

Measures of Central Tendency

7.1. DEFINITION

Measures of central tendency or measures of location give an idea about the concentration of the values (observations) in the central part of the distribution. A measure of central tendency is a statistical average or a single value which represents entire distribution.

The following are some important measures of central tendency.

- (i) Arithmetic mean
- (ii) Median
- (iii) Mode
- (iv) Geometric mean
- (v) Harmonic mean.

7.2. FEATURES OF GOOD AVERAGE (OR) REQUISITES FOR AN IDEAL MEASURE OF CENTRAL TENDENCY

According to Prof. Yule, the following are the characteristics to be satisfied by an ideal measure of central tendency.

1. It should be rigidly defined.
2. It should be easy to calculate.
3. It should be based on all the observations.
4. It should not be affected by extreme observations.
5. It should not be affected by fluctuations of sampling.
6. It should be capable of further mathematical treatment.

7.3. ARITHMETIC MEAN (OR) MEAN

(AU 2016, 2018)

Definition. Arithmetic mean of a set of observations is the sum of all observations divided by the number of observations. It is denoted by \bar{x} .

Mean for an ungrouped data

If x_1, x_2, \dots, x_n are ' n ' observations, then the mean is defined as

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

Mean for grouped data

If the variable x_i has the frequency f_i , then the mean is defined as

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

where $N = \sum_{i=1}^n f_i$ is total frequency.

Merits and Demerits of Arithmetic Mean

Merits

1. It is rigidly defined.
2. It is easy to understand and easy to calculate.
3. It is based on all the observations.
4. It is suitable for further mathematical treatment.
5. It is more accurate and more reliable.

Demerits

1. It cannot be obtained by inspection nor located through a frequency graph.
2. Arithmetic mean cannot be used if we are dealing with qualitative characteristics which cannot be measured quantitatively, example intelligence, beauty, honesty, etc.
3. If a single observation is missed or lost, arithmetic mean cannot be obtained with more accuracy.
4. It is affected very much by extreme values.
5. It cannot be calculated for open-end classes.
6. It may lead to wrong conclusions if the details of the data from which arithmetic mean is computed are not given.

PROBLEMS

PROBLEM 1. The weights of 6 competitors in a game are given below :

58, 62, 56, 63, 55, 61 kgs.

Find arithmetic mean of weight of the competitors.

SOLUTION.

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{6} \\ &= \frac{58 + 62 + 56 + 63 + 55 + 61}{6} = 59.17 \text{ kgs.}\end{aligned}$$

∴ The arithmetic mean of weight of the competitors is 59.17 kgs.

PROBLEM 2. Find arithmetic mean to the following data :

x	1	2	3	4	5
f	5	8	10	12	6

SOLUTION

x_i	f_i	$f_i x_i$
1	5	5
2	8	16
3	10	30
4	12	48
5	6	30
	$N = 41$	$\sum_{i=1}^n f_i x_i = 129$

$$N = \sum_{i=1}^n f_i = 41$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^n f_i x_i = \frac{129}{41} = 3.15$$

PROBLEM 3. Find arithmetic mean to the following data :

Marks	10-20	20-30	30-40	40-50	50-60
Frequency	5	8	25	22	10

SOLUTION

<i>Class Interval</i>	<i>Frequency</i> f_i	<i>Mid-point</i> x_i	$f_i x_i$
10-20	5	15	75
20-30	8	25	200
30-40	25	35	875
40-50	22	45	990
50-60	10	55	550
	$N = 70$		$\sum_{i=1}^n f_i x_i = 2690$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^n f_i x_i = \frac{2690}{70} = 38.43$$

4.3. MEDIAN

(AU 2016, 2018)

· Median of a distribution is the value of the variable which divides it into two equal parts, *i.e.*, median is the middle value of the distribution. Median is also called as a 'Positional Average'.

Median for an ungrouped data

If the number of observations is odd, then the median is the middle value after the values have been arranged in ascending or descending order of magnitude.

If the number observations is even, then median is the arithmetic mean of the two middle terms.

Median for a grouped data

If x_i/f_i , $i = 1, 2, \dots, n$ is frequency distribution, then

Case 1. For discrete frequency distribution, we can find the median through the following steps.

1. Arrange the data in ascending or descending order of magnitude.
2. Find cumulative frequencies
3. Find $\frac{N}{2}$, consider the cumulative frequency just greater than $\frac{N}{2}$.
4. The corresponding value of x is median.

Case 2. For continuous frequency distribution, the median is obtained by considering the median class. The class corresponding to cumulative frequency (*c.f.*) just greater than $\frac{N}{2}$ is called median class. The value of median is obtained by the following formula :

$$\text{Median} = l + \left(\frac{\frac{N}{2} - C}{f} \right) \times h$$

where

- l : lower limit of median class
- f : frequency of median class
- h : magnitude of the median class
- C : cumulative frequency preceeding to the median class.

Merits and Demerits of Median

Merits

1. It is rigidly defined.
2. It is easy to understand and easy to calculate.
3. It is not at all affected by extreme values.
4. It can be calculated for the distributions with open-end classes.

Demerits

1. In case of even number of observations median cannot be determined exactly.
2. It is not based on all the observations.
3. It is not suitable for further mathematical treatment.
4. As compared with mean, median is affected much by fluctuations of sampling.

PROBLEMS

PROBLEM 1. Find the median of the values, 25, 30, 35, 15, 10.

SOLUTION. By arranging the data into ascending order, 10, 15, 25, 30, 35.

The observations are odd, then the median is middle value of the arrangement.

Median is 3rd term value

∴ Median = 25.

PROBLEM 2. Find the median of the values 26, 8, 6, 12, 15, 32.

SOLUTION. By arranging into ascending order, 6, 8, 12, 15, 26, 32.

The observations are even, then the median is arithmetic mean of middle two terms.

∴ Median = $\frac{12 + 15}{2} = 13.5$

PROBLEM 3. Obtain median to the following data :

<i>x</i>	5	8	11	14	17	20	23
<i>f</i>	2	8	12	20	10	6	3

SOLUTION

<i>x</i>	<i>f</i>	<i>Cumulative frequency</i>
5	2	2
8	8	10
11	12	22
14	20	42
17	10	52
20	6	58
23	3	61
N = 61		

$$N = \sum_{i=1}^n f_i = 61, \quad \frac{N}{2} = 30.5$$

Cumulative frequency just greater than $\frac{N}{2}$ is 42 and the value of *x* corresponding to 42 is 14.

∴ Median = 14.

PROBLEM 4. Find median to the following data.

Class interval	40-50	50-60	60-70	70-80	80-90
Frequency	5	12	23	8	2

SOLUTION

<i>Class interval</i>	<i>Frequency</i> f_i	<i>Cumulative frequency</i> <i>(C.f.)</i>
40–50	5	5
50–60	12	17
l 60–70	23	40
70–80	8	48
80–90	2	50
	$N = 50$	

$$N = \sum_{i=1}^n f_i = 50, \quad \frac{N}{2} = 25$$

Cumulative frequency just greater than $\frac{N}{2}$ is 40, corresponding class is 60–70.

\therefore The median class is 60–70

$\therefore l = 60, c = 17, f = 23, h = 10$

$$\begin{aligned} \text{Median} &= l + \left(\frac{\frac{N}{2} - c}{f} \right) \times h \\ &= 60 + \left(\frac{25 - 17}{23} \right) \times 10 = 60 + 3.4782 = 63.4782 \end{aligned}$$

4.4. MODE

(AU 2016, 2018)

Mode is defined as the value which occurs most frequently in a set of observations.

Mode for an ungrouped data

Mode for an ungrouped data is calculated by the value of the variable which repeats maximum number of times.

Mode for a grouped data

Case 1. In case of discrete frequency distribution, mode can be obtained by inspection. The value of the variable having the maximum frequency is known as modal value.

Case 2. In case of continuous frequency distribution, the mode is calculated by using the following formula.

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right) \times h$$

where l : lower limit of modal class.

h : width of the modal class interval.

f_1 : frequency of modal class.

f_0 : frequency of preceding the modal class.

f_2 : frequency of succeeding the modal class.

Merits and Demerits of Mode

Merits

1. It is easy to calculate.
2. It is not at all affected by extreme values.
3. It can be calculated for open-end classes.
4. Mode can be calculated even if the frequency distribution has class intervals of unequal magnitude provided modal class.

Demerits

1. It is not based on all observations.
2. It is not capable for further mathematical treatment.
3. As compared with mean, mode is affected much by fluctuations of sampling.
4. If mode is ill-defined, it is not possible to find clearly defined mode.

PROBLEMS

PROBLEM 1. Obtain mode of the values

10, 12, 15, 20, 12, 16, 18, 15, 12, 10, 16, 20, 12, 24.

SOLUTION

Mode is the value which occurs the maximum number of times. In this series the value (observation) 12 repeated 4 times which is maximum frequently occurred than any other observations.

\therefore Mode = 12.

PROBLEM 2. Find mode to the following frequency distribution :

x	1	2	3	4	5	6	7	8
f	4	9	16	25	22	15	7	3

SOLUTION. The maximum frequency is 25.

The value of x corresponding to maximum frequency is 4. Hence mode is 4.

