Central Tendency

Numerical Measures of Data

- Data sets are usually either a sample or to a population

- Our target: ultimately to use numerical descriptive measures to make inferences

- Most methods measure one of two data characteristics

- **Central tendency** - This measures the extent to which all the values are **grouped around** a typical or central value.

- **Variation or Dispersion -** This measures the amount of **dispersion or scattering** of values **away** from a central value.

Measures of Central Tendency

- In most datasets i.e. population or sample i.e. the values show a distinct tendency to **group or cluster** around a central point.

- This tendency of clustering the values around the center of the series is usually called central tendency.

- The numerical measure of this tendency of concentration is variously known as
  - The measure of central tendency
  - The measure of location
  - The measure of average.

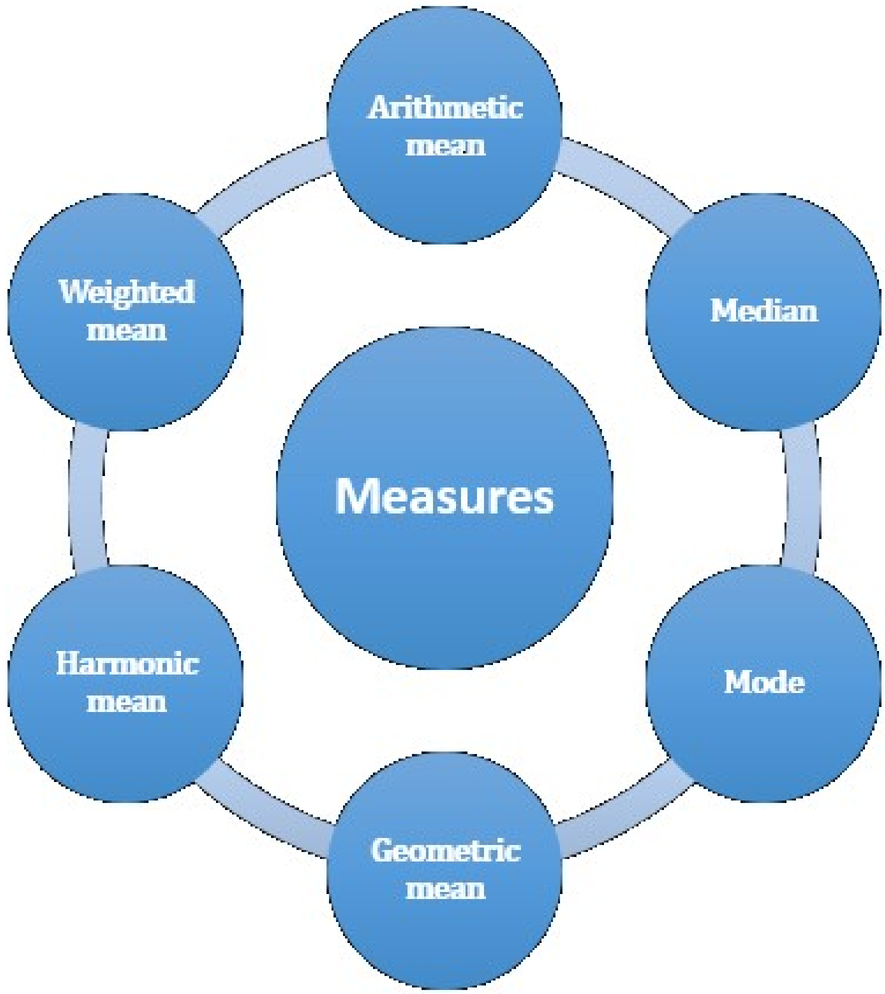
Necessity of measuring the central tendency

I. They give us an idea about the concentration of the values in the central part of the distribution.
II. It is the value of the variable, which is typical of the whole set.
III. It represents all relevant information contained in the data in as few numbers as possible.
IV. They give precise information, not information of a vague general type.

Characteristics of a good measure of central tendency

i. It should be easy to understand.
ii. It should be easy to calculate.
iii. It should be based upon all observations.
iv. It should be rigidly defined.
v. It should not be unduly affected by extreme values.
vi. It should be suitable for further algebraic treatment.
vii. It should be less affected by sampling fluctuation.

