Measure of Central Tendency

(Measure of Location)

F. M. ARIFUR RAHMAN

SENIOR LECTURER, DEPARTMENT OF MATHEMATICAL & PHYSICAL SCIENCES

Content

- Measure of central tendency
  - Mean
  - Median
  - Mode
- Calculation of the measures using raw (ungrouped) and grouped data
- Median like measures- Quartiles, Deciles, Percentiles

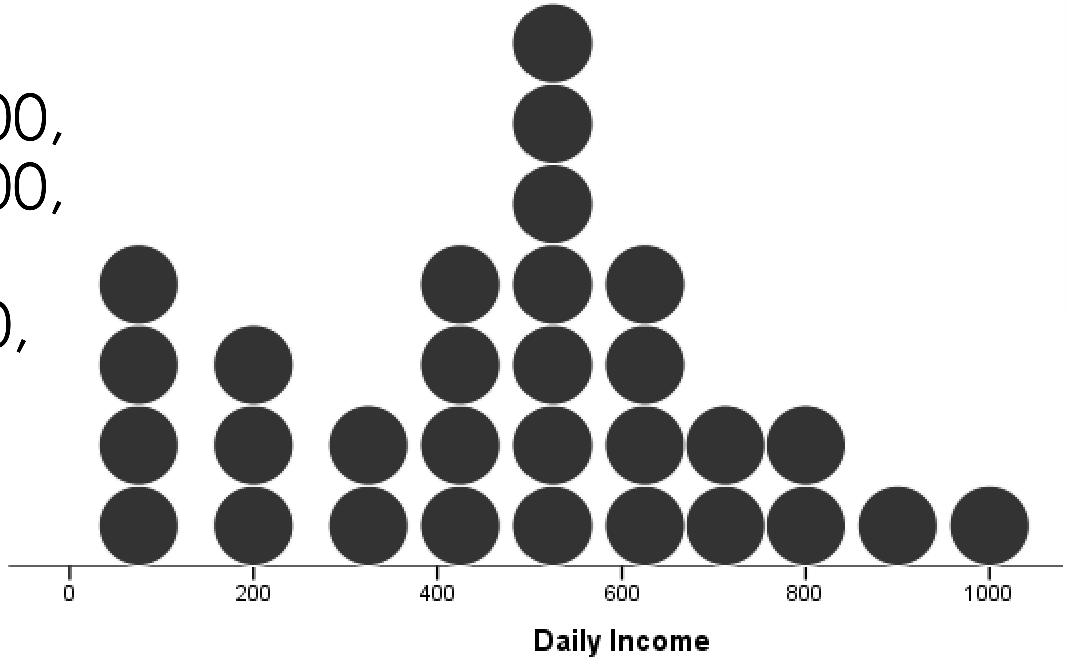
Measures of central tendency

- In every data set, the data have the tendency to occur mostly in a central location.
- The measures used to find and describe those locations, are collectively known as measures of central tendency

Measures of central tendency

Daily income of 30 respondents-

50, 100, 500, 1000, 400, 100, 200, 500, 200, 500, 800, 900, 700, 500, 600, 450, 600, 500,, 450, 400, 350, 650, 300, 200, 800, 700, 50, 550, 600, 500



