
Measures of Central Tendency

Introduction

- Measures of central tendency (average) helps to condense a mass of data into a single representative value
- An average is a single value intended to represent a data set as a whole
- Best averages are:
 - based on all the observations
 - simple to understand and easy to interpret
 - easily manipulated algebraically
 - little affected by fluctuations of sampling
 - should not influenced by extreme values

- Objectives of measuring central tendency:
 - To get one single value that describes the entire data set
 - To facilitate comparison between different data sets

The summation notations

- A variable X has values $x_1, x_2, \dots, x_n$. The sum $x_1 + x_2 + \dots + x_n$ can be written in a compact form as:

$$x_1 + x_2 + \dots + x_n = \sum_{i=1}^n x_i$$

Properties of summation notations:

where k, a and b are any constant

$$1. \sum_{i=1}^n k = nk$$

$$2. \sum_{i=1}^n kx_i = k \sum_{i=1}^n x_i$$

$$3. \sum_{i=1}^n (a + bx_i) = na + b \sum_{i=1}^n x_i$$

$$4. \sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i$$

Types of central tendency

- Measures of Central Tendency
 - ✓ Median
 - ✓ Mode
 - ✓ Mean

Mean

$$\bar{X}$$

Arithmetic mean

- Let $x_1, x_2, \dots, x_n$ are the values of a variable X . The arithmetic mean is:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{OR}$$

$$\bar{X} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i}$$

Where x_i is the **class mark (class mid point)**

The population mean is:

$$\mu = \frac{\sum X_i}{N}$$

Example: Obtain the mean of the following number
2, 7, 8, 2, 7, 3, 7

Solution:

Xi	fi	Xifi
2	2	4
3	1	3
7	3	21
8	1	8
Total	7	36

$$\bar{X} = \frac{\sum_{i=1}^4 f_i X_i}{\sum_{i=1}^4 f_i} = \frac{36}{7} = 5.15$$

Example for Arithmetic Mean for Grouped Data
calculate the mean for the following age distribution.

Class	Frequen cy
6- 10	35
11- 15	23
16- 20	15
21- 25	12
26- 30	9
31- 35	6

Solutions

Class	fi	Xi	Xifi
6- 10	35	8	280
11- 15	23	13	299
16- 20	15	18	270
21- 25	12	23	276
26- 30	9	28	252
31- 35	6	33	198
Total	100		1575

$$\bar{X} = \frac{\sum_{i=1}^6 f_i X_i}{\sum_{i=1}^6 f_i} = \frac{1575}{100} = 15.75$$

Special properties of Arithmetic mean

1. The sum of the deviations of a set of items from their mean is always zero.

$$\sum_{i=1}^n (X_i - \bar{X}) = 0.$$

2. The sum of the squared deviations of a set of items from their mean is the minimum

$$\sum_{i=1}^n (X_i - \bar{X})^2 < \sum_{i=1}^n (X_i - A)^2, A \neq \bar{X}$$

3. If $\bar{X}_1$ is the mean of n_1 observations, if $\bar{X}_2$ is the mean of n_2 observations, ... if $\bar{X}_k$ is the mean of n_k observations, then the mean of all the observations in all groups often called the combined mean is given by:

$$\bar{X}_c = \frac{\bar{X}_1 n_1 + \bar{X}_2 n_2 + \dots + \bar{X}_k n_k}{n_1 + n_2 + \dots + n_k} = \frac{\sum_{i=1}^k \bar{X}_i n_i}{\sum_{i=1}^k n_i}$$

Example: In a class there are 30 females and 70 males. If females averaged 60 in an examination and boys averaged 72, find the mean for the entire class.

Solutions:

Female

$$\bar{X}_1 = 60$$

$$n_1 = 30$$

Males

$$\bar{X}_2 = 72$$

$$n_2 = 70$$

$$\bar{X}_c = \frac{\bar{X}_1 n_1 + \bar{X}_2 n_2}{n_1 + n_2} = \frac{\sum_{i=1}^2 \bar{X}_i n_i}{\sum_{i=1}^2 n_i}$$

$$\Rightarrow \bar{X}_c = \frac{30(60) + 70(72)}{30 + 70} = \frac{6840}{100} = 68.40$$

4. If a wrong figure has been used when calculating the mean the correct mean can be obtained without repeating the whole process using:

$$\text{Correct Mean} = \text{Wrong Mean} + \frac{(\text{Correct Val} - \text{Wrong Val})}{n}$$

Where n is total number of observations.

Example: An average weight of 10 students was calculated to be 65. Later it was discovered that one weight was misread as 40 instead of 80 kg. Calculate the correct average weight.

Solutions:

$$\text{Correct Mean} = \frac{\text{Correct Val} + \text{Wrong Val}}{n}$$

$$\text{Correct Mean} = 65 + \frac{(80 - 40)}{10} = 65 + 4 = 69 \text{ kg}$$

5. The effect of transforming original series on the mean.

- a. If a constant k is added/ subtracted to/from every observation then the new mean will be *the old mean $\pm k$ respectively*.
- b. If every observations are multiplied by a constant k then the new mean will be *$k \times \text{old mean}$*

Weighted Arithmetic mean

used to calculate the average when the relative importance of the observations differs. This relative importance is known as weight

If $x_1, x_2, \dots, x_n$ have weights $w_1, w_2, \dots, w_n$, respectively, then the weighted arithmetic mean becomes:

$$\bar{x}_w = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

- **Example:** Suppose that a student obtained the following grades in the first semester of freshman program.