Different measures of central tendency



Arithmetic Mean

Arithmetic Mean

- We can obtain the arithmetic mean of a series of observations by adding the values of the observations and then dividing the sum by the number of observations.

- Arithmetic mean (AM) for
    - Sample observation is denoted by $\bar{x}$
    - Population mean is denoted by μ

- If there are n values $x_1, x_2, \ldots, \ldots, \ldots, x_n$ for a variable X, then the AM (denoted by $\bar{x}$) is defined as

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + \ldots + \ldots + x_n}{n}; (i = 1, 2, \ldots, \ldots, n)$$

Example

Banglatel is studying the number of minutes used by clients in a particular cell phone rate plan. A random sample of 12 clients showed the following number of minutes used last month.

90	77	94	89	119	112
91	110	92	100	113	83

What is the mean (arithmetic mean) number of minutes used?

Solution

Average use of the rate plan

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \ldots + \ldots + \ldots + x_n}{n} = \frac{90 + 77 + \ldots + \ldots + 112 + 91 + \ldots + \ldots + 113 + 83}{12} = 97.5$$

Thus the arithmetic mean number of minutes used last month by the sample of cell phone users is 97.5 minutes.

Group Data With Frequencies

For a group data as given in the following table

Values:	x_1	x_2	…	…	x_n
Frequencies :	f_1	f_2	…	…	f_n

Such that $f_1 + f_2 + f_3 + \ldots + \ldots + \ldots + f_k = n$, then the AM (denoted by $\bar{x}$) is defined as

$$\overline{x} = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \ldots + f_k x_k}{n}; (i = 1, 2, \ldots, \ldots, k)$$

Example

Calculate the mean for the following frequency distribution for n=100:

Class Interval	Frequency
0-10	10
10-20	20
20-30	40
30-40	20
40-50	10

Solution

Class Interval	Frequency(f_i)	Mid values (x_i)	$(f_i)*(x_i)$
0-10	10	5	50
10-20	20	15	300
20-30	40	25	1000
30-40	20	35	700
40-50	10	45	450
Total	100		2500

Arithmetic mean,

$$\overline{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + f_4x_4 + f_5x_5}{n}$$

$$= \frac{50+300+1000+700+450}{100} = 25$$

Test Yourself

The following data represent the distribution of the age of employees within two different divisions of publishing company. Determine which company have relatively aged group of employees.

Age of employees	Number of employees of division	
	X	Y
20-30	6	13
30-40	19	30
40-50	9	24
50-60	10	0
60-70	2	4

Solution

Age of employees	Mid values(x_i)	Frequency(f_{xi})	Frequency(f_{yi})	(f_{xi})*(x_i)	(f_{yi})*(x_i)
20-30	25	6	13	150	325
30-40	35	19	30	665	1050
40-50	45	9	24	405	1080
50-60	55	10	0	550	0
60-70	65	2	4	130	260
Total		46	71	1900	2715

Arithmetic mean age of employee division X = $\frac{\sum_{i=1}^{5} f_{xi} * x_i}{\sum_{i=1}^{5} f_{xi}} = \frac{1900}{46} = 41.3$

Arithmetic mean age of employee division Y = $\frac{\sum_{i=1}^{5} f_{yi} * x_i}{\sum_{i=1}^{5} f_{yi}} = \frac{2715}{71} = 38.2$

Since, A.M of X group of employees > A.M of Y group of employees, X group of employees are relatively aged more.

Arithmetic Mean

Merits	Demerits
1. Rigidly defined. 2. Easy to understand and calculate. 3. Based upon all observation. 4. Most amenable to algebraic treatment. 5. Not based on position in the series.	1. Cannot be defined graphically. 2. Cannot be used in case of qualitative data. 3. Affected very much by extreme values. 4. May not occur in the series. 5. Difficult to calculate in the case of the data with open-end class.

When NOT to use Arithmetic Mean

1. In highly - skewed distributions.

2. When the distribution is unevenly spread; concentration being small or large at irregular points.

3. When an average rate of growth or change over a period of time is required.

4. When the observation are from geometric progression.

5. When averaging rates (that is speed, fluctuations in the prices of articles, etc.)

6. When there are very large and very small values of observations.

Median

Median

- If the values of a series are arranged in an ascending or descending order of magnitude, then the **middle most value** in this arrangement is called the median of the series.

