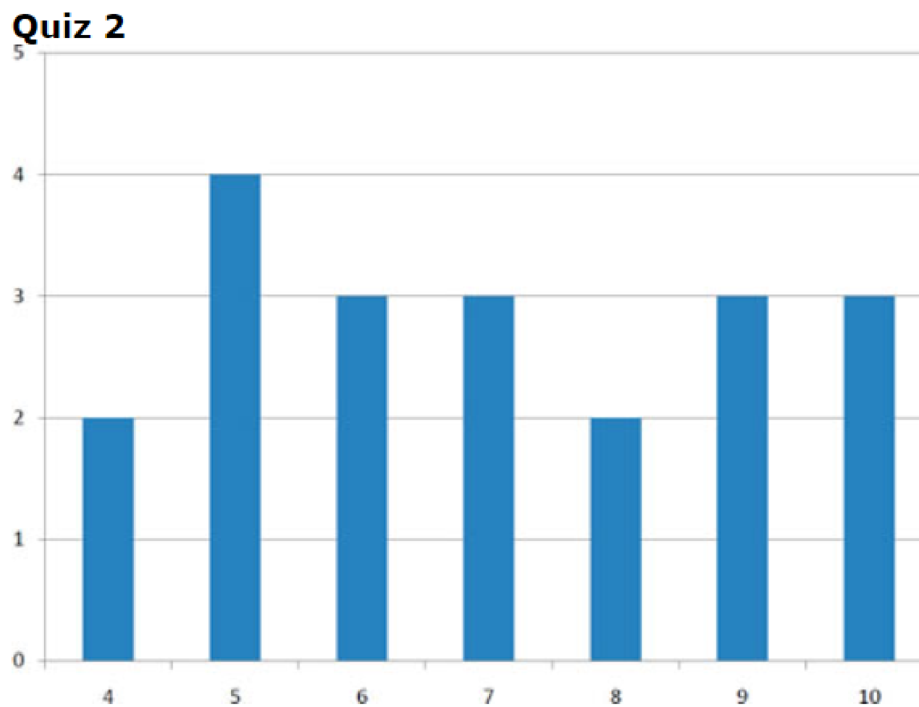
- Variability refers to how "spread out" a group of scores is.

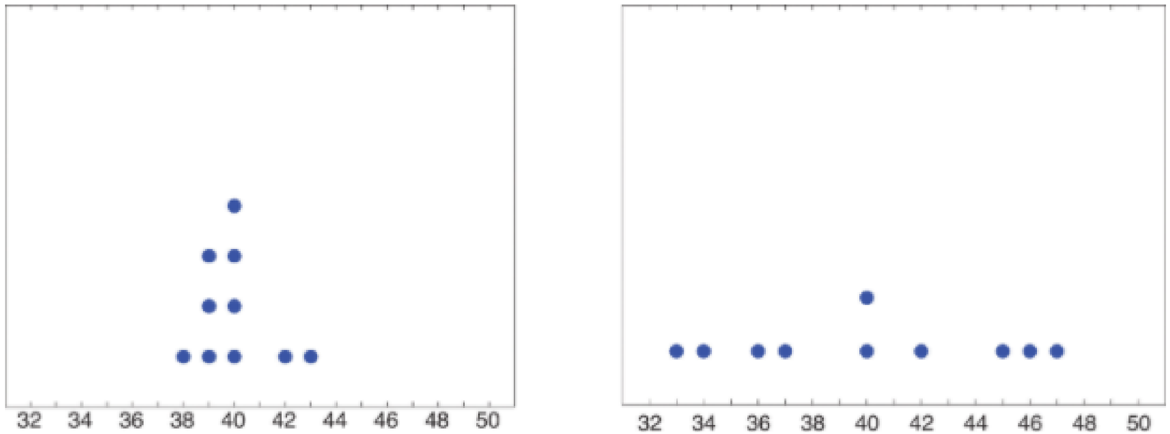
- In statistics, variability, dispersion, and spread are synonyms that denote the width of the distribution.

- The differences among students were much greater on Quiz 2 than on Quiz 1.









(a) Set I (b) Set II

- The two sets of ten measurements each center at the same value: they both have mean, median, and mode 40.

- Nevertheless, a glance at the figure shows that they are markedly different.
- In Data Set I the measurements vary only slightly from the center, while for Data Set II the measurements vary greatly.

- These new quantities are called measures of variability, and we will

Absolute measures of dispersion

1. Positional measures

   Range, Semi-Interquartile Range, Quartile deviation

2. Calculated measures

   Mean deviation, Standard deviation

Relative measures of dispersion

1. Coefficient of range
2. Coefficient of quartile deviation
3. Coefficient of mean deviation
4. Coefficient of standard deviation
5. Coefficient of variation

Range

- Range is the simplest method of studying dispersion.
- It is simply the distance between the largest (L) and the smallest (S) value in a group of items.

$$\mathit{The\ Range\ of\ the\ Data\ set} =$$

$$\mathit{Largest\quad number} - \mathit{Smallest\ Number}$$

Mean deviation (about the mean)

- The deviations of the variety of values from the mean are another method of measuring variability.