PROBLEM 3. Find mode to the following data :

$C.I.$	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Frequency	5	7	10	18	20	12	8	2

SOLUTION

C.I.	Frequency
0-5	5
5-10	7
10-15	10
15-20	18 f_0
20-25	20 f_1
25-30	12 f_2
30-35	8
35-40	2

The maximum frequency is 20 in the class interval 20-25, therefore modal class is 20-25.

$$\therefore l = 20, f_1 = 20, f_0 = 18, f_2 = 12, h = 5$$

$$\therefore \text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right) \times h = 20 + \left(\frac{20 - 18}{2 \times 20 - 18 - 12} \right) \times 5 = 21$$

4.5. GEOMETRIC MEAN

(AU 2016, 2017, 2018)

Definition

The geometric mean of a set of n observations is the n th root of their product. It is denoted by G.

Geometric mean for an ungrouped data

If x_1, x_2, \dots, x_n are 'n' observations, then Geometric mean G is calculated by

$$G = (x_1 \cdot x_2 \cdot \dots \cdot x_n)^{1/n}$$

or
$$\log G = \frac{1}{n} (\log x_1 + \log x_2 + \dots + \log x_n)$$

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

$$G = \text{Antilog} \left[\frac{1}{n} \sum_{i=1}^n \log x_i \right]$$

Geometric mean for a grouped data

If $x_i/f_i; i = 1, 2, \dots, n$ is a grouped frequency distribution, the geometric mean G is calculated by the following formula

$$G = \left[x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_n^{f_n} \right]^{\frac{1}{N}}$$

where $N = \sum_{i=1}^n f_i$

or
$$\log G = \frac{1}{N} [f_1 \log x_1 + f_2 \log x_2 + \dots + f_n \log x_n] = \frac{1}{N} \sum_{i=1}^n f_i \log x_i$$

$$\therefore G = \text{Antilog} \left[\frac{1}{N} \sum_{i=1}^n f_i \log x_i \right]$$

Merits and Demerits of Geometric Mean

Merits

1. It is rigidly defined.
2. It is based on all items.
3. It is suitable for finding averages for ratios, rates and percentages.
4. It is capable for further mathematical treatment.

Demerits

1. Geometric mean is not easy to calculate and understand for a non-mathematics person.
2. If any observations is zero, geometric mean becomes zero and if any one is negative, it becomes imaginary regarding the magnitude of the item.

PROBLEMS

PROBLEM 1. Find the geometric mean of 16, 64 and 108.

SOLUTION. The geometric mean G is given by

$$G = \sqrt[3]{16 \times 64 \times 108} = 48$$

PROBLEM 2. The birth rate of a country in the years 1980 and 1990 are 8% and 10% respectively. Find average birth rate.

SOLUTION. The geometric mean of the birth rates in the country is

$$G = \sqrt{108 \times 110} = 104.9$$

\therefore The average birth rate according to geometric mean is 4.9%.

PROBLEM 3. Find the geometric mean to the following :

x	3	8	10	12
f	15	20	24	8

SOLUTION

x_i	f_i	$\log x_i$	$f_i \log x_i$
3	15	0.4771	7.1565
8	20	0.9031	18.062
10	24	1	24
12	8	1.0792	8.6336
$N = 67$			$\sum_{i=1}^n f_i \log x_i = 57.8521$

$$N = \sum_{i=1}^n f_i = 67$$

$$G = \text{Antilog} \left[\frac{1}{N} \sum_{i=1}^n f_i \log x_i \right]$$

$$= \text{Antilog} \left[\frac{57.8521}{67} \right] = \text{Antilog} [0.8635] = 7.303$$

∴ The geometric mean $G = 7.303$.

PROBLEM 4. Find the geometric mean to the following data.

Class interval	4–8	8–12	12–16	16–20	20–24
Frequency	2	4	8	5	3

SOLUTION

Class interval	Frequency f_i	Mid-value x_i	$\log x_i$	$f_i \log x_i$
4–8	2	6	0.7782	1.5564
8–12	4	10	1.0000	4.0000
12–16	8	14	1.1461	9.1688
16–20	5	18	1.2553	6.2765
20–24	3	22	1.3424	4.0272
	$N = 22$			$\sum_{i=1}^n f_i \log x_i$ $= 25.0289$

$$N = \sum_{i=1}^n f_i = 22$$

The geometric mean G is given by

$$G = \text{Antilog} \left[\frac{1}{N} \sum_{i=1}^n f_i \log x_i \right]$$

$$= \text{Antilog} \left[\frac{25.0289}{22} \right] = \text{Antilog} [1.1377] = 13.7309$$

∴ The geometric mean of the given data is $G = 13.7309$.

4.6. HARMONIC MEAN

(AU 2016, 2018)

Definition

Harmonic mean of a number observations is the reciprocal of the arithmetic mean of the reciprocals of the given values. It is denoted by H .

Harmonic mean for an ungrouped data

Let x_1, x_2, \dots, x_n be 'n' observations, then harmonic mean H is given by

$$H = \frac{1}{\frac{1}{n} \sum_{i=1}^n \left(\frac{1}{x_i} \right)}$$

Harmonic mean for a grouped frequency distribution

If $x_i/f_i, i = 1, 2, \dots, n$ is a grouped frequency distribution, then harmonic mean H is calculated by

$$H = \frac{1}{\frac{1}{N} \sum_{i=1}^n \left(\frac{f_i}{x_i} \right)}$$

where,
$$N = \sum_{i=1}^n f_i$$

Merits and Demerits of Harmonic Mean

Merits

1. It is rigidly defined.
2. It is defined on all the observations.
3. It is suitable for further mathematical treatment.
4. It is not affected by fluctuations of sampling.
5. It gives greater importance to small items and also it is useful only when small items have to be given a greater weightage.

Demerits

1. It is not easy to understand.
2. It is difficult to calculate.
3. It gives greater importance to small items.

PROBLEMS

PROBLEM 1. A person travel from Vijayawada to Tirupati at 80 km per hour and returns at 70 km per hour. Find the average speed of the journey.

SOLUTION. The average speed of the journey by using harmonic mean is

$$H = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}} = \frac{1}{\frac{1}{2} \left[\frac{1}{80} + \frac{1}{70} \right]} = 74.67$$

The harmonic mean (average speed of journey) H is 74.67 km per hour.

PROBLEM 2. Find the harmonic mean to the following frequency distribution.

<i>C.I.</i>	10–20	20–30	30–40	40–50	50–60
<i>f</i>	8	15	30	17	10

SOLUTION

<i>C.I.</i>	<i>f_i</i>	<i>Mid-value</i> <i>x_i</i>	$\frac{f_i}{x_i}$
10–20	8	15	0.5333
20–30	15	25	0.6000
30–40	30	35	0.8571
40–50	17	45	0.3778
50–60	10	55	0.1818
	N = 80		$\sum_{i=1}^n \left(\frac{f_i}{x_i}\right) = 2.5500$

$$N = \sum_{i=1}^n f_i = 80,$$

The harmonic mean H is given by

$$H = \frac{1}{\frac{1}{N} \sum_{i=1}^n \left(\frac{f_i}{x_i}\right)} = \frac{1}{\frac{2.5500}{80}} = 31.3725$$

The harmonic mean H = 31.3725.

4.7. MEDIAN AND MODE THROUGH GRAPH

In a grouped frequency distribution, two measures of central tendency median and mode can also be determined graphically. Now we shall discuss the procedures of locating median and mode graphically.

4.8. LOCATING MEDIAN GRAPHICALLY

Median can be easily located graphically with the help of ogive curve or cumulative frequency curve. To obtain ogive curve, first we calculate less than c.f. and more than c.f. Then obtain less than ogive by plotting less than c.f. against upper limits of corresponding classes and more than ogive by plotting more than c.f. against lower limits of corresponding classes. Now draw either less than ogive or more than ogive.

To locate median graphically, mark a point corresponding to $\frac{N}{2}$ along Y-axis. From

this point draw a line parallel to X-axis which is meeting ogive. Now draw a line perpendicular to X-axis from this point which gives the value of median.

Alternative Procedure

First draw less than ogive and more than ogive curves for the given frequency distribution. We can easily observe that these two ogive curves intersect at a particular point. From the point of intersection of these ogive curves, draw a perpendicular line to the X-axis. The value of the point where this perpendicular line meets the X-axis gives the median.

PROBLEM 1. The profits of 100 companies during the year 2018-19 are given below. Calculate median graphically.