Cours e	Mats	Bio	Chem	Phy	Flen
Credit hours	4	3	3	4	3
Grade	A	C	B	B	C

$$\bar{x}_w = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} = 49/17 = 2.88$$

Properties of arithmetic mean

- Computed for numerical data, it always exists, and unique
- It depends on all observations
- The sum of deviations of the values from the mean is zero
- It is greatly affected by extreme values
- It is not greatly affected by fluctuations in sampling

Geometric mean (GM)

- Is useful in finding the average of percentages, ratios, indexes, or rates.

Example: $GM = \sqrt[n]{x_1 x_2 \dots x_n}$ A company's year-to-year changes in fuel consumption expenditures were 5, 10, 20, 40 and 60 percent. Determine the average yearly percent change in expenditure

Properties of geometric mean

- It depends on all observations
- It is not much affected by extreme values

$$G.M = \sqrt[5]{5 \times 10 \times 20 \times 40 \times 60} = 18.88$$

The Harmonic Mean

The harmonic mean of $X_1, X_2, X_3 \dots X_n$ is denoted by H.M and given by:

$$H.M = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}}$$

In a case of frequency distribution:

$$H.M = \frac{n}{\sum_{i=1}^k \frac{f_i}{X_i}} \quad n = \sum_{i=1}^k f_i$$

If observations $X_1, X_2, \dots, X_n$ have weights $W_1, W_2, \dots, W_n$ respectively, then their harmonic mean is given by

$$H.M = \frac{\sum_{i=1}^n W_i}{\sum_{i=1}^n W_i / X_i}$$

The Harmonic Mean is useful and appropriate in finding average speeds and average rates.

Example: A cyclist pedals from his house to his college at speed of 10 km/hr and back from the college to his house at 15 km/hr. Find the average speed.

Solution: Here the distance is constant

The simple H.M is appropriate for this problem.

$$X_1 = 10 \text{ km/hr}$$

$$X_2 = 15 \text{ km/hr}$$

$$\text{H.M} = \frac{2}{\frac{1}{10} + \frac{1}{15}} = 12 \text{ km/hr}$$

The Mode



- It is the most frequent value in the data set

Properties of mode

- is easy to calculate and understand
- is not affected by extreme values
- is not based on all observations
- is not used in further analysis of data
- is not unique

Mode for Grouped data

If data are given in the shape of continuous frequency distribution, the mode is defined as:

$$\hat{X} = L_{mo} + w \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right)$$

$\hat{X}$ = the mode of the distribution

w = the size of the modal class

$$\Delta_1 = f_{mo} - f_1$$

$$\Delta_2 = f_{mo} - f_2$$

f_{mo} = frequency of the modal class

f_1 = frequency of the class preceding the modal class

f_2 = frequency of the class following the modal class

*The modal class is a class with the highest frequency

■ Example: Following is the distribution of the size of certain farms selected at random from a district.

Calculate the modal class and its frequency.

Size of farms	No. of farms
5-15	8
15-25	12
25-35	17
35-45	29
45-55	31
55-65	5
65-75	3

■ Solutions:

45- 55 is the modal class since it is a class with the highest frequency.

$$L_{mo} = 45$$

$$w = 10$$

$$\Delta_1 = f_{mo} - f_1 = 2$$

$$\Delta_2 = f_{mo} - f_2 = 26$$

$$f_{mo} = 31$$

$$f_1 = 29$$

$$f_2 = 5$$

$$\Rightarrow \hat{X} = 45 + 10 \left(\frac{2}{2 + 26} \right) \\ = 45.71$$

The Median



- It is the middle value for data arranged in order of magnitude
- Median for raw (ungrouped) data:

$$\left\{ \left(\frac{n+1}{2} \right)^{th} \text{ obs if } n \text{ is odd} \right.$$

Properties of median

- It is an average of position $\left\{ \frac{1}{2} \left(\left(\frac{n}{2} \right)^{th} + \left(\frac{n+2}{2} \right)^{th} \right) \text{ obs if } n \text{ is even} \right.$
- It is affected by the number of observations than by extreme values

Median for grouped data

If data are given in the shape of continuous frequency distribution, the median is defined

$$\bar{X} = L_{\text{med}} + \frac{w}{f_{\text{med}}} \left(\frac{n}{2} - c \right)$$

Where

L_{med} = lower class boundary of the median class

w = the size of the median class

n = total number of observations

c = the cumulative frequency less than type preceeding the median class

f_{med} = the frequency of the median class

*The median class is the class with the smallest cumulative frequency (less than type) greater than or equal to $\frac{n}{2}$

Ex Find the median of the following

dis

Class	Frequency
40-44	7
45-49	10
50-54	22
55-59	15
60-64	12
65-69	6
70-74	3

Solutions:

Clas s	Frequ ency	Cumu.Fre q(less than type)
40- 44	7	7
45- 49	10	17
50- 54	22	39
55- 59	15	54
60- 64	12	66
65- 69	6	72
70- 74	3	75

$$\frac{n}{2} = \frac{75}{2} = 37.5$$

39 is the first cumulative frequency to be greater than or equal to 37.5
 $\Rightarrow$ 50-54 is the median class

$$L_{\text{med}} = 49.5 \quad w = 5$$

$$n = 75 \quad c = 17 \quad f_{\text{med}} = 22$$

$$\begin{aligned} \Rightarrow \bar{X} &= L_{\text{med}} + f_{\text{med}} \frac{w}{n} \left(\frac{n}{2} - c \right) \\ &= 49.5 + \frac{5}{22} (37.5 - 17) \\ &= 54.16 \end{aligned}$$

Example: Calculate the mean time spent by automobile workers.