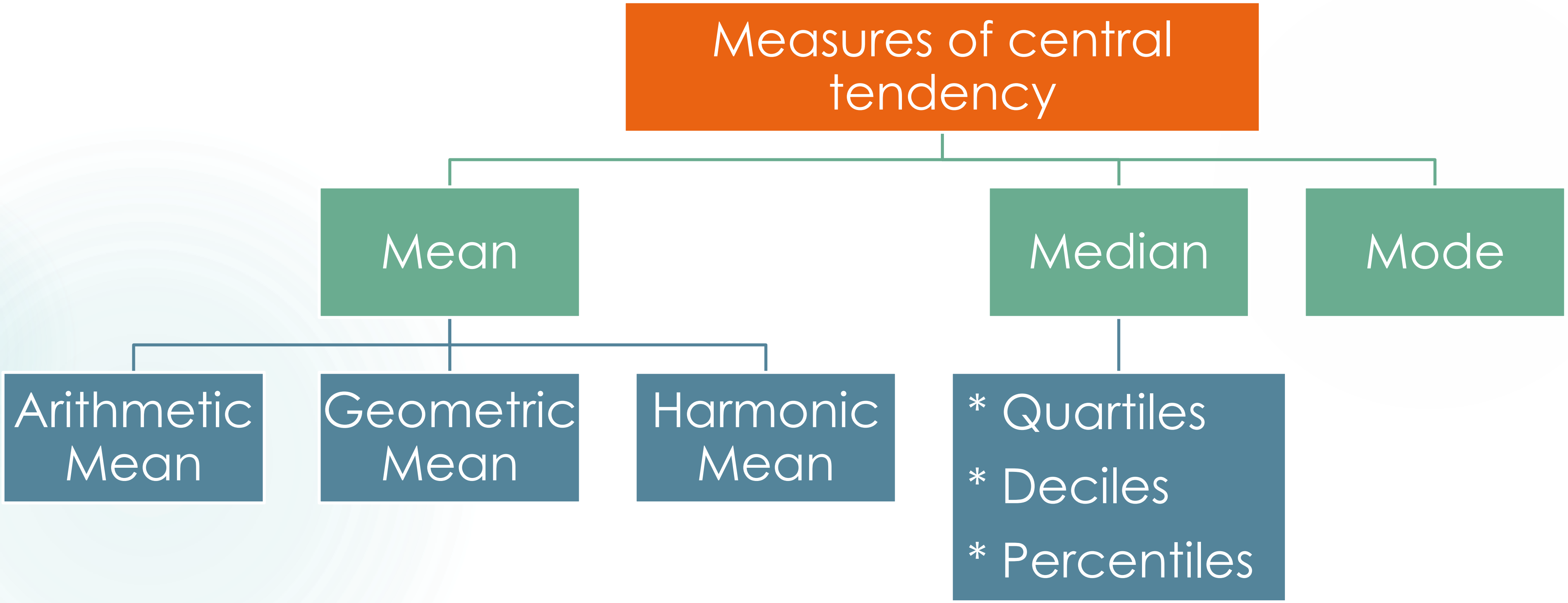
Measures of central tendency

- Measure of central tendency provide a very convenient way of describing a set of scores with a **single number** that describe the **performance of the group**

- It is also defined as a single value that is used to describe the **center of the data**.

Source: Slideshare.net

Measures of central tendency



Mean (Arithmetic Mean)

It is simply the sum of the **numbers divided by the number of numbers** in a set of data. This is also known as average.

For example, consider the values-

5, 3, 9, 2, 7, 5, 8

Mean (Arithmetic Mean)

It is simply the sum of the **numbers divided by the number of numbers** in a set of data. This is also known as average.

For example, consider the values-

5, 3, 9, 2, 7, 5, 8

$$Mean = \frac{5+3+9+2+7+5+8}{7}$$

$$= \frac{39}{7} = 5.57\ unit$$

Mean (Arithmetic Mean)

Formulas:

For raw or ungrouped data-

For Population: let, $X_1, X_2, \ldots, X_N$ are values of a variable from a population of size N. Then,

$$Population\ mean, \mu = $$

$$= \frac{X_1 + X_2 + \ \ldots + X_N}{N} = \frac{\sum_{i=1}^{N} X_i}{N}$$

(Parameter)

For Sample: let, $x_1, x_2, \ldots, x_n$ are values of a variable from a sample of size n. Then,

$$Sample\ mean, \bar{x} = $$

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

(Statistic)

Mean (Arithmetic Mean)

Note:

- Mean cannot be calculated for Nominal & Ordinal level of data
- Mean is easily affected by extreme values
- $\sum_i (x_i - \bar{x}) = 0$

Median

Median is the number present in the **middle** when the numbers in a set of data are **arranged in ascending or descending order**. If the number of numbers in a data set is **even**, then the median is the **mean of the two middle numbers**.