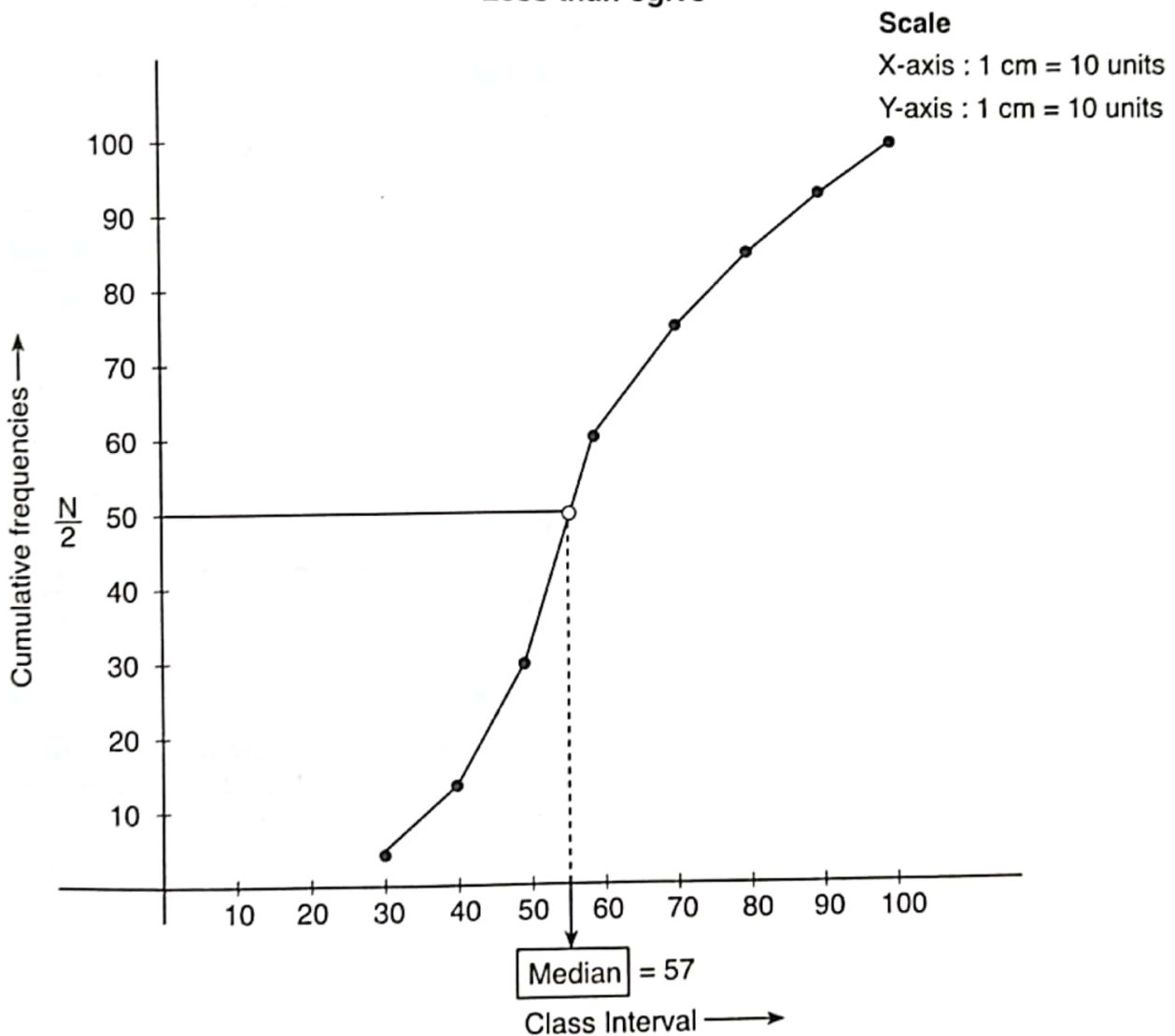
<i>Profits (in crores of Rupees)</i>	<i>No. of Companies</i>
20-30	4
30-40	8
40-50	18
50-60	30
60-70	15
70-80	10
80-90	8
90-100	7

SOLUTION

<i>Profits (in crores of ₹)</i>	<i>No. of Companies</i>	<i>Less than cumulative frequencies</i>
20-30	4	4
30-40	8	12
40-50	18	30
50-60	30	60
60-70	15	75
70-80	10	85
80-90	8	93
90-100	7	100
	N = 100	

<i>Upper limit of C.I.</i>	<i>Less than c.f.</i>
30	4
40	12
50	30
60	60
70	75
80	85
90	93
100	100

Less than ogive



Median located from the less than ogive curve is approximately 57.

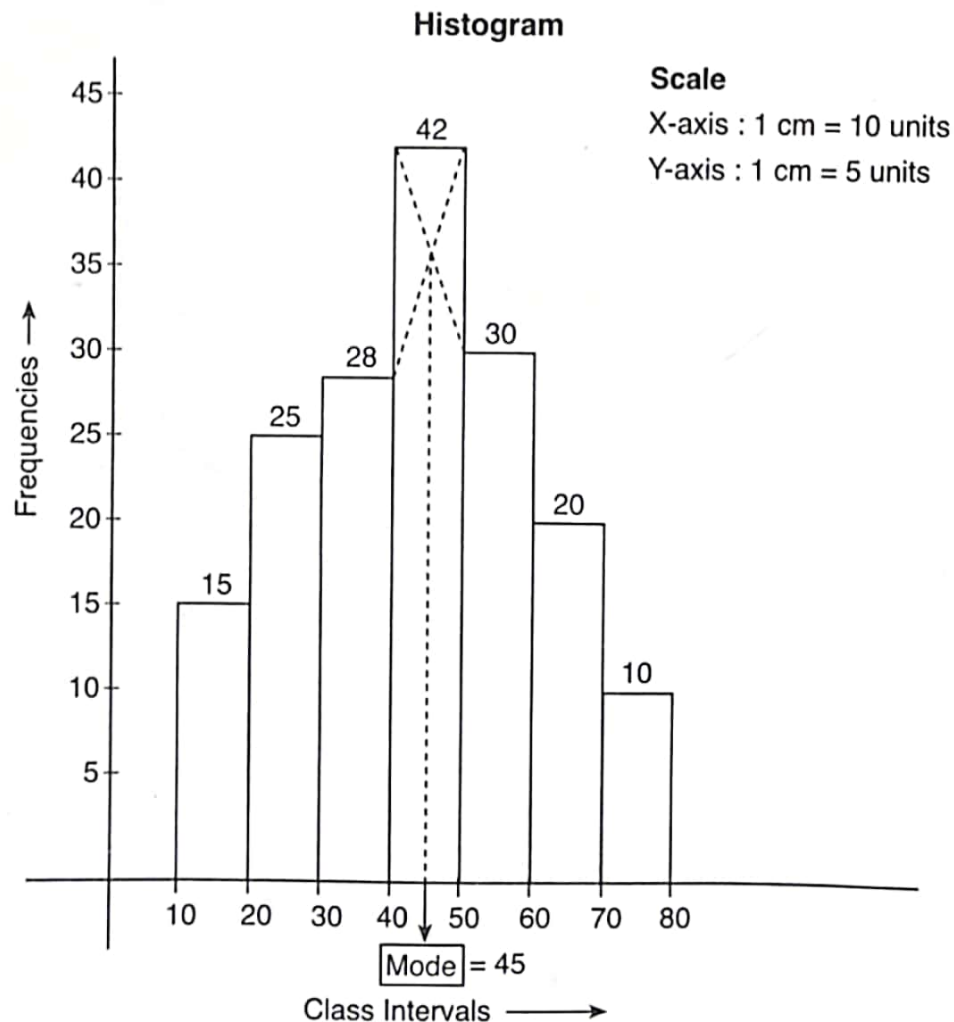
4.9. LOCATING MODE GRAPHICALLY

To locate mode graphically, first we construct histogram for the given grouped frequency distribution. Then draw two straight lines diagonally on the inside of the model class bars, starting from each upper corner of the bar to upper corner of adjacent bar. Now draw a perpendicular line from the intersection point of the two diagonal lines to the X-axis which gives us the value of the mode.

PROBLEM 1. Obtain mode graphically to the following data.

C.I.	10-20	20-30	30-40	40-50	50-60	60-70	70-80
f	15	25	28	42	30	20	10

SOLUTION



First histogram was constructed, then we observed the model class, which is 40-50. Now draw two straight lines diagonally inside of the model class bars from the upper corners of the adjacent bars. Then a perpendicular line is drawn from the intersection of the two diagonal lines on X-axis. This point on X-axis gives us mode 45 approximately.

Measures of Dispersion

5.1. INTRODUCTION

Measures of central tendency give us an idea about the concentration of the central part of the distribution by using a single value. If we know the average alone we cannot form a complete idea about the distribution. This will be clear from the following example.

Consider two series

Series-I 8, 9, 10, 11, 12

Series-II 2, 6, 10, 14, 18

If we observe the two series, the mean as well as median are same which are equal to 10. Therefore with the central tendency we cannot form a complete idea. Hence there is another measure required to get an idea about the distribution called measure of dispersion.

Definition. Dispersion means scatteredness. Measures of dispersion study about the homogeneity or heterogeneity of the distribution.

Series I is less scattered and Series-II is more scattered in the above example.

Characteristics of an ideal measure of Dispersion

According to Prof. Yule, the following are the characteristics of an ideal measure of dispersion.

1. It should be rigidly defined.
2. It should be easy to calculate.
3. It should be easy to understand.
4. It should be based on all the observations.

5. It should be suitable for further mathematical treatment.
6. It should be affected least by fluctuations of sampling.
7. It should be readily comprehensible.

The following are the measures of dispersion

1. Range
2. Quartile deviation
3. Mean deviation
4. Standard deviation.

5.2. RANGE

(AU 2017)

The range is the difference between two extreme observations of the distribution.

$$\text{Range} = \text{Maximum Value} - \text{Minimum Value}$$

Problem 1

Find the range of the values

7, 9, 15, 14, 6, 8

Solution

$$\text{Range} = \text{Maximum Value} - \text{Minimum Value} = 15 - 6 = 9$$

Merits and Demerits of Range

Merits

1. It is simple to understand.
2. It is easy to calculate.
3. It is useful in the statistical quality control, forecasting, price analysis etc.

Demerits

1. It is based on only two observations.
2. It cannot be calculated for open end classes.
3. It is not suitable for further mathematical treatment.
4. It is much affected by the extreme values.
5. It is very rarely used measure.

5.3. QUARTILE DEVIATION

(AU 2017)

The quartile deviation is given by

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

where Q_1 is first quartile
 Q_3 is third quartile.

It is also known as semi-interquartile range.

Quartile deviation for an ungrouped data

For an ungrouped data, first quartile is obtained by $\left(\frac{n+1}{4}\right)^{\text{th}}$ observation and third quartile is $3\left(\frac{n+1}{4}\right)^{\text{th}}$ observation. And then $Q.D. = \frac{Q_3 - Q_1}{2}$.