- We have numbers $x_1, x_2, \ldots\ldots x_n$. Then

$$Mean\, deviation = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

Example:

Find the mean deviation of the set of numbers,
2 ,3,6,8,11

$$\bar{x} = \frac{2+3+6+8+11}{5} = 6$$

$$Mean\, deviation = \frac{\sum_{i=1}^{n} \left| x_i - \bar{x} \right|}{n} |$$

$$= \frac{|2-6|+|3-6|+|6-6|+|8-6|+|11-6|}{5}$$

$$= \frac{14}{5} = 2.8$$

Inter quartile range

Inter quartile range ; I_{QR}

$$I_{QR} = Q_3 - Q_1$$

Inter Quartile Deviation *(semi Interquartile Range)*

The dependence of range on extreme items can be avoided by adopting this measure.

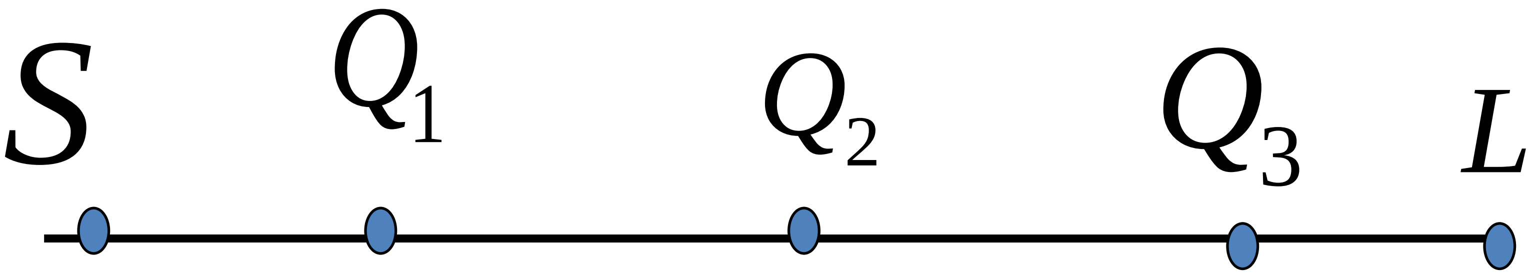
$$\boxed{Q_D = \frac{(Q_3 - Q_1)}{2}}$$

Coefficient of quartile deviation

Quartile deviation is an absolute measure of dispersion. The relative measures corresponding to quartile deviation is called the coefficient of quartile deviation and is expressed as follows.

$$\boxed{CQD = \frac{(Q_3 - Q_1)}{(Q_3 + Q_1)}}$$

Five number summaries



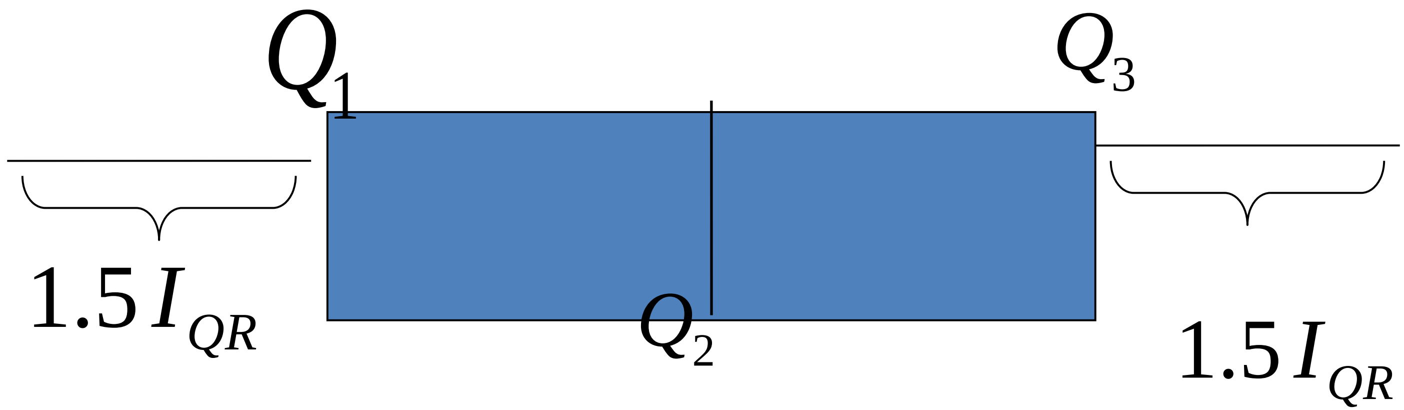
S - smallest observation

L - Largest observation

Q_1, Q_2, Q_3 - Quartiles

Box plot

- The box plot is a graphical display, that simultaneously displays important features of the data, such as,
  - location
  - Central Tendency (Median)
  - Spread or variability(Variance and Stranded deviation, Inter quartile Range)
  - unusually observations (outliers)