For example, consider the values-

5, 3, 9, 2, 7, 5, 8

Median

Organizing in **ascending** order,

2, 3, 5, 5, 7, 8, 9

Median

Organizing in **ascending** order,

$$\underbrace{2,\ 3,\ 5}_{\text{50\% observations}},\ 5,\ \underbrace{7,\ 8,\ 9}_{\text{50\% observations}}$$

Here, n= 7 (an odd number)

Median = 5

Median

Consider – 3, 5, 5 , 7, 8, 9

Median

Consider – 3, 5, 5 , 7, 8, 9

Organizing in **ascending** order,

$$\underbrace{3,\ \ 5,\ \ 5,}_{\text{50\% observations}} \mid \underbrace{7,\ \ 8,\ \ 9}_{\text{50\% observations}}$$

Median

Consider – 3, 5, 5 , 7, 8, 9

Organizing in **ascending** order,

$$\underbrace{3,\ 5,\ 5,}_{\text{50\% observations}}\ \Big|\ \underbrace{7,\ 8,\ 9}_{\text{50\% observations}}$$

Here, n= 6 (an even number)

$$Median = \frac{3rd\ value + 4th\ value}{2}$$

$$= \frac{5+7}{2} = \frac{12}{2} = 6$$

Median

Formulas:

For raw or ungrouped data- (Sample)

For Odd n	For Even n
$Median = \left(\frac{n+1}{2}\right)^{th} value$	$Median = \frac{1}{2}\left[\left(\frac{n}{2}\right)^{th} value + \left(\frac{n}{2}+1\right)^{th} value\right]$

Median

Note:

- To find median, data has to be at least in ordinal level of measurement

- Median is not affected by extreme values

Mode

Mode is the value that occurs most frequently in a set of data

Notes: Mode can be computed for all levels of data.

For example, consider the values-

5, 3, 9, 2, 7, 5, 8

Mode

consider the values-

5, 3, 9, 2, 7, 5, 8

Mode

consider the values-

5, 3, 9, 2, 7, 5, 8

Value **5** occurred maximum 2 times

Mode = 5

Mode

Formulas:

For raw or ungrouped data-

Find the value occurred most of the times in the data

Example (raw data)

Weekly income of 6 respondents (in taka)-

2500, 3900, 3500, 5000, 4000, 3500

Find Mean, Median and Mode. Interpret the results.

Class task (raw data)

Weekly income of 6 respondents (in taka)-

2500, 3900, 3500, 5000, 4000, 3500

$$\boldsymbol{Mean, \bar{x}} = \frac{2500 + 3900 + 3500 + 5000 + 4000 + 3500}{6} = \frac{22400}{6}$$

$$= 3733.33 \approx 3734\, taka$$

Interpretation: Average weekly income of the respondents is 3734 taka

Example (raw data)

Weekly income of 6 respondents (in taka)-

2500, 3900, 3500, 5000, 4000, 3500

Organizing the values in ascending order-

2500, 3500, 3500, 3900, 4000, 5000

Example (raw data)

$$\boldsymbol{Median} = \frac{1}{2}\left[\left(\frac{n}{2}\right)^{th} value + \left(\frac{n}{2}+1\right)^{th} value\right]$$

$$= \frac{1}{2}[3^{rd} value + 4^{th} value] = \frac{1}{2}[3500 + 3900] = 3700\ taka$$

Interpretation: 50% of the respondents have weekly income less than or equal to 3700 taka, and 50% of the respondents have weekly income higher than or equal to 3700 taka.

Example (raw data)