- Median is usually denoted by ***Me.***

- The median is generally the best average in open-end grouped distribution, especially where if plotted as a frequency curve one gets a J or reverse J shaped curve

Determination of Median: Ungrouped Data

Let n be the number of observations

a. When ***n is odd the*** value of the $\frac{n+1}{2}th$ observation will be the median.

b. When ***n is even*** the median will be the AM of the values of $\frac{n}{2}th$ and $(\frac{n}{2}+1)th$ observation in the series.

Example: n is odd

The ages of a family of seven members are given as 12, 7, 2, 34, 17, 21 and 19. Find the median age.

Step 1	Count the total number of elements, n=? Here n= 7, an odd number
Step 2	Arrange the values in ascending order : 2, 7, 12, 17, 19, 21, 34
Step 3	Median: Me = Value of $\frac{n+1}{2}$ th observation = Value of $\frac{7+1}{2}$ th observation =value of 4^{th} observation = 17
Step 4	Median Age of the family is 17 years

Example: n is even

The ages of a family of eight members are given as 12, 7, 2, 34, 17, 40, 21 and 19. Find the median age.

Step 1	Count the total number of elements, n=? Here n= 8, an even number
Step 2	Arrange the values in ascending order : 2, 7, 12, 17, 19, 21, 34, 40
Step 3	Median: Me = AM of the values of $\frac{n}{2}th$ and $(\frac{n}{2}+1)th$ observation = AM of the values of $4th$ and $5th$ observation = $\frac{17+19}{2}$ = 18
Step 4	Median Age of the family is 18 years

Determination of Median: Grouped Data

$$\text{Formula, } Me = L_0 + \frac{(\frac{n}{2} - F_{-Me})}{f_{Me}} * W_{Me}$$

- Me = Median
- L_0 = Lower limit of the Median class
- F_{-Me} = Cumulative frequency of the pre median class.
- f_{Me}= Frequency of the median class.
- W_{Me} = Width of the median class.
- n = Total number of observation.

MEDIAN CLASS: the class that contains $\frac{n}{2}th$ observation of the given data.

Example

The Table displays summary information of the parent of 50 students. Compute the median income of the parents.

Income of parent (in thousand taka)	Frequency
Below 20	3
20-40	4
40-60	6
60-80	8
80-100	12
100-120	10
120 and over	7
Total	50

- Step 1: Compute the cumulative frequencies.
- Step 2: Determine $\frac{n}{2}$th value, one half of the total number of cases.
- Step 3: Locate the median class.
- Step 4: Determine the lower limit (L_0) of the median class.
- Step 5: Sum the frequencies of all the classes prior to the median class. This is F_{-Me} .
- Step 6: Determine the frequency of the median class f_{Me}..
- Step 7: Determine the width of the median class, W_{Me} .

Test Yourself

The following data represents the amount (in thousands taka) of loan requirements of the people of two different upazilla. Using median, comment on which upazilla has the greater average demand of loans.

Upazilla 1	42	12	26	18	9	35	28	39	8
Upazilla 2	8	15	10	18	22	20	26	42	35

Solution

Here, n = 9 (odd)

Arranging Upazilla 1 observations in ascending order:

8, 9, 12, 18, 26, 28, 35, 39, 42

Therefore, median of Upazilla 1 = $\frac{9+1}{2}$ th observation = 26

Arranging Upazilla 2 observations in ascending order:

8, 10, 15, 18, 20, 22, 26, 35, 42

Therefore, median of Upazilla 2 = $\frac{9+1}{2}$ th observation = 20

Since, median of Upazilla 1 > median of Upazilla 2, Upazilla 1 has the greater average demand of loans.

Test Yourself

The following table gives the data pertaining to kilowatt hours of electricity consumed by 100 randomly selected flat owners of Japan garden city.