Quartile deviation for a grouped data

For a grouped data, the first and third quartile are calculated by

$$Q_1 = l_1 + \left(\frac{\frac{N}{4} - c_1}{f_1}\right) \times h$$

where l_1 : lower limit of first quartile class
 c_1 : cumulative frequency of the class preceding the first quartile class.
 f_1 : frequency of first quartile class
 h : width of the quartile class or class interval.

$$Q_3 = l_3 + \left(\frac{\frac{3N}{4} - c_3}{f_3}\right) \times h$$

where l_3 : lower limit of third quartile class.
 c_3 : cumulative frequency of the class preceding the third quartile class.
 f_3 : frequency of third quartile class.

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

Merits and Demerits of Quartile deviation

Merits

1. It is simple and easy to calculate.
2. It is not affected by extreme values.
3. It can be calculated for the data with open end classes.

Demerits

1. It is not based on all the observations.
2. It is not suitable for further mathematical treatment.
3. It is affected by fluctuations of sampling.

PROBLEMS

PROBLEM 1. Find quartile deviation of the values

2, 4, 10, 8, 6, 12, 5.

SOLUTION. By arranging into ascending order,

2, 4, 5, 6, 8, 10, 12

First quartile = $\left(\frac{n+1}{4}\right)^{\text{th}}$ observation = $\left(\frac{7+1}{4}\right)^{\text{th}}$ = 2nd observation = 4

$$\therefore Q_1 = 4$$

Third quartile = $3 \left(\frac{n+1}{4}\right)^{\text{th}}$ observation

$$Q_3 = 3 \times 2 = 6^{\text{th}} \text{ term} = 10$$

$$\therefore \text{Q.D.} = \frac{Q_3 - Q_1}{2} = \frac{10 - 4}{2} = 3$$

PROBLEM 2. Find quartile deviation to the following data :

<i>C.I.</i>	0-20	20-40	40-60	60-80	80-100
<i>f</i>	10	25	40	15	10

SOLUTION

<i>C.I.</i>	<i>f_i</i>	<i>Cumulative frequency (c.f.)</i>
0-20	10	10 → <i>c₁</i>
20-40	25 → <i>f_i</i>	35
40-60	40	75
60-80	15	90
80-100	10	100
	N = 100	

$$N = \sum_{i=1}^n f_i = 100$$

First Quartile

$\frac{N}{4} = 25$, just greater than $\frac{N}{4}$ cumulative frequency is 35, the corresponding class is 20-40. Therefore first quartile class is 20-40

$$l_1 = 20, c_1 = 10, f_1 = 25, h = 20$$

$$\therefore Q_1 = l_1 + \left(\frac{\frac{N}{4} - c_1}{f_1}\right) \times h = 20 + \left(\frac{25 - 10}{25}\right) \times 20 = 32$$

Third Quartile

$\frac{3N}{4} = 75$, just greater than $\frac{3N}{4}$ cumulative frequency is 90, the corresponding class is 60–80. Therefore third quartile class is 60–80.

$$l_3 = 60, c_3 = 75, f_3 = 15, h = 20.$$

$$\therefore Q_3 = l_3 + \left(\frac{\frac{3N}{4} - c_3}{f_3} \right) \times h = 60 + \left(\frac{75 - 75}{15} \right) \times 20 = 60$$

$$\therefore \text{Q.D.} = \frac{Q_3 - Q_1}{2} = \frac{60 - 32}{2} = 14$$

5.4. MEAN DEVIATION

(AU 2017, 2018)

Mean deviation is the arithmetic mean of absolute deviations from the mean or median or mode. It is usually denoted by M.D.

Mean deviation for an ungrouped data

$$\text{M.D.} = \frac{1}{n} \sum_{i=1}^n |x_i - A|,$$

where A is either mean or median or mode.

Mean deviation for a grouped data

If $x_i | f_i, i = 1, 2, \dots, n$ is the grouped frequency distribution, then the mean deviation is given by

$$\text{M.D.} = \frac{1}{N} \sum_{i=1}^n f_i |x_i - A|$$

where $N = \sum_{i=1}^n f_i$, A is usually mean or median or mode.

Merits and Demerits of Mean deviation

Merits

1. It is simple to understand and easy to calculate.
2. It is rigidly defined.
3. It is based on all the observations.
4. It is less affected by the extreme values.
5. It is a good measure for comparative studies.

Demerits

1. It is not an accurate measure of dispersion.
2. It is not suitable for further mathematical treatment.
3. It is rarely used and it is not as popular as standard deviation.

PROBLEMS

PROBLEM 1. Find mean deviation of the values 2, 5, 8, 11, 14 from their respective mean.

SOLUTION.

$$\bar{x} = \text{Mean} = \frac{2 + 5 + 8 + 11 + 14}{5} = 8$$

$$\text{M.D.} = \frac{1}{n} \sum_{i=1}^n |x_i - A|$$

Consider $A = \bar{x} = 8$

$$\begin{aligned} \therefore \text{M.D.} &= \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| = \frac{1}{n} \sum_{i=1}^n |x_i - 8| \\ &= \frac{1}{5} [|2 - 8| + |5 - 8| + |8 - 8| + |11 - 8| + |14 - 8|] \\ &= 3.6 \end{aligned}$$

PROBLEM 2. Calculate mean deviation about median to the following data.

C.I.	10-20	20-30	30-40	40-50	50-60
f	5	10	15	10	5

SOLUTION

C.I.	f _i	Mid value x _i	C.f.	f _i x _i - A = f _i x _i - Median
10-20	5	15	5	100
20-30	10	25	15	100
30-40	15	35	30	0
40-50	10	45	40	100
50-60	5	55	45	100
N = 45				$\sum_{i=1}^n f_i x_i - A = 400$

$$\text{Median} = l + \left(\frac{\frac{N}{2} - c}{f} \right) \times h = 30 + \left(\frac{22.5 - 15}{15} \right) \times 10 = 35$$

\therefore M.D. about Median

$$\begin{aligned} \text{M.D.} &= \frac{1}{N} \sum_{i=1}^n f_i |x_i - A|, \text{ where } A \text{ is median} \\ &= \frac{400}{45} = 8.89 \end{aligned}$$

Definition

Standard deviation is the positive square root of the arithmetic mean of the squares of the deviations of the given values from their arithmetic mean. It is denoted by σ or S.D.

Standard deviation for an ungrouped data

For ungrouped data, the standard deviation σ is given by

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

where \bar{x} is the mean of the given data.

Standard deviation for a grouped data

If x_i/f_i , $i = 1, 2, \dots, n$ is the grouped frequency distribution, then the standard deviation is given by

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

where $N = \sum_{i=1}^n f_i$

Merits and Demerits of Standard Deviation**Merits**

1. It is rigidly defined.
2. It is based on all the observations.
3. It is suitable for further mathematical treatment.
4. It is less affected by the fluctuations of sampling.
5. The actual sign of the deviations are used in the calculation of standard deviation.
6. Standard deviation is the basis for calculating the variance, correlation coefficient etc.

Demerits

1. It is not easy to understand and difficult to calculate.
2. It gives more weight to extreme values.
3. It cannot be used for the comparative studies.

PROBLEMS

PROBLEM 1. Compute standard deviation of the values 10, 6, 8, 12, 20.

SOLUTION

$$\text{mean} = \bar{x} = \frac{10 + 6 + 8 + 12 + 20}{5} = \frac{56}{5} = 11.2$$

$$\begin{aligned} \text{S.D.} = \sigma &= \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \sqrt{\frac{1}{5} [(10 - 11.2)^2 + (6 - 11.2)^2 + (8 - 11.2)^2 + (12 - 11.2)^2 + (20 - 11.2)^2]} \\ &= \sqrt{\frac{116.8}{5}} = 4.83 \end{aligned}$$

PROBLEM 2. Find the standard deviation of 14, 15, 16, 17, 18, 19, 20. (AU 2016)

SOLUTION.

$$\begin{aligned} \text{mean} = \bar{x} &= \frac{14 + 15 + 16 + 17 + 18 + 19 + 20}{7} \\ &= 17 \end{aligned}$$

$$\begin{aligned} \text{S.D.} = \sigma &= \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \sqrt{\frac{1}{7} [(14 - 17)^2 + (15 - 17)^2 + (16 - 17)^2 + (17 - 17)^2 + (18 - 17)^2 \\ &\quad + (19 - 17)^2 + (20 - 17)^2]} \\ &= \sqrt{\frac{28}{7}} = 2 \end{aligned}$$

PROBLEM 3. Calculate standard deviation to the following grouped frequency distribution.

C.I.	0-10	10-20	20-30	30-40	40-50	50-60
f	2	8	16	28	12	4

SOLUTION

C.I.	f_i	Mid value x_i	$f_i x_i$	$f_i (x_i - \bar{x})^2$	$f_i x_i^2$
0-10	2	5	10	1504.81	50
10-20	8	15	120	2430.44	1,800
20-30	16	25	400	883.28	10,000
30-40	28	35	980	184.94	34,300
40-50	12	45	540	1896.06	24,300
50-60	4	55	220	2037.62	12,100
	$N = 70$		$\sum_{i=1}^n f_i x_i$ $= 2270$	$\sum_{i=1}^n f_i (x_i - \bar{x})^2$ $= 8937.15$	$\sum_{i=1}^n f_i x_i^2$ $= 82550$

$$N = \sum_{i=1}^n f_i = 70$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^n f_i x_i = \frac{2270}{70} = 32.43$$

$$\text{S.D.} = \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^n f_i (x_i - \bar{x})^2} = \sqrt{\frac{8937.15}{70}} = 11.29$$

Alter :

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^n f_i x_i^2 - (\bar{x})^2} = \sqrt{\frac{82550}{70} - (32.43)^2} = 11.29$$

5.6. VARIANCE

Definition. Variance is arithmetic mean of squares of the deviations from their arithmetic mean. It is denoted by σ^2 . The square of standard deviation is called variance. Variance of the variable X is denoted by $\text{var}(X)$ or σ_X^2 .