Standard Deviation and Variance

- A commonly used measure of dispersion is the **standard deviation**, which is simply the square root of the **variance**.
- The variance and the standard deviation are measures of how spread out a distribution is. In other words, they are measures of va

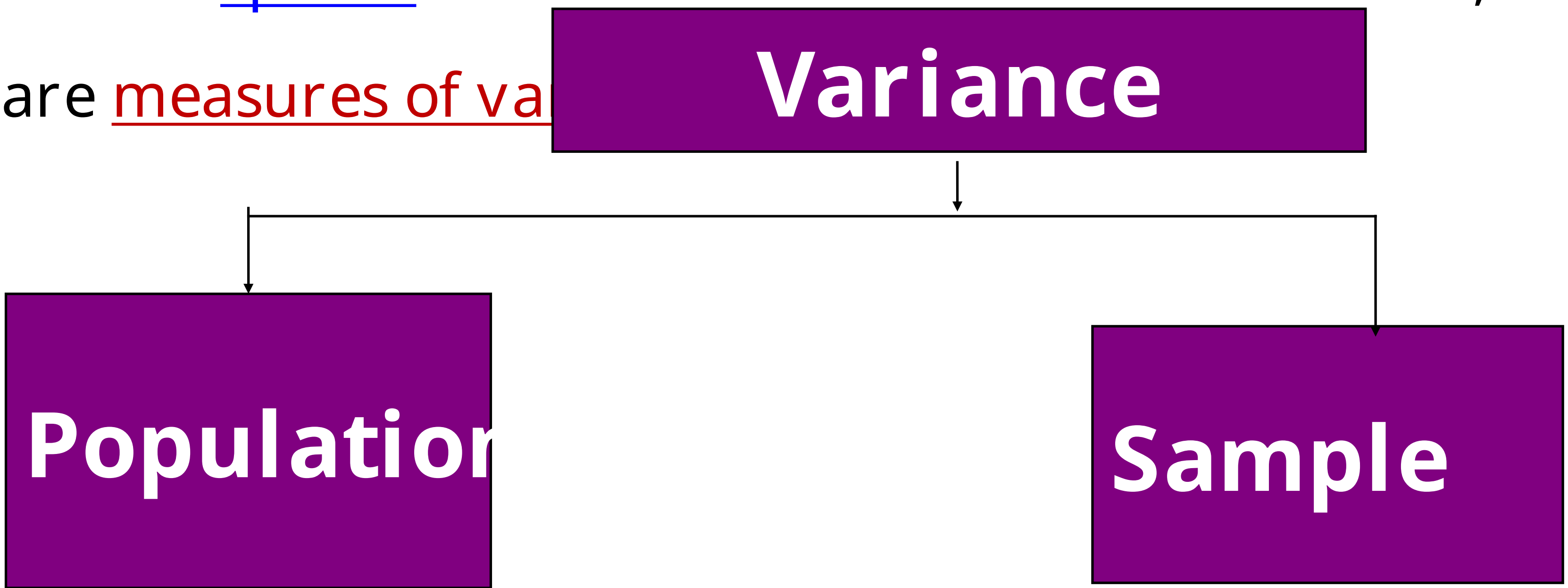


Table 1: Mean and Standard Deviation				
Measure Name	Symbol for Population	Symbol for Sample	Computation for Population	Computation for Sample
Mean	μ	$\bar{x}$	$\mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$	$\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$
Standard Deviation	σ	s_x	$\sigma = \sqrt{\dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$ or the equivalent form $\sigma = \sqrt{\dfrac{\sum_{i=1}^{N} (x^2) - \dfrac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}}{N}}$	$s_x = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$ or the equivalent form $s_x = \sqrt{\dfrac{\sum_{i=1}^{n} (x^2) - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}}$

Definition

The **sample variance** *of a set of n sample data is the number* s^2 *defined by the formula*

$$s^2 = \frac{\Sigma(x - \overline{x})^2}{n-1}$$

which by algebra is equivalent to the formula

$$s^2 = \frac{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2}{n-1}$$

The **sample standard deviation** *of a set of n sample data is the square root of the sample variance, hence is the number s given by the formulas*

$$s = \sqrt{\frac{\Sigma(x - \overline{x})^2}{n-1}} = \sqrt{\frac{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2}{n-1}}$$

Variance and Standard deviation for sample

- *The square root* of the variance is known as the stranded deviation.