Weekly income of 6 respondents (in taka)-

2500, 3900, 3500, 5000, 4000, 3500

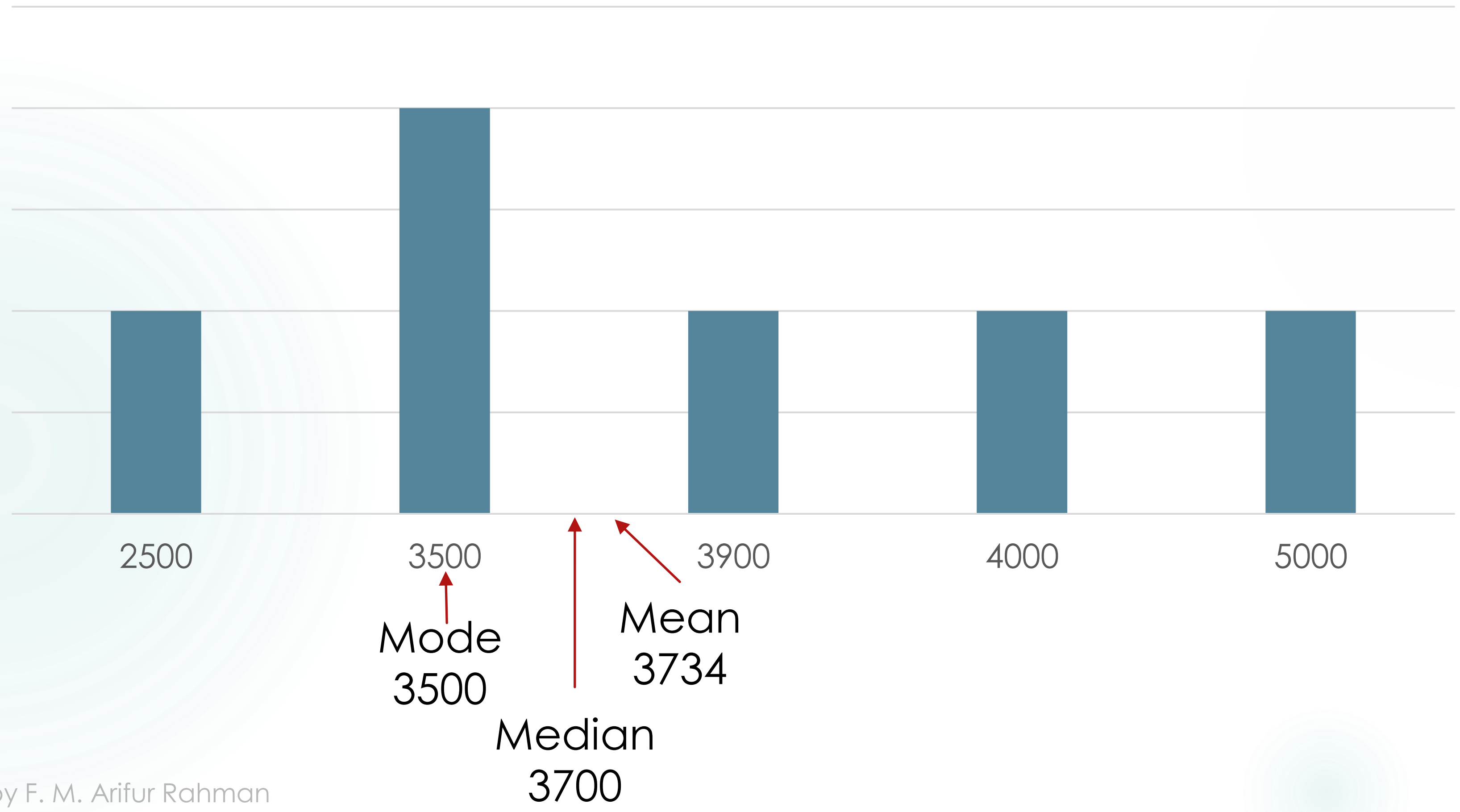
Value 3500 occurs highest 2 times in the data.

So, **Mode** = 3500

Interpretation: In the data, weekly income 3500 taka is occurred highest number of times.

Example (raw data)

In graphs-



Mean (Arithmetic Mean)

Formulas:

<u>For grouped data- </u>(Population)

X_i	f_i
X_1	f_1
X_2	f_2
…	…
X_K	f_K

let, $\mathbf{X_1}$, $\mathbf{X_2}$, **…,** $\mathbf{X_K}$ are values of a variable from a population of size N and they occurred $\mathbf{f_1}$, $\mathbf{f_2}$, **…,** $\mathbf{f_K}$ times respectively. Then,

$$Population\ mean, \mu = \frac{\sum_{i=1}^{K} f_i X_i}{N}$$

Mean (Arithmetic Mean)

Formulas:

<u>For grouped data- </u>(Sample)

x_i	f_i
x_1	f_1
x_2	f_2
…	…
x_k	f_k

let, $\mathbf{x_1, x_2, \ldots, x_k}$ are values of a variable from a sample of size n and they occurred $\mathbf{f_1, f_2, \ldots, f_k}$ times respectively. Then,

$$Sample\ mean, \bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{n}$$

Median

Formulas:

<u>For grouped data- </u>(Sample)

$$Median = L_m + \frac{\frac{n}{2} - F}{f_m} * c$$

Where,

L_m= Lower class limit of the median class

n= Total frequency

F= Cumulative frequency of the pre-median class

f_m= frequency of the median class

c= class interval

Mode

Formulas:

<u>For grouped data-</u>

$$Mode = L_o + \frac{\Delta_1}{\Delta_1 + \Delta_2} * c$$

Mode

Where,

L_o = Lower class limit of the modal class

Δ_1 = Excess of modal frequency over frequency of the next lower class (Pre-modal class) = difference between the frequencies of the modal class and pre-modal class.