Time	Class mark	Number of workers
15.5-16.5	18.5	3
16.5-27.5	24.5	6
27.5-33.5	30.5	8
33.5-39.5	36.5	4

The mean time spent is 30.74 minutes.

Example: Find the median for the following frequency distribution.

The median 24.5.

Class boundary	Frequency
5.5 – 10.5	1
10.5 – 15.5	2
15.5 – 20.5	3
20.5 – 25.5	5
25.5 – 30.5	4
30.5 – 35.5	3
35.5 – 40.5	2

Example: Calculate the modal time spent by the automobile workers.

Time (in Minutes)	Number of workers
16-21	3
22-27	6
28-33	8
34-39	4
40-45	3
46-51	1
Total	25

- The mode is

Measure of position or location

Measure of position include standard scores and Quantile(Quartile, decile and percentile). They are used to locate the relative position of a data value in data set.

- **Quartile**

Quartile divided the distribution into four groups.

Quartiles are denoted by Q_1, Q_2, Q_3

To find Q_i ($i=1,2,3$) =

For grouped data we have the following formula

$$Q_i = L_{Q_i} + \frac{W}{f_{Q_i}} \left(\frac{iN}{4} - c \right), \quad i=1,2,3$$

Measure of position or location

Where

- L_{Q_i} lower class boundary of the quartile class
- W the size of the quartile class
- N total number of observation
- c the cumulative frequency preceding the quartile class
- f_{Q_i} the frequency of the quartile class

The quartile class is a class with the smallest cumulative frequency $\frac{N}{4}$ greater than or equal to

Measure of position or location

Deciles

Deciles divide the distribution into 10 groups.

Deciles are denoted by $D_1, D_2, \dots, D_9$

■ To find D_i ($i=1,2,3,\dots,9$) $\frac{iN}{10} =$

■ For grouped data we have the following formula

$$D_i = L_{D_i} + \frac{W}{f_{D_i}} \left(\frac{iN}{10} - c \right), \quad i = 1, 2, 3, \dots, 9$$

Measure of position or location

Where

- L_{D_i} lower class boundary of the Decile class
- W the size of the Decile class
- N total number of observation
- c the cumulative frequency preceding the Decile class
- f_{D_i} the frequency of the Decile class

The Decile class is a class with the smallest cumulative frequency $\frac{iN}{10}$ greater than or equal to

Measure of position or location

Percentiles

Percentiles divide the distribution into 100 groups. Percentiles are denoted by $P_1, P_2, P_3, \dots, P_{99}$

■ To find P_i ($i=1,2,3,\dots,99$) =

■ For grouped data we have the following formula

$$P_i = L_{P_i} + \frac{W}{f_{P_i}} \left(\frac{\frac{iN}{100} - c}{f_{P_i}} \right), \quad i = 1, 2, 3, \dots, 99$$

Measure of position or location

Where

- L_{P_i} lower class boundary of the Percentile class
- W the size of the Percentile class
- N total number of observation
- c the cumulative frequency preceding the Percentile class
- f_{P_i} the frequency of the Percentile class

The Percentile class is a class with the smallest cumulative frequency greater than or equal to $\frac{iN}{100}$

Measures of Variation (Dispersion)

■ Introduction

- We have the distribution of yield of two rice varieties from 5 plots each

Variety 1: 45 42 42 41 40

Variety 2: 54 48 42 33 30

- The mean yield of both varieties is 42 kg
- Objectives of measuring variation:
 - To describe dispersion (variability) in a data
 - To compare the spread in two or more distributions
 - To determine the reliability of an average

Absolute and relative measures

- **Absolute measures of dispersion** are expressed in the same unit of measurement in which the original data are given
- **Relative measures dispersion**: It is the ratio of a measure of absolute dispersion to an appropriate measure of central tendency
- It is used when
 - two sets of data are expressed in different units
 - the average sizes are very different
- It is a unit less measure

1.3 Types of measures of variation

The most commonly used measures of dispersions are:

- Range and relative range
- Quartile deviation and coefficient of Quartile deviation
- Mean deviation and coefficient of Mean deviation
- Standard deviation and coefficient of variation.