Variance for an Ungrouped data

For an ungrouped data, the variance σ^2 is given by

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (\text{or}) \quad \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$$

where \bar{x} is the mean of the given data

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

Variance for a grouped data

If $x_i/f_i, i = 1, 2, \dots, n$ is the grouped frequency distribution, then variance is given by

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^n f_i (x_i - \bar{x})^2 \quad (\text{or}) \quad \frac{1}{N} \sum_{i=1}^n f_i x_i^2 - \bar{x}^2$$

where \bar{x} is mean

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

and $N = \sum_{i=1}^n f_i$ is total frequency

Merits and Demerits of Variance

Merits

1. It is rigidly defined
2. It is based on all the observations.
3. It is suitable for further mathematical treatment.
4. It is less affected by the fluctuations of sampling.

Demerits

1. It is not easy to understand and difficult to calculate.
2. It gives more weight to extreme values.
3. It cannot be used for comparative studies.

PROBLEM 1. Calculate variance to the following data.

4, 8, 10, 15, 23, 24

SOLUTION.

$$\text{mean} = \bar{x} = \frac{4 + 8 + 10 + 15 + 23 + 24}{6} = \frac{84}{6} = 14$$

$$\begin{aligned} \text{Variance} = \sigma^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \frac{1}{6} [(4 - 14)^2 + (8 - 14)^2 + (10 - 14)^2 + (15 - 14)^2 + (23 - 14)^2 + (24 - 14)^2] \\ &= \frac{334}{6} = 55.67 \end{aligned}$$

(or)

$$\begin{aligned} \text{Variance} = \sigma^2 &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \\ &= \frac{1}{6} [4^2 + 8^2 + 10^2 + 15^2 + 23^2 + 24^2] - (14)^2 \\ &= \frac{1510}{6} - 196 = 55.67 \end{aligned}$$

PROBLEM 2. Calculate variance to the following data.

<i>C.I.</i>	10-14	14-18	18-22	22-26	26-30	30-34
<i>f</i>	4	9	15	22	7	3

SOLUTION

<i>C.I.</i>	f_i	<i>Mid value</i> x_i	$f_i x_i$	$f_i x_i^2$
10–14	4	12	48	576
14–18	9	16	144	2,304
18–22	15	20	300	6,000
22–26	22	24	528	12,672
26–30	7	28	196	5,488
30–34	3	32	96	3,072
	$N = 60$		$\sum_{i=1}^n f_i x_i = 1,312$	$\sum_{i=1}^n f_i x_i^2 = 30,112$

$$N = \sum_{i=1}^n f_i = 60,$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^n f_i x_i = \frac{1,312}{60} = 21.87$$

$$\begin{aligned} \text{Variance } \sigma^2 &= \frac{1}{N} \sum_{i=1}^n f_i x_i^2 - (\bar{x})^2 \\ &= \frac{30,112}{60} - (21.87)^2 = 23.57 \end{aligned}$$

6.1. MOMENTS

Moments are statistical averages. They are of two kinds.

1. Non-central moments
2. Central moments

Non-Central Moments

(AU 2016, 2018)

The r th moment about any point 'A' are called non-central moments. Non-central moments are denoted by μ_r' and is given by

$$\mu_r' = \frac{1}{N} \sum_{i=1}^n f_i (x_i - A)^r$$

where $N = \sum_{i=1}^n f_i$

Let $d_i = x_i - A,$

$$\mu_r' = \frac{1}{N} \sum_{i=1}^n f_i d_i^r$$

If we take the arbitrary point $A = 0$, the non-central moments are called raw moments, these are moments about origin.

The r th raw moment is given by

$$\mu_r' = \frac{1}{N} \sum_{i=1}^n f_i x_i^r$$

If $r = 1,$ $\mu_1' = \frac{1}{N} \sum_{i=1}^n f_i x_i = \bar{x}$

$$\therefore \mu_1' = \bar{x}, \text{ the mean in the raw moments}$$

$$r = 2, \quad \mu_2' = \frac{1}{N} \sum_{i=1}^n f_i x_i^2$$

$$r = 3, \quad \mu_3' = \frac{1}{N} \sum_{i=1}^n f_i x_i^3$$

$$r = 4, \quad \mu_4' = \frac{1}{N} \sum_{i=1}^n f_i x_i^4$$

Central Moments (AU 2016, 2018)

The r th moment about the mean \bar{x} are called central moments. Central moments are denoted by μ_r and is given by

$$\mu_r = \frac{1}{N} \sum_{i=1}^n f_i (x_i - \bar{x})^r$$

Let $z_i = x_i - \bar{x},$

$$\mu_r = \frac{1}{N} \sum_{i=1}^n f_i z_i^r$$

If $r = 0, \quad \mu_0 = \frac{1}{N} \sum_{i=1}^n f_i = 1$

$$r = 1, \quad \mu_1 = \frac{1}{N} \sum_{i=1}^n f_i (x_i - \bar{x}) = \frac{1}{N} \sum f_i x_i - \bar{x} \cdot \frac{1}{N} \sum f_i = \bar{x} - \bar{x} = 0$$

$$\mu_1 = 0$$

$\therefore \mu_0 = 1, \mu_1 = 0$

$$r = 2, \quad \mu_2 = \frac{1}{N} \sum_{i=1}^n f_i (x_i - \bar{x})^2 = \sigma^2$$

$\therefore \mu_2 = \sigma^2$, the variance of the variable x .

Note : If $d_i = x_i - A$

Multiply by f_i and taking summations,

$$\sum_{i=1}^n f_i d_i = \sum_{i=1}^n f_i (x_i - A)$$

Divide by N ,

$$\frac{1}{N} \sum_{i=1}^n f_i d_i = \frac{1}{N} \sum_{i=1}^n f_i (x_i - A)$$

$$\mu_1' = \frac{1}{N} \sum_{i=1}^n f_i x_i - A \frac{1}{N} \sum_{i=1}^n f_i$$

$$\mu_1' = \bar{x} - A$$

$\therefore \bar{x} = A + \mu_1'$

6.3. EFFECT OF CHANGE OF ORIGIN AND SCALE ON MOMENTS

(AU 2018)

Statement. Moments are independent of changing the origin but not on scale.

Proof :

$$\begin{aligned}\text{Let } u &= \frac{x-a}{h} \\ x &= a + hu \\ \bar{x} &= a + h\bar{u} \\ x - \bar{x} &= h(u - \bar{u})\end{aligned}$$

Non-central moments

$$\begin{aligned}\mu_r' &= \frac{1}{N} \sum_{i=1}^n f_i (x_i - a)^r = \frac{1}{N} \sum_{i=1}^n f_i (h u_i)^r & [\because x_i - a = h u_i] \\ &= h^r \cdot \frac{1}{N} \sum_{i=1}^n f_i u_i^r\end{aligned}$$

Central moments

$$\begin{aligned}\mu_r &= \frac{1}{N} \sum_{i=1}^n f_i (x_i - \bar{x})^r = \frac{1}{N} \sum_{i=1}^n f_i [h(u_i - \bar{u})]^r \\ &= h^r \cdot \frac{1}{N} \sum_{i=1}^n f_i (u_i - \bar{u})^r\end{aligned}$$

Thus the r th moment of the variable x (both central and non-central moments) h^r times r th moment of u .

\therefore Moments are independent of changing the origin but dependent of changing the scale.

6.4. SHEPPARD'S CORRECTION FOR MOMENTS

In case of grouped frequency distribution, while calculating moments we assume that the frequencies are concentrated at the middle point of the class intervals. This assumption is generally true for symmetrical distributions or even for moderately skewed distributions in which the class intervals are not greater than one-twentieth $\left(\frac{1}{20} \text{ th}\right)$ of the range of distribution. This assumption leads to a symmetric error in the even moments.

$$\mu_2 (\text{corrected}) = \mu_2 - \frac{h^2}{12}$$

$$\mu_3 (\text{corrected}) = \mu_3 (\because \text{ odd moment})$$

$$\mu_4 (\text{corrected}) = \mu_4 - \frac{1}{2} h^2 \mu_2 + \frac{7}{240} \cdot h^4.$$

Uses of Moments

- Moments are useful to find statistical averages. Some important and frequently used measures are listed below.
 - First moment about origin is mean (μ_1')
 - Second moment about mean is variance (μ_2).
- The measure of skewness is calculated by using third central moment (μ_3)
- The measure of kurtosis is calculated by using fourth central moment (μ_4).

6.5. SKEWNESS

(AU 2017)

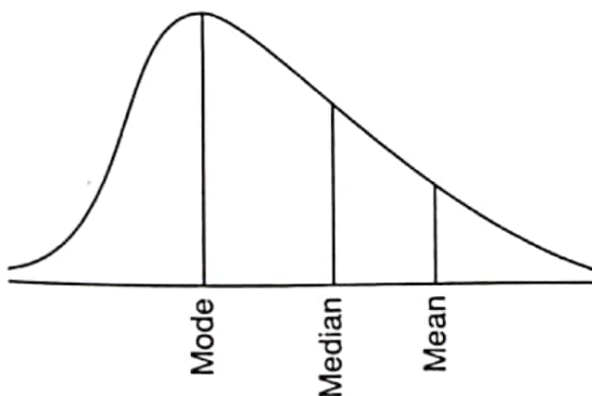
Definition

Skewness means lack of symmetry. It gives an idea about the shape of the curve of the given data.