$$S\tan derd\ deviation\ (S) = \sqrt{Varaince}$$

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

$$S = \sqrt{\frac{\sum_{i=1}^{n} x_i^{\ 2} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}}$$

Example

- Find the sample variance and the sample standard deviation of Data Set I and Data set II

Data Set I:	40	38	42	40	39	39	43	40	39	40
Data Set II:	46	37	40	33	42	36	40	47	34	45

$$s = \sqrt{\frac{\Sigma(x-\overline{x})^2}{n-1}} = \sqrt{\frac{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2}{n-1}}$$

Data Set I:	40	38	42	40	39	39	43	40	39	40
Data Set II:	46	37	40	33	42	36	40	47	34	45

To use the defining formula (the first formula) in the definition we first compute for each observation x its deviation $x-\overline{x}$ from the sample mean. Since the mean of the data is $\overline{x}=40$, we obtain the ten numbers displayed in the second line of the supplied table.

x	46	37	40	33	42	36	40	47	34	45
$x-\overline{x}$	6	−3	0	−7	2	−4	0	7	−6	5

Then

$$\Sigma(x-\overline{x})^2=6^2+(-3)^2+0^2+(-7)^2+2^2+(-4)^2+0^2+7^2+(-6)^2+5^2=224$$

so

$$s^2=\frac{\Sigma(x-\overline{x})^2}{n-1}=\frac{224}{9}=24.\overline{8}$$

and

$$s=\sqrt{\frac{\Sigma(x-\overline{x})^2}{n-1}}=\sqrt{\frac{\Sigma x^2-\frac{1}{n}(\Sigma x)^2}{n-1}}$$

$$s=\sqrt{24.\overline{8}}\approx 4.99$$

Example

Find the sample variance and the sample standard deviation of the ten GPAs

1.90 3.00 2.53 3.71 2.12 1.76 2.71 1.39 4.00 3.33

$$s = \sqrt{\frac{\Sigma(x-\overline{x})^2}{n-1}} = \sqrt{\frac{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2}{n-1}}$$

Example

Find the sample variance and the sample standard deviation of the ten GPAs

1.90   3.00   2.53   3.71   2.12   1.76   2.71   1.39   4.00   3.33

$$\Sigma x = 1.90 + 3.00 + 2.53 + 3.71 + 2.12 + 1.76 + 2.71 + 1.39 + 4.00 + 3.33 = 26.45$$

and

$$\begin{aligned}\Sigma x^2 &= 1.90^2 + 3.00^2 + 2.53^2 + 3.71^2 + 2.12^2 + 1.76^2 \\ &\qquad +2.71^2 + 1.39^2 + 4.00^2 + 3.33^2 \\ &= 76.7321\end{aligned}$$

the shortcut formula gives

$$s^2 = \frac{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2}{n-1} = \frac{76.7321 - \frac{(26.45)^2}{10}}{10-1} = \frac{6.77185}{9} = .75242\overline{7}$$

and

$$s = \sqrt{.75242\overline{7}} \approx .867$$

Example (01) :

Calculate the variance and standard deviation

8,3,5,112,10

Measures of Skewness

- Skewness, like measures of central value and dispersion is a measure for the study of frequency distributions.
- Thus skewness measures the degree of departure from symmetry.
- When a distribution is symmetrical, skewness is absent and the values of the mean, median and mode coincide.
- The mean and median are pulled away from the mode, either to the right or to the left.
- When mean and median are pulled towards right, the skewness is positive and otherwise negative.

Measure of skewness

There are three important measures of skewness.

1. The Karl Pearson’s coefficient of skewness

2. The bowly’s coefficient of skewness

3. Coefficient of skewness, based on moments

The Karl Pearson's coefficient of skewness

The Karl Pearson's coefficient of skewness is denoted by ,

$$S_{kp} = \frac{mean - mode}{s\tan dard\ Deviation}$$

It is also known as Pearson's coefficient of skewness. If the mode is ill defined,

$$S_{kp} = \frac{3(mean - median)}{s\tan dard\ Deviation}$$

- If $S_{kp} = \mathbf{0}$, the distribution of data is symmetric.
- If $S_{kp} > 0$, the distribution of data is skewed to the right.
- If $S_{kp} < 0$, the distribution of data is skewed to the left.