The Range and Relative range

Range = Max - Min

- Range is the crudest absolute measures of variation
- It is widely used in the construction of quality control charts and description of daily temperature

$$RR = \text{Range}/(\text{Max} + \text{Min})$$

■ Range for grouped data:

If data are given in the shape of continuous frequency distribution, the range is computed as:

$$R = UCL_k - LCL_1, \quad \begin{array}{l} UCL_k \text{ is upper class limit of the last class} \\ LCL_1 \text{ is lower class limit of the first class} \end{array}$$

Or

$$R = X_k - X_1, \quad \begin{array}{l} X_k \text{ is class mark of the last class} \\ X_1 \text{ is class mark of the first class} \end{array}$$

Properties of range

- It is affected by extreme values
- It does not take into account all observations
- It is easy to calculate and simple to understand

Example:

- Find the relative range of the above two distributions.
- If the range and relative range of a series are 4 and 0.25 respectively. Then what is the value of: a) Smallest observation b) Largest observation

The Quartile Deviation (Semi-inter quartile range), Q.D

- The inter quartile range is the difference between the third and the first quartiles of a set of items and semi-inter quartile range is half of the inter quartile range.

$$Q.D = \frac{Q_3 - Q_1}{2}$$

- **Coefficient of Quartile Deviation (C.Q.D)**

$$C.Q.D = \frac{(Q_3 - Q_1)/2}{(Q_3 + Q_1)/2} = \frac{2 * Q.D}{Q_3 + Q_1} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

- **Example:** Compute Q.D and its coefficient for the following distribution.

Values	Freq.
140- 150	17
150- 160	29
160- 170	42
170- 180	72
180- 190	84
190- 200	107
200- 210	49
210- 220	34
220- 230	31
230- 240	16
240- 250	12

Solutions:

$$Q_1 = 174.90, \quad Q_2 = 190.23, \quad Q_3 = 203.83$$

$$\Rightarrow Q.D = \frac{Q_3 - Q_1}{2} = \frac{203.83 - 174.90}{2} = 14.47$$

$$C.Q.D = \frac{2 * Q.D}{Q_3 + Q_1} = \frac{2 * 14.47}{203.83 + 174.90} = 0.07\%$$

Q.D or C.Q.D includes only the middle 50% of the observation.

The mean, median and mode deviation

- MD – is the average of the absolute deviations taken from a central value, generally the mean, median and mode

Mean Deviation about the mean

$$M.D(\bar{X}) = \frac{\sum_{i=1}^n |X_i - \bar{X}|}{n}$$

For the case of frequency distribution it is given as:

$$M.D(\bar{X}) = \frac{\sum_{i=1}^k f_i |X_i - \bar{X}|}{n}$$

Mean Deviation about the median.

$$M.D(\tilde{X}) = \frac{\sum_{i=1}^n |X_i - \tilde{X}|}{n}$$

- For the case of frequency distribution, it is given as:

$$M.D(\tilde{X}) = \frac{\sum_{i=1}^k f_i |X_i - \tilde{X}|}{n}$$

- **Mean Deviation about the mode.**

$$\mathbf{M.D}(\hat{\mathbf{X}}) = \frac{\sum_{i=1}^n |\mathbf{x}_i - \hat{\mathbf{x}}|}{\mathbf{n}}$$

- For the case of frequency distribution it is given as:

$$M.D(\hat{X}) = \frac{\sum_{i=1}^k f_i |X_i - \hat{X}|}{n}$$

Examples:

1. The following are the number of visit made by ten mothers to the local doctor's surgery. 8, 6, 5, 5, 7, 4, 5, 9, 7, 4

Find mean deviation about mean, median and mode.

Solutions:

First calculate the three averages

$$\bar{X} = 6, X = 5.5, X = 5$$

Xi	4	4	5	5	5	6	7	7	8	9	total
$ X_i - 6 $											
$ X_i - 5.5 $	2	2	1	1	1	0	1	1	2	3	14
$ X_i - 5 $	1.5	1.5	0.5	0.5	0.5	0.5	1.5	1.5	2.5	3.5	14
	1	1	0	0	0	1	2	2	3	4	14

$$\Rightarrow M.D(\bar{X}) = \frac{\sum_{i=1}^{10} |X_i - 6|}{10} = \frac{14}{10} = 1.4$$

$$M.D(\tilde{X}) = \frac{\sum_{i=1}^{10} |X_i - 5.5|}{10} = \frac{14}{10} = 1.4$$

$$M.D(\hat{X}) = \frac{\sum_{i=1}^{10} |X_i - 5|}{10} = \frac{14}{10} = 1.4$$

Example

2. Calculate the mean deviation about the mean and median for the students score in a test: 6,7,7,10,10.

- From the mean is 1.6; and from the median is 1.4.

Properties of mean deviation

- simple to understand and compare as compared to standard deviation
- less affected by extreme values than standard deviation

Example 3. Find mean deviation about mean, median and mode for the following distributions.

Class	Frequency
40-44	7
45-49	10
50-54	22
55-59	15
60-64	12
65-69	6
70-74	3

Coefficient of Mean Deviation (C.M.D)

$$C.M.D = \frac{M.D}{\text{Average about which deviations are taken}}$$

$$\Rightarrow C.M.D(\bar{X}) = \frac{M.D(\bar{X})}{\bar{X}}$$

$$C.M.D(\tilde{X}) = \frac{M.D(\tilde{X})}{\tilde{X}}$$

$$C.M.D(\hat{X}) = \frac{M.D(\hat{X})}{\hat{X}}$$

Example: calculate the C.M.D about the mean, median and mode for the data in example 1 above.