Consumption (in K-watt hours)	0-100	100-200	200-300	300-400	400-500
No. of users	6	25	36	20	13

Calculate

i. Mean consumption of electricity

ii. Median use of electricity

Solution

Consumption (in K-watt hours)	Mid Value(x_i)	No. of users(f_i)	(f_i)*(x_i)	Cumulative Frequency
0-100	50	6	300	6
100-200	150	25	3750	31
200-300	250	36	9000	67
300-400	350	20	7000	87
400-500	450	13	5850	100
Total		100	25900	

i) Mean consumption of electricity $= \frac{\sum_{i=1}^{5} f_i * x_i}{\sum_{i=1}^{5} f_i} = \frac{25900}{100} = 259$

[*continued in next page*]

ii) Median = $\frac{100}{2} = 50th$ Observation

Median class = (200-300)

Lower Limit of the median class (L_0) = 200

Sum of the frequencies of all classes prior the median class (F_{-Me}) = 31

Frequency of median class (f_{Me}) = 36

Width of the median class (W_{Me}) = 300-200 = 100

Median, $Me = L_0 + \frac{(\frac{n}{2} - F_{-Me})}{f_{Me}} * W_{Me}$

$= 200 + \frac{\frac{100}{2} - 31}{36} * 100$

$= 252.78$ (Answer)

Median

Merits	Demerits
1. Rigidly defined. 2. Easy to understand and calculate. 3. Not affected very much by extreme values. 4. Can be calculated in the case of the data with open-end class. 5. Can be defined graphically.	1. In case of even number of observations, it is not defined exactly. 2. Not based on all observations. 3. Not easy for algebraic treatment. 4. For calculating median, it is necessary to arrange the data in either ascending or descending order.

Mode

Mode

- Mode: The value of the variable that occurs most frequently; that is for which the frequency is a maximum.

- Generally speaking, mode can be used to describe qualitative data.

- Mode is particularly useful average for discrete data.

- For ungrouped data / categorical variable:

  Mode is the value of the variable for which the **frequency is highest**.

Mode: Ungrouped Data

For the data sets:

i. 7, 8, 6, 7, 9, 7, and 4: Here ‘7’ appears highest 3 times, hence mode is ‘7’and the data is unimodal.

ii. 6, 4, 8, 5, 8, 1, 2, 5, 4, 7, 5, 2, 4, and 3: here ‘5’ and ‘4’ both occur highest 3 times hence the mode ‘5’ and ‘4’ and the data is bimodal.

iii. 1, 5, 7, 2, 6, 9, and 4: there is no mode.

iv. Consider the following table representing the frequency distribution of religion

Religion	Muslim	Hindu	Buddhist	Christian	Others
Frequency	18	75	12	4	2

- Here the highest frequency ‘75’ occurs for the category ‘Hindu’. Hence mode for the given data is Hindu.

Determination of Mode: Grouped Data

Formula, $Mo = L_0 + \left\{ \frac{(f_0 - f_{-1})}{(f_0 - f_{-1}) + (f_0 - f_1)} \right\} * W$

- Mo = Mode
- L_0 = Lower limit of the Modal class
- f_0 = Frequency of the modal class.
- f_{-1} = Frequency of the pre modal class.
- f_1 = Frequency of the post modal class.
- W = Width of the modal class.

Example

The Table displays summary information of the parent of 50 students. Compute the mode of the parents' income.

Income of parent (in thousand taka)	Frequency
Below 20	3
20-40	4
40-60	6
60-80	8
80-100	12
100-120	10
120 and over	7
Total	50

- Step 1: Locate the modal class.
- Step 2: Determine the lower limit(L_0) of the modal class.
- Step 3: Determine the frequency(f_0) of the modal class.
- Step 4: Determine the frequency(f_{-1}) of the pre modal class.
- Step 5: Determine the frequency(f_1) of the post modal class.
- Step 6: Determine the width of the modal class, W.

Mode

Merits	Demerits
1. Most typical and representative value of a distribution. 2. Not at all affected by extreme values. 3. Can be calculated in the case of the data with open-end class. 4. Easy to understand and calculate. 5. Can be defined graphically.	1. Not clearly defined in case of bimodal or multi modal distribution. 2. Not based on all observation. 3. Not suitable for further algebraic treatment. 4. Affected by sampling fluctuations.