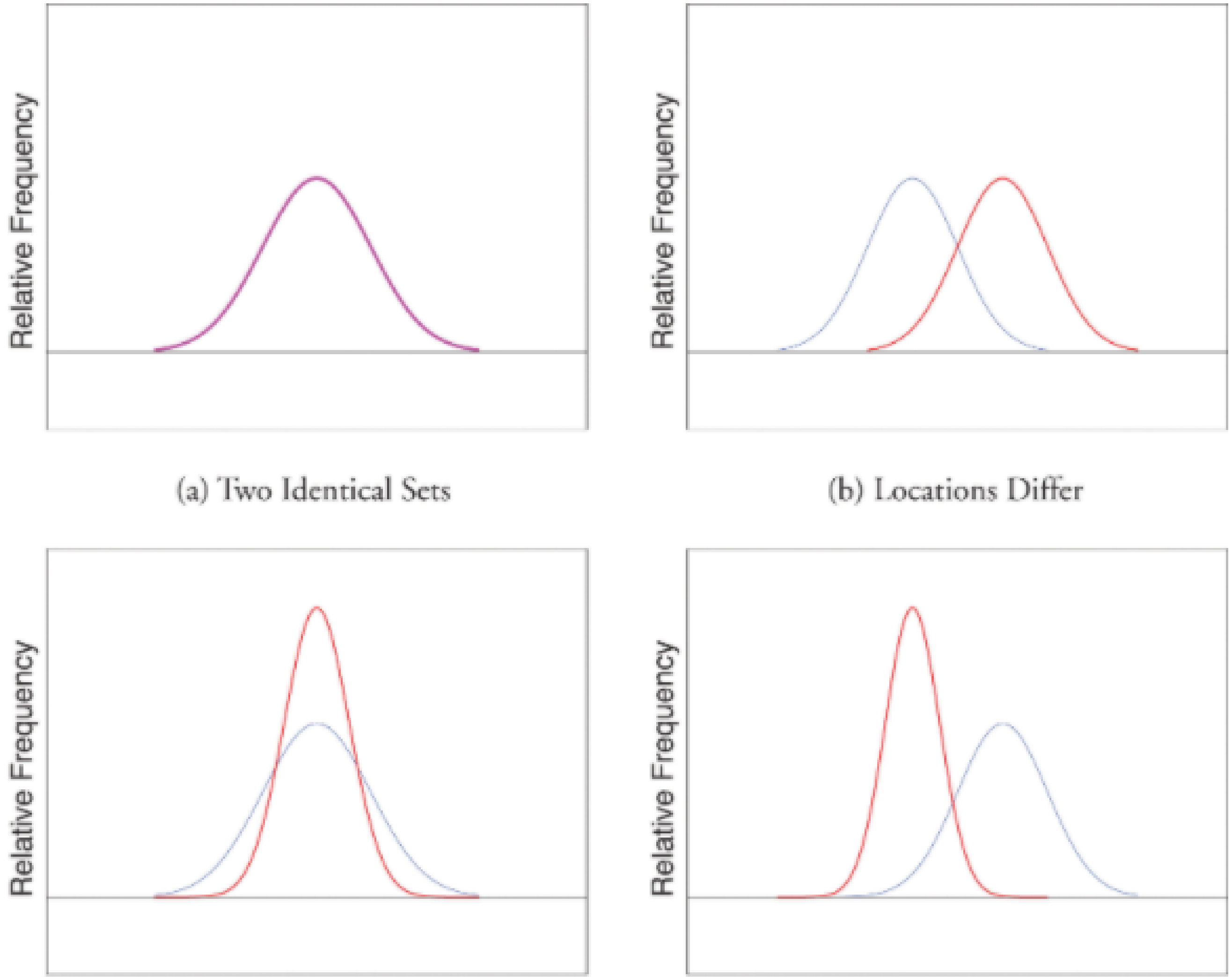
The numerical value given by this formula usually varies between -1 and +1.