$$\Rightarrow C.M.D(\bar{X}) = \frac{M.D(\bar{X})}{\bar{X}} = \frac{1.4}{6} = 0.233$$

$$C.M.D(\tilde{X}) = \frac{M.D(\tilde{X})}{\tilde{X}} = \frac{1.4}{5.5} = 0.255$$

$$C.M.D(\hat{X}) = \frac{M.D(\hat{X})}{\hat{X}} = \frac{1.4}{5} = 0.28$$

■ The Variance

In order to have a more meaningful statistic to measure the variability, statistics use measures called variance and standard deviation

Variance and Standard Deviation

Population Variance

$$\begin{aligned}\sigma^2 &= \frac{\sum_{i=1}^N (x - \mu)^2}{N} \\ &= \frac{\sum_{i=1}^N x^2 - \frac{\left(\sum_{i=1}^N x\right)^2}{N}}{N} \\ \sigma &= \sqrt{\sigma^2}\end{aligned}$$

Sample Variance

$$\begin{aligned}s^2 &= \frac{\sum_{i=1}^n (x - \bar{x})^2}{(n - 1)} \\ &= \frac{\sum_{i=1}^n x^2 - \frac{\left(\sum_{i=1}^n x\right)^2}{n}}{(n - 1)} \\ s &= \sqrt{s^2}\end{aligned}$$

Variance and Standard Deviation

For the case of frequency distribution it is expressed as

Population Variance

Sample

$$\sigma^2 = \frac{1}{N} \sum f_i (X_i - \mu)^2, \quad i = 1, 2, \dots, k$$

$$S^2 = \frac{1}{n-1} \sum f_i (X_i - \bar{X})^2, \quad i = 1, 2, \dots, k$$

$$S^2 = \frac{\sum_{i=1}^k f_i X_i^2 - n\bar{X}^2}{n-1}$$

Variance – is average of the squared deviations from the mean

Standard deviation - is the square root of the variance

The following steps are used to calculate the sample variance:

1. Find the arithmetic mean.
2. Find the difference between each observation and the mean.
3. Square these differences.
4. Sum the squared differences.
5. Since the data is a sample, divide the number (from step 4 above) by the number of observations minus one, i.e., $n-1$ (where n is equal to the number of observations in the data set).

Example: The height of nine students was measured in inches and the data is presented below.

Height(x): 69 66 67 69 64 63 65 68 72

Calculate the population variance and standard deviations.

Variance = 7.11 inch²; S.D = 2.66 inch

- The variance and Standard deviation of a grouped data is calculated by using the Class Mark.

Examples: Find the variance and standard deviation of the following sample data

1. 5, 17, 12, 10.
2. The data is given in the form of frequency distribution.

Class Frequency

40-44	7
45-49	10
50-54	22
55-59	15
60-64	12
65-69	6
70-74	3

Properties of Variance and Standard deviation

- The unit of measurement of the variance is the square of the unit of measurement of the observed values.
- It is based on all observations
- Standard deviation is considered to be the best measure of dispersion and is used widely
- If the variance/standard deviation is large, the data is more dispersed
- The variance and standard deviation are used quite often in inferential statistics

Coefficient of variation (CV)

- it is the corresponding relative measure of standard deviation
- It is used to compare the variability of two or more different groups

$CV = \frac{s}{\bar{x}} \times 100\%$
 Less coefficient of variation – is said to be less variable or more consistent or more uniform or more homogeneous.

Example: The students of Biology and Chemistry took *Stat 273* course. The following information was recorded

Departments	Mean	S.D
Biology	79	23
Chemistry	64	11

CV (Biology) = 29.11%

CV (Chemistry) = 17.2%

➤ Chemistry students are more homogenous.

The Standard Score (Z - score)

- is a measure that describes the relative position of a single value in the entire distribution of values
- It also gives the number of standard deviations a particular observation lie above or below the mean

- If Z is +ve - the observation lies above the mean
- If Z is -ve - the observation lies below the mean
- If Z is 0 - the observation is equal to the mean

$$Z = \frac{X - \bar{X}}{S}$$

Example: Two sections were given an exam in certain course. The average score was 72 with standard deviation of 6 for *section 1* and 85 with standard deviation of 5 for *section 2*. Student A from *section 1* scored 84 and student B from *section 2* scored 90. Who performed better relative to his/her group?