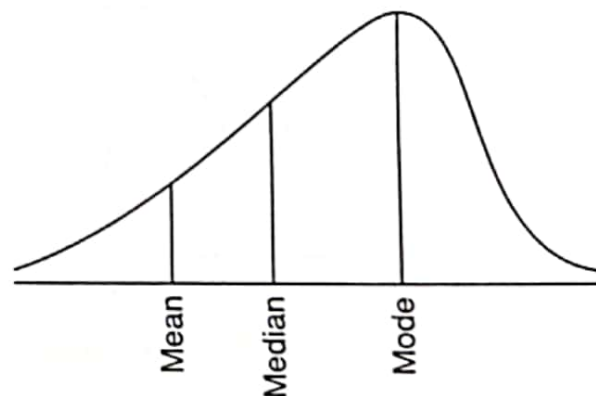
A distribution is said to be skewed if

- Mean, median and mode fall at different points *i.e.* Mean \neq Median \neq Mode.
- Quartiles are not equidistant from median, *i.e.*, $Q_3 - Q_2 \neq Q_2 - Q_1$.
- The curve drawn with the help of the given data is not symmetrical but stretched more to one side than to the other (as explained through the following curves).

(a) Rightly stretched curve
(mode < median < mean)



(b) Leftly stretched curve
(mean < median < mode)



Measures of Skewness (AU 2016)

They are several measures for skewness. (Denote coefficient of skewness by S_K)

Absolute measures of skewness

- $S_K = \text{Mean} - \text{Mode}$
- $S_K = \text{Mean} - \text{Median}$
- $S_K = (Q_3 - \text{Md}) - (\text{Md} - Q_1)$ where $\text{Md} = Q_2$

Karl-Pearson's Coefficient of Skewness

1.
$$S_K = \frac{\text{Mean} - \text{Mode}}{\text{Standard deviation}} = \frac{\text{Mean} - \text{Mode}}{\text{S.D.}}$$

2. If mode is illdefined, then by using empirical mode as

$$\text{Mode} = 3 \text{ Median} - 2 \text{ Mean}$$

$$\therefore S_K = \frac{3 (\text{Mean} - \text{Median})}{\text{S.D.}}$$

Limits of this coefficient are ± 3 .

Bowley's Coefficient of Skewness

It is also known as quartile coefficient of skewness.

$$S_K = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)} = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

or

$$S_K = \frac{Q_3 + Q_1 - 2 \text{ Md}}{Q_3 - Q_1}$$

Coefficient of Skewness based on moments

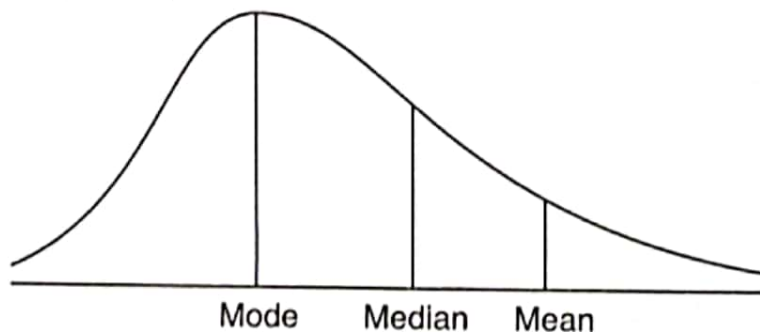
$$S_K = \frac{\sqrt{\beta_1} (\beta_2 + 3)}{2 (5\beta_2 - 6\beta_1 - 9)}$$

where $\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \beta_2 = \frac{\mu_4}{\mu_2^2}$

Skewness may be positive or negative and there may be no skewness called symmetrical. These three types can be explained through the following curves.

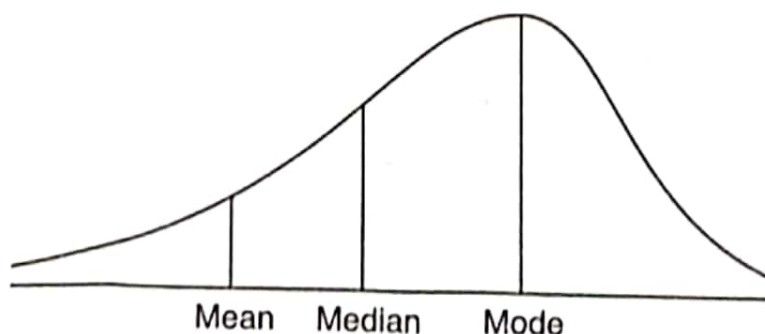
1. Positively skewed distribution

$$\text{Mean} > \text{Median} > \text{Mode}$$



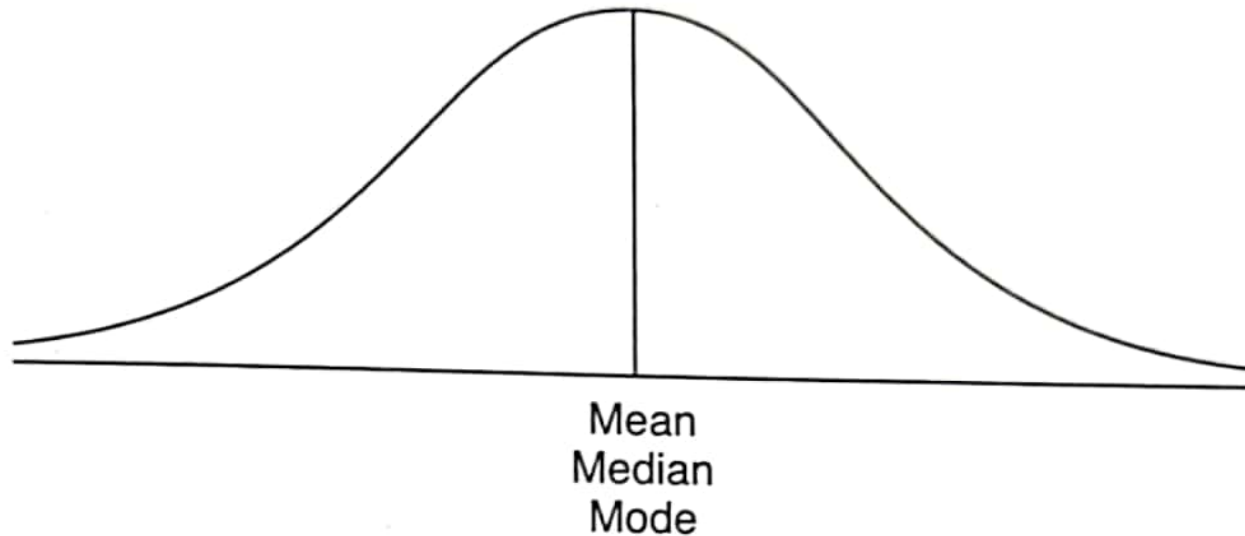
2. Negatively skewed distribution

$$\text{Mean} < \text{Median} < \text{Mode}$$



3. Symmetrical distribution

Mean = Median = Mode



It can be observed from the curves, the skewness is said to be positive if the curve stretched right and $\text{mean} > \text{median} > \text{mode}$. The skewness can be negative if the curve stretched left and $\text{mean} < \text{median} < \text{mode}$. The distribution is said to be symmetrical if the mean, median and mode coincide at the same point *i.e.* they are equal.

Theorem 1 Show that $K_3 = 0$ for a symmetrical distribution.

6.6. KURTOSIS

(AU 2016, 2017)

The measures of central tendency, dispersion and skewness may not give a complete idea about the distribution, therefore another statistical measure is required to know about flatness or peakedness of the curve called Kurtosis. It is discovered by Prof. Karl Pearson. Pearson called Kurtosis as convexity of curve.

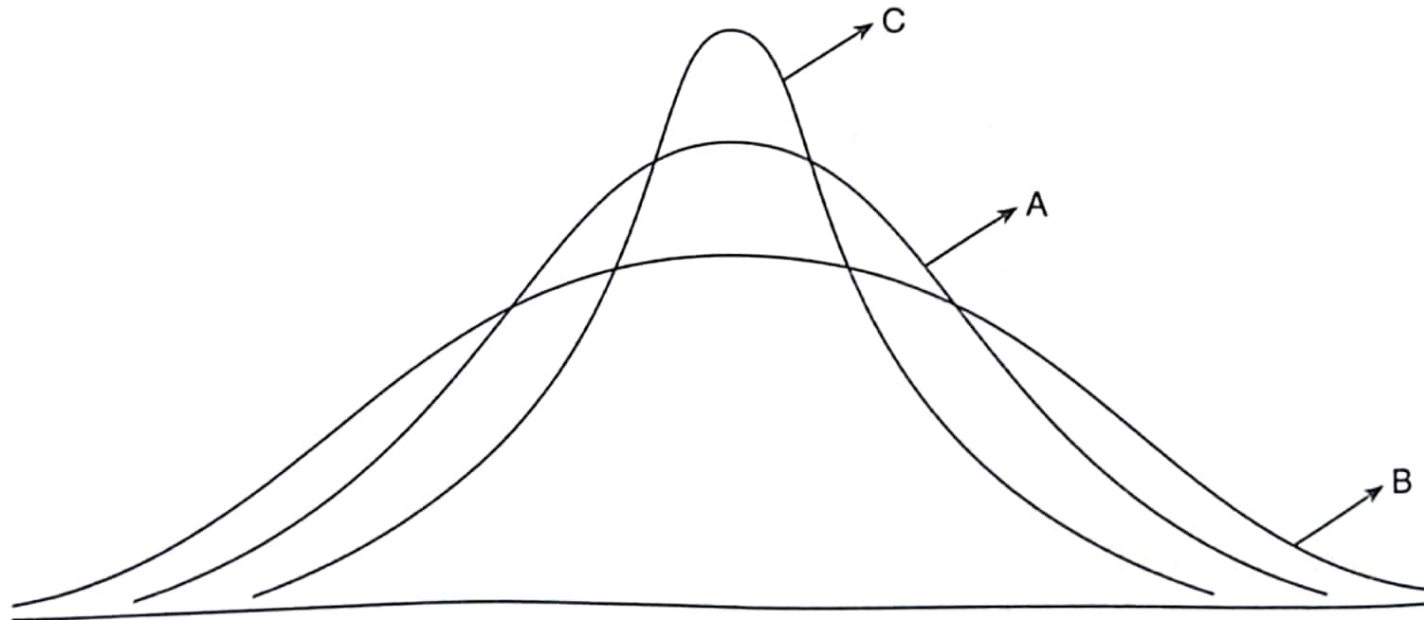
Kurtosis gives an idea about the flatness or peakedness of the curve. It is measured by the coefficient β_2 or γ_2 .