Δ_2 = Excess of modal frequency over frequency of the next higher class (Post-modal class) = difference between the frequencies of the modal class and post-modal class.

c= class interval

Example (grouped data)

Monthly income ('000 tk)	No. of respondents
5-30	7
30-55	10
55-80	6
80-105	4
105-130	3

Find Mean, Median and Mode. Interpret the results.

Example (grouped data)

Monthly income ('000 tk)	No. of respondents (f_i)	Class Midpoint (x_i)	$f_i x_i$
5-30	7	17.5	122.5
30-55	10	42.5	425
55-80	6	67.5	405
80-105	4	92.5	370
105-130	3	117.5	352.5
Total	**30**		**1675**

$$Mean, \bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{n} = \frac{1675}{30} = 55.83\ (thousand\ taka)$$

Example (grouped data)

Monthly income ('000 tk)	No. of respondents (f_i)	Cumulative frequency
5-30	7	7
30-55	10	17
55-80	6	23
80-105	4	27
105-130	3	30
Total	**30**	

50% of the respondents have monthly family income less than or equal to 50,000 taka and 50% of the respondents have monthly family income higher than or equal to 50,000 taka

$$Median = L_m + \frac{\frac{n}{2} - F}{f_m} * c = 30 + \frac{15 - 7}{10} * 25 = 50 \; thousand \; taka$$

Example (grouped data)

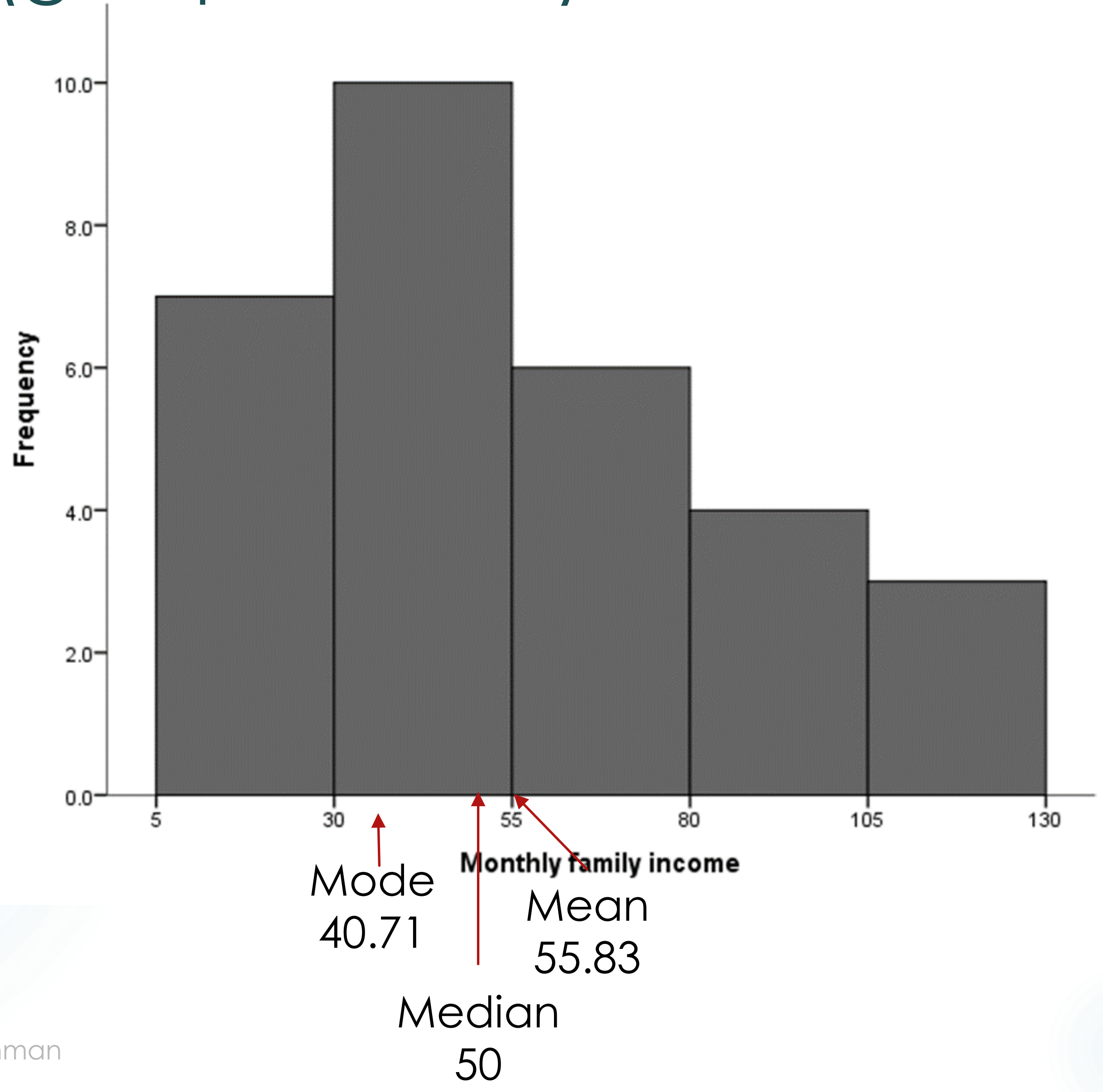
Monthly income ('000 tk)	No. of respondents (f_i)
5-30	7
30-55	10
55-80	6
80-105	4
105-130	3
Total	**30**