Z-score for student A = 2.00

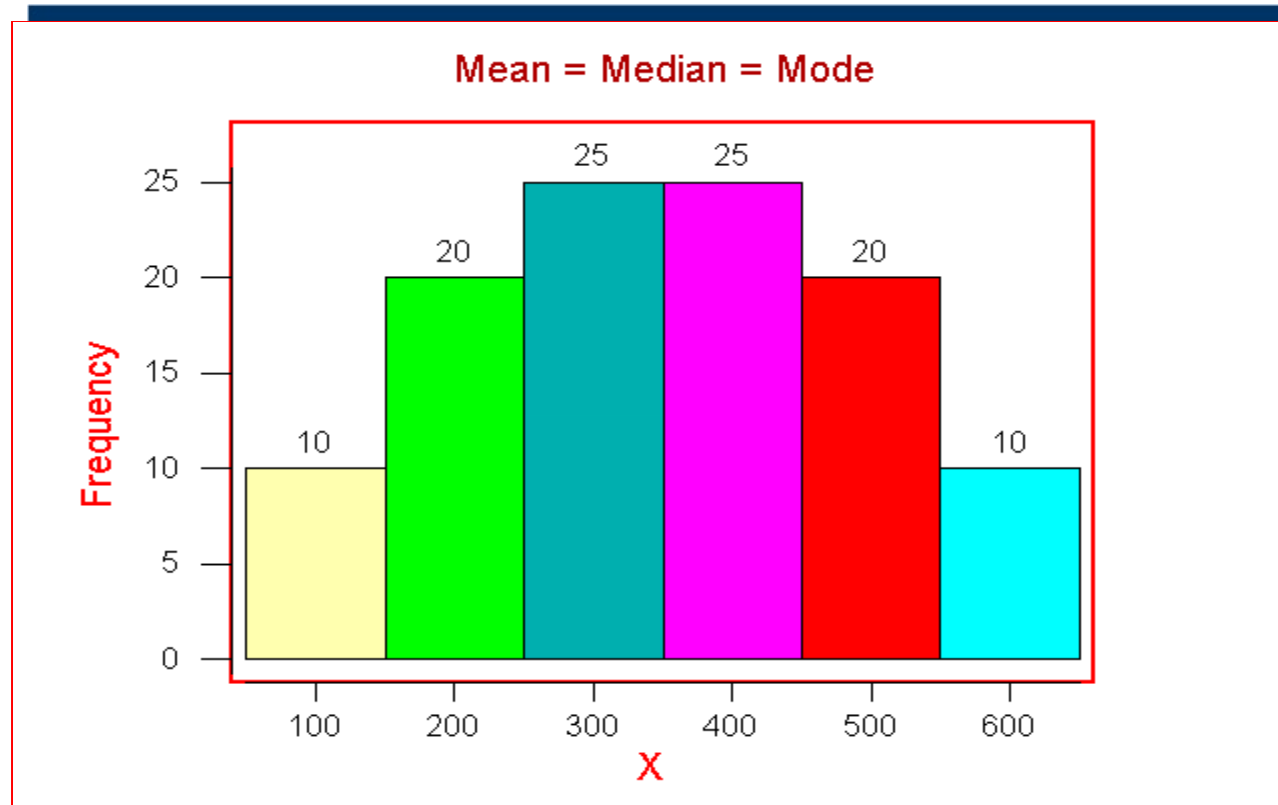
Z-score for student B = 1.00

Measure of Shape

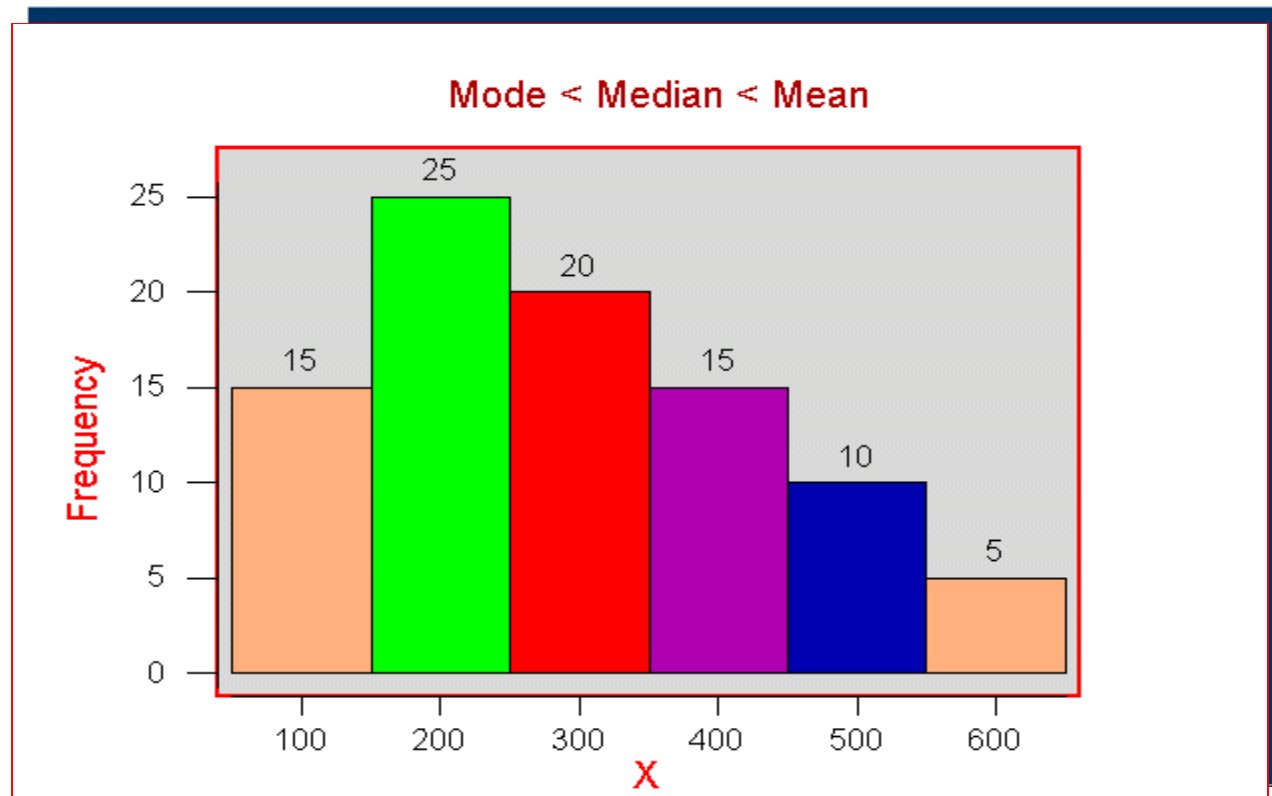
- **Skewness**

- Measure of asymmetry of a frequency distribution
 - Symmetric or unskewed
 - Skewed to right
 - Skewed to left