$$\beta_2 = \frac{\mu_4}{\mu_2^2}, \quad \gamma_2 = \beta_2 - 3.$$

Another Definition

If two or more distributions have same average (measure of central tendency), dispersion and skewness, then another measure required to compare the distributions called Kurtosis.

The following figure explains the curve of kurtosis.



Type A curve is neither peaked nor flat is called the normal curve or mesocurtic curve, for such a curve, $\beta_2 = 3$ and $\gamma_2 = 0$.

Type B curve is flatter than the normal curve is known as platykurtic, for this curve $\beta_2 < 3$ and $\gamma_2 < 0$.

Type C curve is more peaked than the normal curve is called leptokurtic, for such a curve $\beta_2 > 3$ and $\gamma_2 > 0$.

PROBLEMS

PROBLEM 1. The first four moments of a distribution about the value 5 of the variable are 2, 20, 40 and 50. Calculate mean, variance, β_1 and β_2 of the distribution.

SOLUTION. Given that, $A = 5$,

$$\mu_1' = 2, \mu_2' = 20, \mu_3' = 40, \mu_4' = 50$$

$$\text{Mean } \bar{x} = \mu_1' + A = 2 + 5 = 7$$

$$\begin{aligned} \text{Variance} = \mu_2 &= \mu_2' - \mu_1'^2 = 20 - (2)^2 = 16 \\ \mu_3 &= \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = 40 - 3(20)(2) + 2(2)^3 = -64 \\ \mu_4 &= \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4 \\ &= 50 - 4(40)(2) + 6(20)(2)^2 - 3(2)^4 = 162 \\ \beta_1 &= \frac{\mu_3^2}{\mu_2^3} = \frac{(-64)^2}{(16)^3} = 1 \\ \beta_2 &= \frac{\mu_4}{\mu_2^2} = \frac{162}{(16)^2} = 0.6328 \end{aligned}$$

PROBLEM 2. The first four moments of a distribution about the value 4 of the variable are -1.5, 17, -30 and 108. Find the moments about mean, β_1 and β_2 . Also find moments about origin. (AU 2016)

SOLUTION. Given that, $A = 4$,

$$\mu_1' = -1.5, \mu_2' = 17, \mu_3' = -30, \mu_4' = 108.$$

Moment about mean

$$\begin{aligned} \mu_2 &= \mu_2' - \mu_1'^2 = 17 - (-1.5)^2 = 14.75 \\ \mu_3 &= \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 \\ &= -30 - 3(17)(-1.5) + 2(-1.5)^3 = 39.75 \\ \mu_4 &= \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4 \\ &= 108 - 4(-30)(-1.5) + 6(17)(-1.5)^2 - 3(-1.5)^4 = 142.3125 \\ \beta_1 &= \frac{\mu_3^2}{\mu_2^3} = \frac{(39.75)^2}{(14.75)^3} = 0.4924 \\ \beta_2 &= \frac{\mu_4}{\mu_2^2} = \frac{142.3125}{(14.75)^2} = 0.6541 \end{aligned}$$

Mean $\bar{x} = \mu_1' + A = -1.5 + 4 = 2.5$

Now moment about origin,

We know $\bar{x} = 2.5, \mu_2 = 14.75, \mu_3 = 39.75, \mu_4 = 142.3125$

and have $A = 0, \bar{x} = 2.5$

then $\mu_1' = \bar{x} - A = 2.5 - 0 = 2.5$

Moments about origin

$$\begin{aligned} \mu_2' &= \mu_2 + \mu_1'^2 = 14.75 + (2.5)^2 = 21 \\ \mu_3' &= \mu_3 + 3\mu_2\mu_1' + \mu_1'^3 = 39.75 + 3(14.75)(2.5) + (2.5)^3 = 166 \\ \mu_4' &= \mu_4 + 4\mu_3\mu_1' + 6\mu_2\mu_1'^2 + \mu_1'^4 \\ &= 142.3125 + 4(39.75)(2.5) + 6(14.75)(2.5)^2 + (2.5)^4 = 1132 \end{aligned}$$

PROBLEM 3. For a distribution the mean is 10, variance is 16, γ_1 is +1 and β_2 is 4. Obtain the first four moments about the origin. Comment the nature of the distribution.

SOLUTION. Given that

Mean $= \bar{x} = 10$, Variance $= \mu_2 = 16$, $\gamma_1 = +1$, $\beta_2 = 4$

First four moments about origin

$A = 0, \mu_1' = \bar{x} - A = 10 - 0 = 10$

and we know

$$\mu_2 = \mu_2' - \mu_1'^2$$

$$16 = \mu_2' - (10)^2$$

$$\mu_2' = 16 + 100 = 116$$

$$\gamma_1 = 1 \Rightarrow \sqrt{\beta_1} = 1, \quad \sqrt{\frac{\mu_3^2}{\mu_2^3}} = 1 \quad \text{or} \quad \frac{\mu_3}{\mu_2^{3/2}} = 1$$

$$\mu_3 = \mu_2^{3/2} = (16)^{3/2} = 64$$

$$\mu_3 = \mu_3' - 3\mu_2' \mu_1' + 2\mu_1'^3$$

$$\begin{aligned} \mu_3' &= \mu_3 + 3\mu_2' \mu_1' - 2\mu_1'^3 \\ &= 64 + 3(116)10 - 2(10)^3 = 1544. \end{aligned}$$

Given, $\beta_2 = 4$

$$\frac{\mu_4}{\mu_2^2} = 4$$

$$\mu_4 = 4 \times 16^2 = 1024$$

and

$$\mu_4 = \mu_4' - 4\mu_3' \mu_1' + 6\mu_2' \mu_1'^2 - 3\mu_1'^4$$

$$\mu_4' = \mu_4 + 4\mu_3' \mu_1' - 6\mu_2' \mu_1'^2 + 3\mu_1'^4$$

$$= 1024 + 4 \times 1544 \times 10 - 6 \times 116 \times (10)^2 + 3(10)^4 = 23184.$$

Comment on the distribution

Since $\gamma_1 = +1$, the distribution is moderately positively skewed.

Since $\beta_2 = 4 > 3$, the distribution is leptokurtic.

PROBLEM 4. Compute the first four central moments to the following data and also find Sheppards correction, β_1 and β_2 .

Class interval	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	2	8	12	40	20	15	3

SOLUTION

Class interval	Fre- quency f_i	Mid- Value x_i	$d_i = \frac{x_i - A}{h}$ $= \frac{x_i - 35}{10}$	$f_i d_i$	$f_i d_i^2$	$f_i d_i^3$	$f_i d_i^4$
0-10	2	5	-3	-6	18	-54	162
10-20	8	15	-2	-16	32	-64	128
20-30	12	25	-1	-12	12	-12	12
30-40	40	35	0	0	0	0	0
40-50	20	45	1	20	20	20	20
50-60	15	55	2	30	60	120	240
60-70	3	65	3	9	27	81	243
	N = 100			$\sum_{i=1}^n f_i d_i$ = 25	$\sum_{i=1}^n f_i d_i^2$ = 169	$\sum_{i=1}^n f_i d_i^3$ = 91	$\sum_{i=1}^n f_i d_i^4$ = 805

$$h = 10, N = \sum_{i=1}^n f_i = 100, d_i = \frac{x_i - A}{h} = \frac{x_i - 35}{10}$$

If h is width of the class interval, then

Non-central moments about point $A = 35$ are

$$\mu_1' = h \cdot \frac{1}{N} \sum_{i=1}^n f_i d_i = 10 \times \frac{25}{100} = 2.5$$

$$\mu_2' = h^2 \cdot \frac{1}{N} \sum_{i=1}^n f_i d_i^2 = 10^2 \times \frac{169}{100} = 169$$

$$\mu_3' = h^3 \cdot \frac{1}{N} \sum_{i=1}^n f_i d_i^3 = 10^3 \times \frac{91}{100} = 910$$

$$\mu_4' = h^4 \cdot \frac{1}{N} \sum_{i=1}^n f_i d_i^4 = 10^4 \times \frac{805}{100} = 80,500$$

Central moments

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - \mu_1'^2 = 169 - (2.5)^2 = 162.75$$

$$\mu_3 = \mu_3' - 3\mu_2' \mu_1' + 2\mu_1'^3 = 910 - 3(169)(2.5) + 2(2.5)^3 = -326.25$$

$$\begin{aligned} \mu_4 &= \mu_4' - 4\mu_3' \mu_1' + 6\mu_2' \mu_1'^2 - 3\mu_1'^4 \\ &= 80,500 - 4(910)(2.5) + 6(169)(2.5)^2 - 3(2.5)^4 = 77698.4375 \end{aligned}$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-326.25)^2}{(162.75)^3} = 0.0247$$

$\beta_1 > 0$, the distribution is positively skewed. (Since $\mu_3 > 0$)

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{77698.4375}{(162.75)^2} = 2.93$$

$\beta_2 = 2.93 < 3$, the distribution is platykurtic distribution.