$$\boxed{Mean = Median = Mode \Rightarrow \ Symmetric}$$

$$\boxed{Median < Mean \Rightarrow \ Positive\ Skewness}$$

$$\boxed{Mean < Median \ \Rightarrow \ Negative\ Skewness}$$

Difference between Two Data Sets



(a) Two Identical Sets

(b) Locations Differ

(c) Variabilities Differ

(d) Locations and Variabilities Differ

The Karl Pearson's coefficient of skewness

The Karl Pearson's coefficient of skewness is denoted by ,

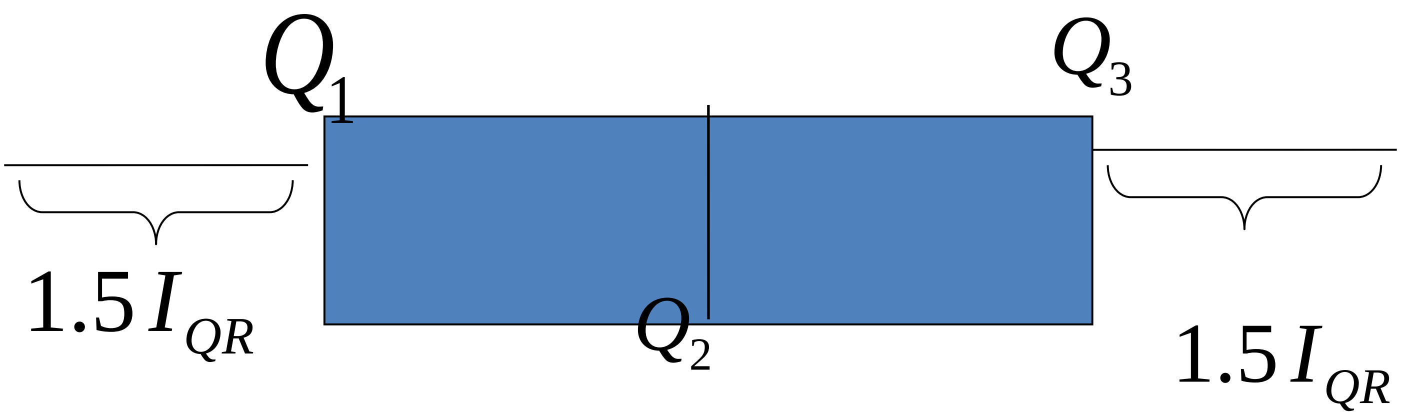
$$\boxed{S_{kp} = \frac{mean - \mathrm{mode}}{s\tan dard\ Deviation}}$$

It is also known as Pearson's coefficient of skewness. If the mode is ill defined,

$$\boxed{S_{kp} = \frac{3(mean - median)}{s\tan dard\ Deviation}}$$

Box plot

- The box plot is a graphical display, that simultaneously displays important features of the data, such as,
  - location
  - Central Tendency (Median)
  - Spread or variability(Variance and Stranded deviation, Inter quartile Range)
  - unusually observations (outliers)



Summary Examples

The Nicotine contents, in milligrams for 50 cigarettes (sample) of a certain brand were selected randomly from their population (1000) and recorded) .

31	22	25	24	28
41	43	47	34	38

Calculate the

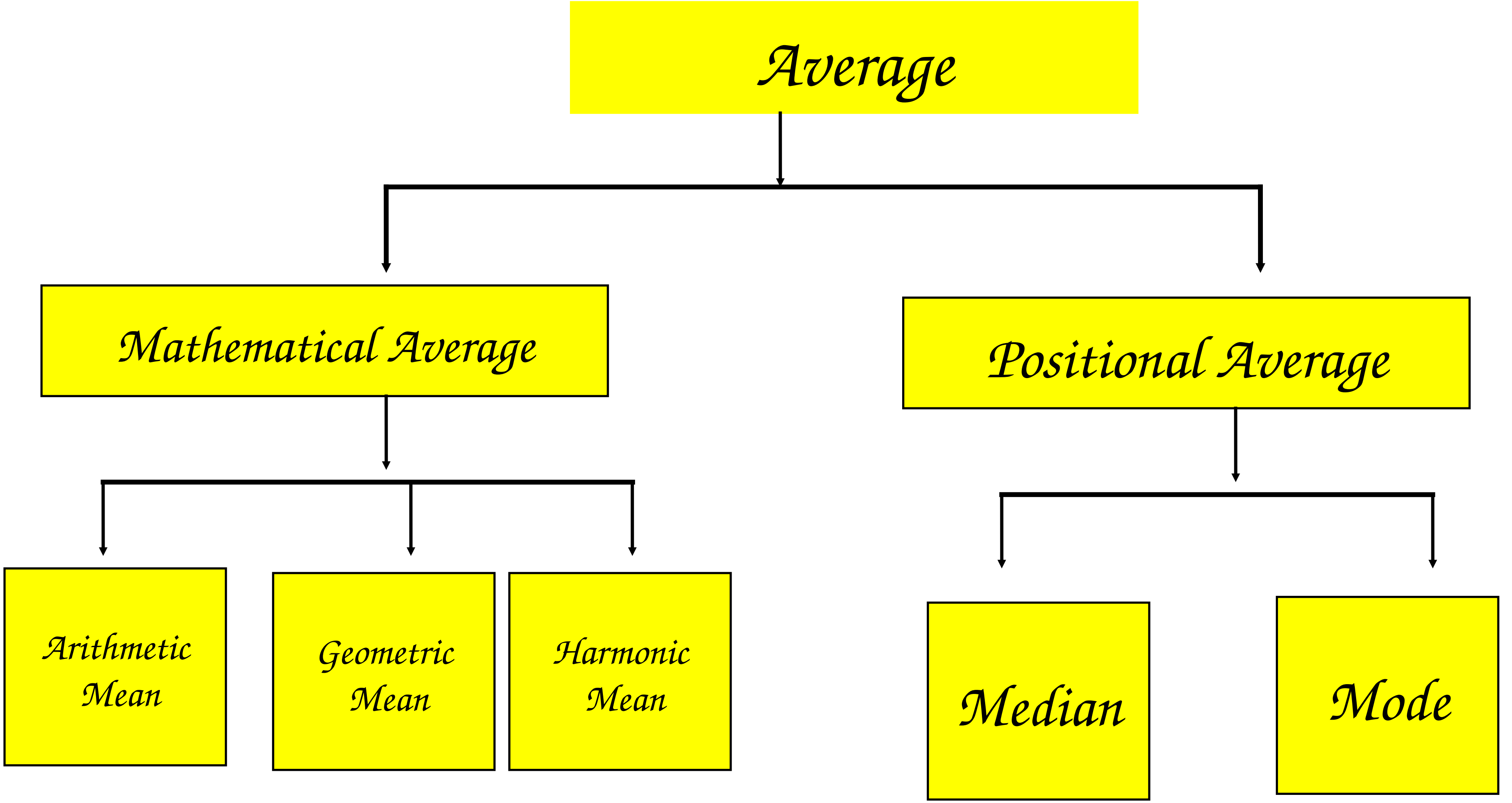
1. Sample mean
2. Median
3. Mode
4. Q1 and Q3
5. sample variance and standard deviation.
6. Discuss the skewness
7. Draw a box plot and identify the outliers