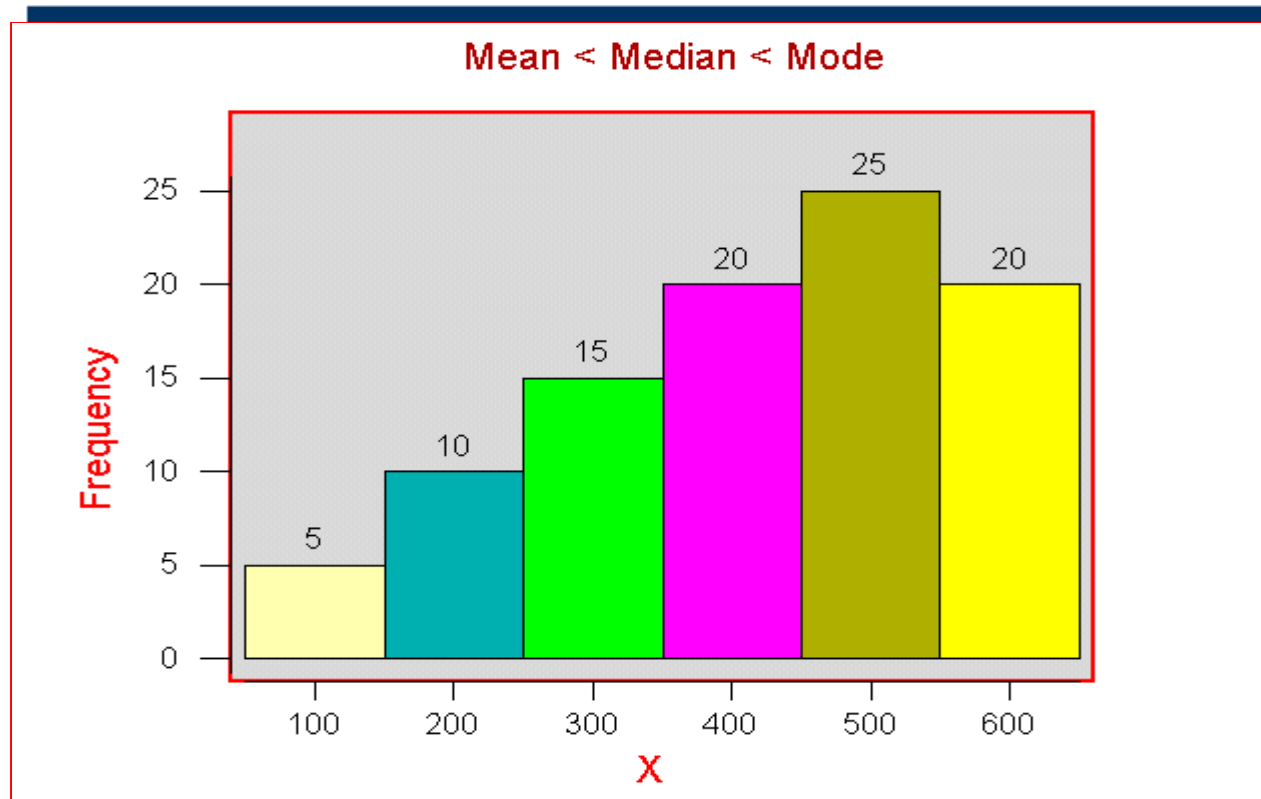
Symmetric



Skewed to right



Skewed to left



Measures of skewness

- Pearson coefficient of skewness (P_{csk})