Comparatively a higher number of the respondents have monthly family income around 40,000 taka

$$Mode = L_o + \frac{\Delta_1}{\Delta_1 + \Delta_2} * c = 30 + \frac{3}{3+4} * 25 = 40.71\ thousand\ taka$$

Example (grouped data)

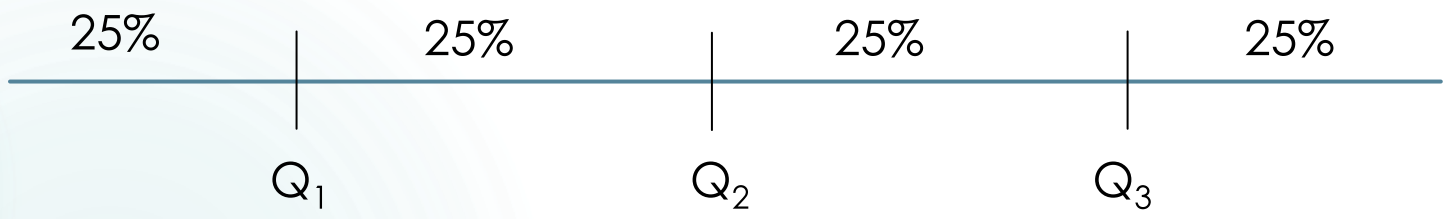
In graphs-



Median like measures

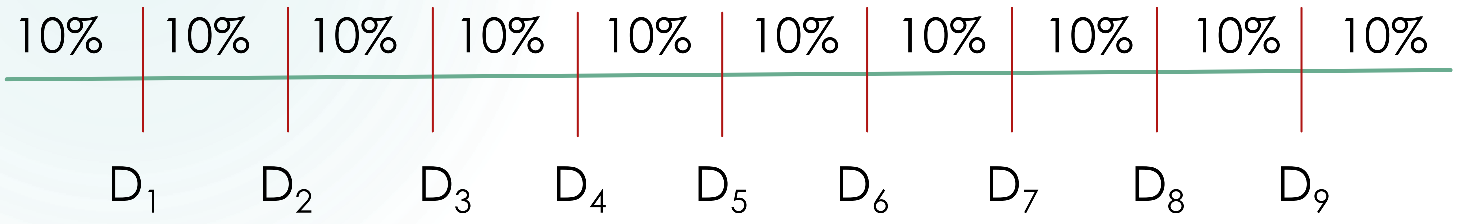
Quartiles:

Quartiles Divides the whole distribution into 4 equal parts



Deciles:

Deciles divide the whole distribution into 10 equal parts



Median like measures

Percentiles:

Percentiles divide the whole distribution into 100 equal parts

Median like measures

Formulas for finding percentiles:

For raw data,

- If $\frac{in}{100}$ is an integer- (for i= 1, 2, 3, …, 99)

$$P_i = \frac{1}{2}\left[\left(\frac{in}{100}\right)^{th} value + \left(\frac{in}{100} + 1\right)^{th} value\right]$$

- If $\frac{in}{100}$ is not an integer-

$$P_i = next\ integer^{th}\ value\ of\ \frac{in}{100}$$

Median like measures

Example:

Weekly income of 6 respondents (in taka)-

2500, 3900, 3500, 5000, 4000, 3500

Find Q1, Q2, Q3.