Geometric mean

Harmonic Mean

Weighted Arithmetic Mean

MEASURES OF CENTRAL TENDENCY



What Is the Geometric Mean?

- The geometric mean is most useful when numbers in the series are not independent of each other or if numbers tend to make large fluctuations.

- Geometric Mean is well defined only for sets of positive real numbers.

- Applications of the geometric mean are most common in business and finance, where it is frequently used when dealing with percentages to calculate growth rates and returns on a portfolio of securities.

Geometric mean

- This is calculated by multiplying all the numbers (call the number of numbers n), and taking the nth root of the total.

Geometric Mean

$$\bar{x}_{\text{geom}} = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

Arithmetic Mean

$$\bar{x}_{\text{arithm}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

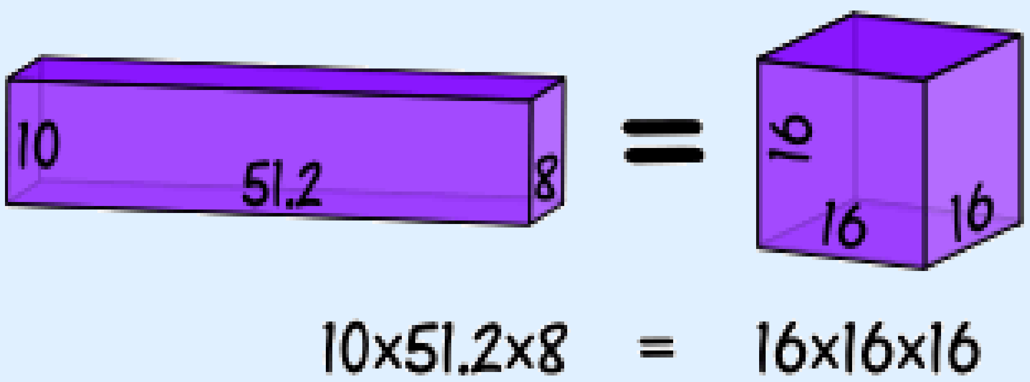
Example: What is the Geometric Mean of 10, 51.2 and 8?

- First we multiply them: $10 \times 51.2 \times 8 = 4096$
- Then (as there are three numbers) take the cube root: $\sqrt[3]{4096} = \mathbf{16}$

In one line:

$$\textbf{Geometric Mean} = \sqrt[3]{(10 \times 51.2 \times 8)} = 16$$

It is like the volume is the same:



Example: What is the Geometric Mean of 1, 3, 9, 27 and 81?

Example: What is the Geometric Mean of 1, 3, 9, 27 and 81?

- First we multiply them: 1 x 3 x 9 x 27 x 81 = 59049
- Then (as there are 5 numbers) take the 5th root: $\sqrt[5]{59049} = \mathbf{9}$

In one line:

$$\textbf{Geometric Mean} = \sqrt[5]{(1 \times 3 \times 9 \times 27 \times 81)} = 9$$

I can't show you a nice picture of this, but it is still true that:

$$1 \times 3 \times 9 \times 27 \times 81 = 9 \times 9 \times 9 \times 9 \times 9$$

Example

1. Calculate the Geometric Mean of 2, 4, 8 and 16

2. Calculate the Geometric Mean of the first eight natural numbers.

3. The Geometric Mean of two numbers is 27. One of the numbers is 81. What is the other number?

Harmonic Mean

- The harmonic mean is a very specific type of average.

- It's generally used when dealing with averages of units, like speed or other rates and ratios.

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

- Add the reciprocals of the numbers in the set. Divide the number of items in the set .

What is the harmonic mean of 1,5,8,10?

What is the harmonic mean of 1,5,8,10?