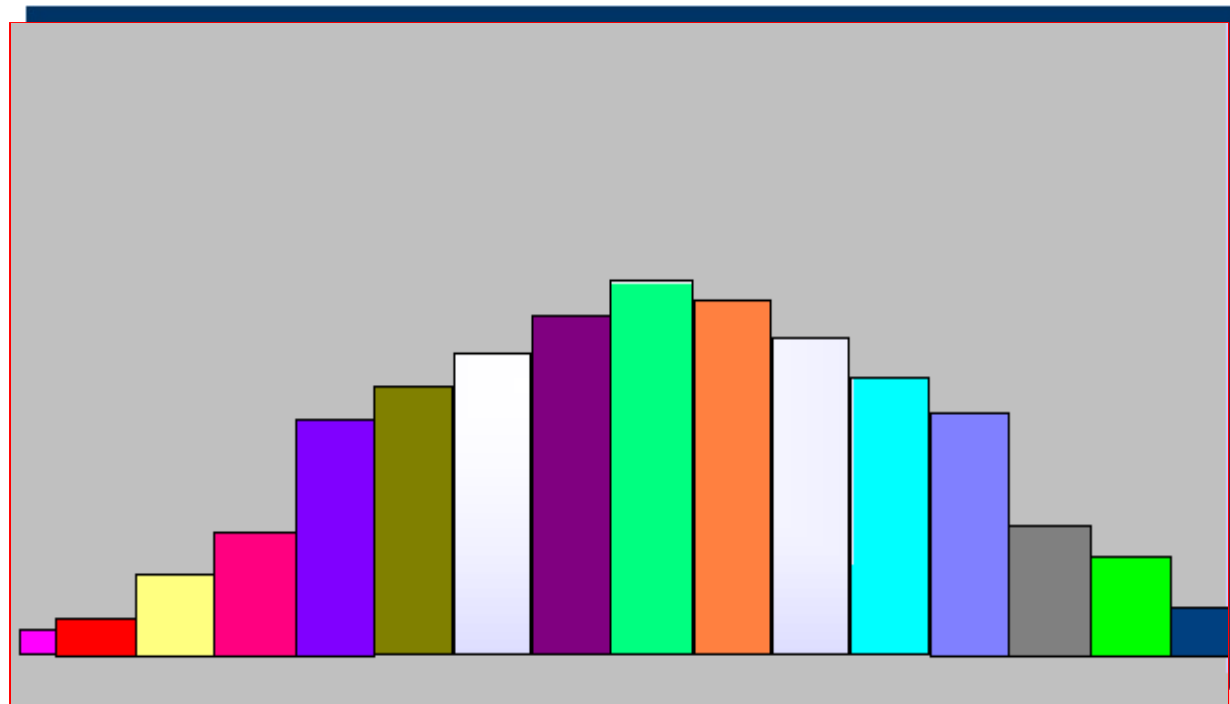
$$P_{\text{csk}} = \frac{\bar{x} - \hat{x}}{s} = \frac{3(\bar{x} - \tilde{x})}{s}$$

- If $P_{\text{csk}} = 0$ - the distribution is symmetric
- If $P_{\text{csk}} < 0$ - the distribution is negatively skewed
- If $P_{\text{csk}} > 0$ - the distribution is positively skewed

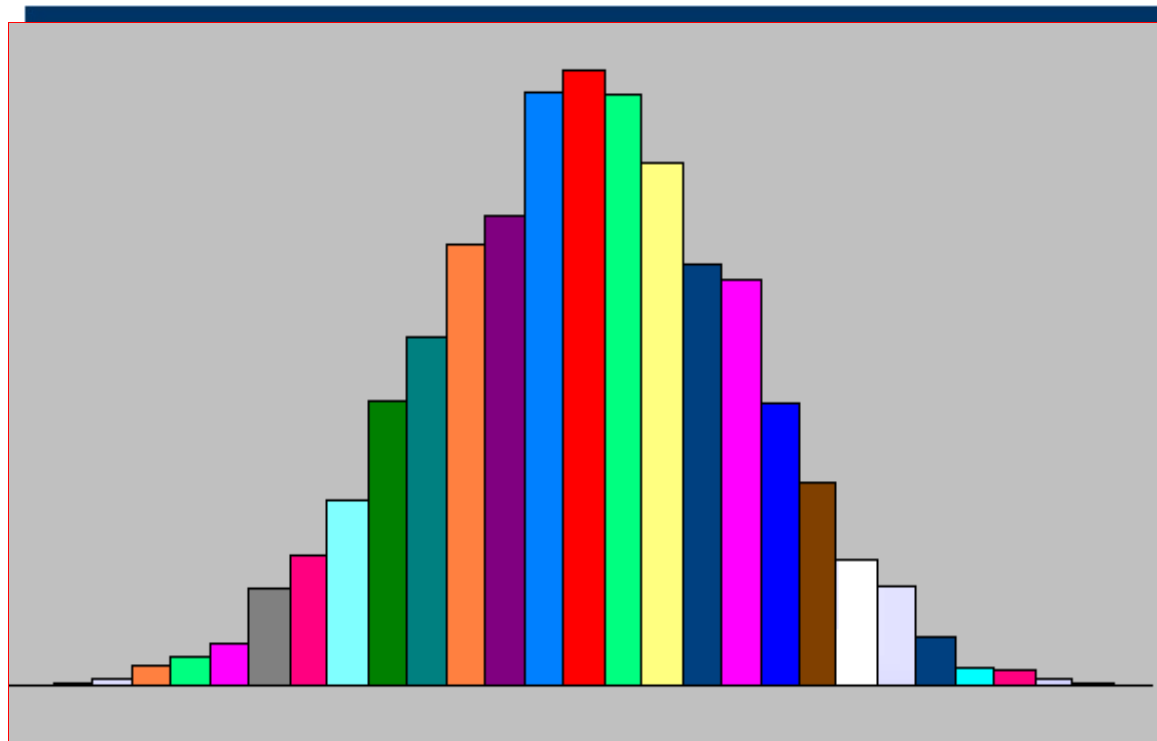
- **Kurtosis**

- Measure of flatness or peakedness of a frequency distribution
 - **Platykurtic** (relatively flat)
 - **Mesokurtic** (normal)
 - **Leptokurtic** (relatively peaked)

Platykurtic - flat distribution



Mesokurtic - not too flat and not too peaked



Leptokurtic - peaked distribution