Sheppard's corrections

$$\mu_2(\text{corrected}) = \mu_2 - \frac{h^2}{12} = 162.75 - \frac{(10)^2}{12} = 154.42$$

$$\mu_3(\text{corrected}) = \mu_3 = -326.25$$

$$\begin{aligned} \mu_4(\text{corrected}) &= \mu_4 - \frac{h^2}{2} \cdot \mu_2 + \frac{7}{240} \cdot h^4 \\ &= 77698.4375 - \frac{(10)^2}{2} \cdot (162.75) + \frac{7}{240} \cdot (10)^4 = 69852.57 \end{aligned}$$

① Relation between Central moments in terms of non-Central moments:

By the definition of Central moments.

$$\mu_r = \frac{1}{N} \sum_{i=1}^n f_i (x_i - \bar{x})^r$$

$$= \frac{1}{N} \sum_{i=1}^n f_i (x_i - A + A - \bar{x})^r$$

$$= \frac{1}{N} \sum_{i=1}^n f_i (d_i - A_i)^r$$

$$= \frac{1}{N} \sum_{i=1}^n f_i [{}^r C_0 d_i^r \mu_0 - {}^r C_1 d_i^{r-1} \mu_1 + {}^r C_2 d_i^{r-2} \mu_2 + \dots]$$

$$= {}^r C_0 \frac{1}{N} \sum_{i=1}^n f_i d_i^r \mu_0 - {}^r C_1 \frac{1}{N} \sum_{i=1}^n f_i d_i^{r-1} \mu_1 + {}^r C_2 \frac{1}{N} \sum_{i=1}^n f_i d_i^{r-2} \mu_2 + \dots$$

$$= ({}^r C_0) \mu_0 - {}^r C_1 \mu_{r-1} \mu_1 + {}^r C_2 \mu_{r-2} \mu_2^{(1)}$$

$$\mu_r = \mu_0 - {}^r C_1 \mu_{r-1} \mu_1 + {}^r C_2 \mu_{r-2} \mu_2^{(1)} \quad \text{--- (1)}$$

Now, If $r=0 \Rightarrow \mu_0 = \mu_0 = 1$

$$r=1 \Rightarrow \mu_1 = \mu_1 - {}^1 C_1 \mu_{1-1} \mu_1 \Rightarrow \mu_1 - (1) \mu_0 \mu_1 \Rightarrow \mu_1 - \mu_1 = 0$$

$$r=2 \Rightarrow \mu_2 = \mu_2 - {}^2 C_1 \mu_{2-1} \mu_1 + {}^2 C_2 \mu_{2-2} \mu_2^{(1)} \Rightarrow \mu_2 - 2\mu_1 \mu_1 + 1 \mu_0 \mu_2^{(1)} \Rightarrow \mu_2 - 2\mu_1^2 + \mu_2^{(1)} \Rightarrow \mu_2 - \mu_1^{(2)}$$

$$r=3 \Rightarrow \mu_3 = \mu_3 - {}^3 C_1 \mu_{3-1} \mu_1 + {}^3 C_2 \mu_{3-2} \mu_2^{(1)} + {}^3 C_3 \mu_{3-3} \mu_3^{(2)} \Rightarrow \mu_3 - 3\mu_2 \mu_1 + 3\mu_1 \mu_2^{(1)} - 1 \mu_0 \mu_3^{(2)}$$

$$\Rightarrow \mu_3 - 3\mu_2 \mu_1 + 3\mu_1 \mu_2^{(1)} - \mu_3^{(2)}$$

$$\Rightarrow \mu_3 - 3\mu_2 \mu_1 + 2\mu_1^{(3)}$$

$$r=4 \Rightarrow \mu_4 = \mu_4 - {}^4 C_1 \mu_{4-1} \mu_1 + {}^4 C_2 \mu_{4-2} \mu_2^{(1)} - {}^4 C_3 \mu_{4-3} \mu_3^{(2)} + {}^4 C_4 \mu_{4-4} \mu_4^{(3)}$$

$$= \mu_4 - 4\mu_3 \mu_1 + 6\mu_2^2 \mu_1^{(1)} - 4\mu_1 \mu_3^{(2)} + 1 \mu_0 \mu_4^{(3)} \Rightarrow \mu_4 - 4\mu_3 \mu_1 + 6\mu_2^2 \mu_1^{(1)} - 4\mu_1 \mu_3^{(2)} + (1) (1) \mu_4^{(3)}$$

$$\Rightarrow \mu_4 - 4\mu_3 \mu_1 + 6\mu_2^2 \mu_1^{(1)} - 3\mu_1^{(4)}$$

[∴ where $d_i = x_i - A$,
 $\mu_1 = \bar{x} - A$
 $A - \bar{x} = -\mu_1$
 $\therefore \sum_{i=1}^n f_i = N$
 $[∴ \mu_0 = 1$
 $4 \mu_1 = 1]$
 $\therefore \frac{1}{N} \sum_{i=1}^n f_i = \mu_1$

(i) Relation between non-central moments in terms of central moments:

By the definition of non-central moments.

$$\mu'_1 = \frac{1}{N} \sum_{i=1}^n f_i (x_i - A)^1$$

$$= \frac{1}{N} \sum_{i=1}^n f_i (x_i - \bar{x} + \bar{x} - A)^1$$

$$= \frac{1}{N} \sum_{i=1}^n f_i (z_i + \mu_1)^1$$

$$= \frac{1}{N} \sum_{i=1}^n f_i [\gamma c_0 z_i^1 \mu_1^0 + \gamma c_1 z_i^1 \mu_1^1 + \gamma c_2 z_i^1 \mu_1^2 + \dots]$$

$$= \gamma c_0 \frac{1}{N} \sum_{i=1}^n f_i z_i^1 \mu_1^0 + \gamma c_1 \frac{1}{N} \sum_{i=1}^n f_i z_i^1 \mu_1^1 + \gamma c_2 \frac{1}{N} \sum_{i=1}^n f_i z_i^1 \mu_1^2 + \dots$$

$$\mu'_1 = (1) \mu_1^{(1)} + \gamma c_1 \mu_{1-1} \mu_1^1 + \gamma c_2 \mu_{1-2} \mu_1^2 + \dots \quad \text{--- (1)}$$

Now
If $\gamma=0 \Rightarrow \mu'_0 = \mu_0 = 1$

$$\gamma=1 \Rightarrow \mu'_1 = \mu_1 + \gamma c_1 \mu_{1-1} \mu_1^1 \Rightarrow \mu_1 + (1) \mu_0 \mu_1^1 \Rightarrow \mu_1 + \mu_1$$

$$\gamma=2 \Rightarrow \mu'_2 = \mu_2 + 2c_1 \mu_{2-1} \mu_1^1 + 2c_2 \mu_{2-2} \mu_1^2 \Rightarrow \mu_2 + 2(1) \mu_1 \mu_1^1 + (1) \mu_0 \mu_1^2 \Rightarrow \mu_2 + 2\mu_1 \mu_1^1 + \mu_0 \mu_1^2$$

$$\Rightarrow \mu_2 + \mu_1 \mu_1^1 \Rightarrow \mu_2 + \mu_1^2$$

$$\gamma=3 \Rightarrow \mu'_3 = \mu_3 + 3c_1 \mu_{3-1} \mu_1^1 + 3c_2 \mu_{3-2} \mu_1^2 + 3c_3 \mu_{3-3} \mu_1^3 \Rightarrow \mu_3 + 3\mu_2 \mu_1^1 + 3(1) \mu_1 \mu_1^2 + (1) \mu_0 \mu_1^3$$

$$\Rightarrow \mu_3 + 3\mu_2 \mu_1 + 3(1) \mu_1^2 + (1) \mu_0^3$$

$$\gamma=4 \Rightarrow \mu'_4 = \mu_4 + 4c_1 \mu_{4-1} \mu_1^1 + 4c_2 \mu_{4-2} \mu_1^2 + 4c_3 \mu_{4-3} \mu_1^3 + 4c_4 \mu_{4-4} \mu_1^4 \Rightarrow \mu_4 + 4\mu_3 \mu_1^1 + 6\mu_2 \mu_1^2 + 2(1) \mu_1^3 + (1) \mu_0 \mu_1^4$$

$$\Rightarrow \mu_4 + 4\mu_3 \mu_1 + 6\mu_2 \mu_1^2 + 2(1) \mu_1^3 + (1) \mu_0 \mu_1^4$$

$$\Rightarrow \mu_4 + 4\mu_3 \mu_1 + 6\mu_2 \mu_1^2 + 2\mu_1 \mu_1^3 + \mu_0 \mu_1^4$$

Where
 $z_i = x_i - \bar{x}$
 $\mu_1 = \bar{x} - A$

$\therefore \frac{1}{N} \sum_{i=1}^n f_i z_i^1 = \mu_1$

$$[\therefore \mu_1 = 1]$$