- Add the reciprocals of the numbers in the set: 1/1 + 1/5 + 1/8 + 1/10 = 1.425
- Divide the number of items in the set by your answer to Step 1.
- There are 4 items in the set, so: 4 / 1.425 = 2.80702

Example :

- Calculate the harmonic mean for the following data:

x	1	3	5	7	9	11
f	2	4	6	8	10	12

Example :

- Calculate the harmonic mean for the following data:

x	1	3	5	7	9	11
f	2	4	6	8	10	12

x	f	1/x	f/x
1	2	1	2
3	4	0.333	1.332
5	6	0.2	1.2
7	8	0.143	1.144
9	10	0.1111	1.111
11	12	0.091	1.092
	N =42		Σ f/x = 7.879

$HM_w = N / [(f_1/x_1) + (f_2/x_2) + (f_3/x_3) + \ldots(f_n/x_n)]$

$HM_w = 42 / 7.879$

$HM_w = 5.331$

Therefore, the harmonic mean, HM_w is 5.331.

Example

- we travel 10 km at 60 km/h, than another 10 km at 20 km/h, what is our average speed?

Example

- we travel 10 km at 60 km/h, than another 10 km at 20 km/h, what is our average speed?

$$\text{Harmonic mean} = 2/(\frac{1}{60} + \frac{1}{20}) = \mathbf{30\ km/h}$$

- Check:
  - the 10 km at 60 km/h takes 10 minutes,
  - the 10 km at 20 km/h takes 30 minutes,
  - so the total 20 km takes 40 minutes, which is 30 km per hour

Example

- During the month the total number of km driven by each truck is given bellow. Find the Harmonic mean

Truck Number	1	2	3	4
Km. driven	40	50	60	75

Example

Truck Number	1	2	3	4
Km. driven	40	50	60	75

x	$1/x$
40	0.02500
50	0.02000
60	0.01677
75	0.01333
	0.07500

$$H.M = \frac{N}{\sum 1/x}$$

$$H.M = \frac{4}{0.07500}$$

$$H.M = 53.33\ Km.$$

Example

Truck Number	1	2	3	4
Km. driven	40	50	60	75

x	$1/x$
40	0.02500
50	0.02000
60	0.01677
75	0.01333
	0.07500

$$H.M = \frac{N}{\sum 1/x}$$

$$H.M = \frac{4}{0.07500}$$

$$H.M = 53.33\ Km.$$

Weighted Arithmetic Mean

$$\bar{x} = \frac{\sum_{i=1}^{n} (x_i * w_i)}{\sum_{i=1}^{n} w_i}$$

- w = the weights.
- x = the value

- If all the values of the data are not equally important, the weighted arithmetic mean is calculated after assigning appropriate weights to the values of the data.
- If all the weights are equal, then the weighted mean is the same as the arithmetic mean.

Problem A: Given that A = 4, B+ = 3.5, B = 3, C+ = 2.5, C = 2, F = 0, determine the grade point average carried by a student in particular semester, based on the following grades.

Course	*Credit hours*	*Grade*
Accounting	4	A
Finance	2	C+
Marketing	3	B+
Statistics	4	B
Management	3	C

x	w	wx
4	4	16
2.5	2	5
3.5	3	10.5
3	4	12
2	3	6
	$\sum w = 16$	$\sum wx = 49.5$

A = 4, B+ = 3.5, B = 3, C+ = 2.5, C = 2, F = 0

Course	*Credit hours*	*Grade*
Accounting	4	A
Finance	2	C+
Marketing	3	B+
Statistics	4	B
Management	3	C

Weighted Arithmetic Mean= 49.5/16=3.094

Therefore the grade average point of the student is 3.094 which falls in B grade.

Problem B: Suppose there are 20 men and 10 women in an office. The average age of 20 men is 30 years while that of the 10 women is 25 years. Find the mean age of the employees working in the office.

Problem B: Suppose there are 20 men and 10 women in an office. The average age of 20 men is 30 years while that of the 10 women is 25 years. Find the mean age of the

Solution:

By putting the values from above problem into the table we get:

Employees	*x*	*w*	*wx*
Men	30	20	600
Women	25	10	250
		$\Sigma w=30$	$\Sigma wx=850$

Weighted Arithmetic Mean= 850/30=28.33

Therefore the average age of employees working in the office is 28.33 years.

Weighted Harmonic Mean

- Calculating weighted harmonic mean is similar to the simple harmonic mean.
- It is a special case of harmonic mean where all the weights are equal to 1.
- If the set of weights such as $w_1, w_2, w_3, \ldots, w_n$ connected with the sample space $x_1, x_2, x_3, \ldots, x_n$, then the weighted harmonic mean is defined by

$$HM_w = \frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} \frac{w_i}{x_i}}$$


THE END