Median like measures

Example:

Organizing in ascending order- 2500, 3500, 3500, 3900, 4000, 5000

So, the quartiles-

$$Q_1 = P_{25} = \ next\ integer^{th} value\ of\ \frac{25n}{100} = next\ integer^{th} value\ of\ \frac{150}{100}$$
$$= next\ integer^{th} value\ of\ 1.5 = 2^{nd}\ value = 3500$$

Median like measures

Example:

$$Q_2 = P_{50} = Median = \frac{1}{2}\left[\left(\frac{n}{2}\right)^{th} value + \left(\frac{n}{2}+1\right)^{th} value\right] = 3700$$

$$Q_3 = P_{75} = next\ integer^{th} value\ of\ \frac{75n}{100} = next\ integer^{th} value\ of\ \frac{450}{100}$$
$$= next\ integer^{th} value\ of\ 4.5 = 5^{th}\ value = 4000$$

Median like measures

Formulas for finding percentiles:

For grouped data-

$$P_i = L_i + \frac{\frac{in}{100} - F_i}{f_i} * c \qquad i = 1, 2, 3, \ldots, 99$$

L_i= Lower class limit of the i^{th} percentile class

n= Total frequency

F_i= Cumulative frequency of the i^{th} pre-percentile class

f_i= frequency of the i^{th} percentile class

c= class interval

Geometric Mean

The geometric mean G of n positive rates x1, x2,…, xn is defined as the nth positive root of the product of the rates. Symbolically,

$$G = (x_1 x_2 x_3 \dots x_n)^{\frac{1}{n}} - 1$$

Where x_i's are rate of changes.

If x_i's are the value for a given time t, then,

$$G = \sqrt[n]{\frac{x_n}{x_0}} - 1$$

G provides average rates of change.

Geometric Mean

Uses:

- Geometric mean is useful when the data is in geometric progression

Example: At year 2000, the profit was 10000.

Year	Profit
2001	12000
2002	15000
2003	20000
2004	18000
2005	22000
2006	27000

Geometric Mean

Example: At year 2000, the profit was 10000.

Year	Profit	Increase rate	Increased rate (x)
2001	12000	0.20	1.20
2002	15000	0.25	1.25
2003	20000	0.33	1.33
2004	18000	-0.10	0.90
2005	22000	0.22	1.22
2006	27000	0.23	1.23

$$G = (x_1 x_2 x_3 \dots x_n)^{\frac{1}{n}} - 1 = (1.20 * 1.25 * 1.33 \ * 0.90 * 1.22 \ * 1.23)^{\frac{1}{6}} - 1$$
$$= 1.18 - 1 = \ 0.18 = 18\% \ increase\ per\ year.$$

Or,

$$G = \left(\frac{y_{2006}}{y_{2000}}\right)^{\frac{1}{6}} - 1 \ = \left(\frac{27000}{10000}\right)^{\frac{1}{6}} = 0.18 = 18\% \ increase\ per\ year$$

Harmonic Mean

Harmonic mean is defined as the reciprocal of the arithmetic mean of the reciprocals of the individual values.

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}$$

For grouped data,

$$H = \frac{\sum f_i}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \ldots + \frac{f_k}{x_k}} = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \ldots + \frac{f_k}{x_k}}$$

Harmonic Mean

Uses:

When dealing with rates, harmonic mean is more appropriate.

Example:

Suppose you are travelling to Narayanganj to Dhaka by car. The distance between Dhaka and Naraynganj is 18km. You drive your car for the first 6km at a speed of 30km per hour and the second 5km at a rate of 40km per hour and the remaining 7 kilometers at a speed of 20km per hour. What is the average speed with which you traveled from Dhaka to Narayanganj?