4 Measures of Dispersion

$0 \leq P(A) \leq 1$

data

4.1 Introduction

The measure of central tendency does not tell us any thing about the spread of the values in a set, because any two sets with vast difference in magnitude of their variability may have the same central tendency. Look at the following two data sets:

Data set a: 8, 7, 5, 8, 6 Data set b: 1, 4, 7, 10, 12

These two data sets have same mean 6.8 but differ in their variations from the central value. There is more variation in the date set b as compared to the data set a. This illustrates the fact that measure of central tendency is not sufficient.

To give a sensible description of data, a numerical quantity called measure of dispersion or variability that describes the spread of the values in a set of data is required.

Two types of measures of dispersion or variability are defined:

i) Absolute measures ii) Relative measures

The absolute measures are defined in such a way that they have units (meters, grams etc) same as those of the original measurements, whereas the relative measures have no units as these are ratios.

The most common measures of absolute variability are:

a) Range b) Quartile Deviation

c) Mean Deviation d) Variance

e) Standard Deviation

These are also called measures of dispersion or measures of spread.

The relative measures are discussed in article 4.2

4.1.1 Range

The range of n values $Y_1, Y_2, \ldots, Y_n$ is defined as the difference between the largest and smallest observation. If $Y_{(1)}$ is smallest in magnitude and $Y_{(n)}$ is largest in magnitude then range denoted by R, is defined by;

$$R = Y_{(n)} - Y_{(1)} \qquad (4.1)$$

This is very simple measure of variability and only takes into account two most extreme observations.

Example 4.1: Calculate range for the following observations (in cms):

84.2, 87.5, 80.7, 92.4, 91.9, 86.5, 85.4

Solution: Here $Y_{(1)} = 80.7 \text{ and } Y_{(7)} = 92.4$

so $R = 92.4 - 80.7$

$= 11.7cm$

Range for Grouped Data: In case of grouped data, it may be calculated by the following formula:

R = mid value of the highest class – mid value of the lowest class

Example 4.2: The following frequency Distribution gives the weights of 90 cotton bales.

Weights	70 – 74	75 – 79	80 – 84	85 – 89	90 – 94	95 – 99
Frequencies	1	7	17	29	20	16

Find its range.

Solution:

Weights	f	Class boundary
70 – 74	1	69.5 – 74.5
75 – 79	7	74.5 – 79.5
80 – 84	17	79.5 – 84.5
85 – 89	29	84.5 – 89.5
90 – 94	20	89.5 – 94.5
95 – 99	16	94.5 – 99.5
Total	90	- -

The mid value of first group is $\frac{69.5+74.5}{2} = 72$ and the mid value of last group is $\frac{94.5+99.5}{2} = 97$, so

$Y_{(n)} = 97.0 = Y_{(1)} = 72.0$

$R = Y_{(n)} - Y_{(1)} = 97.0 - 72.0 = 25$

Merits of range:

i) It is easy to calculate.

ii) It is a useful measure in small samples.

Demerits of range:

i) It is not based on all the observations.

ii) It depends only upon the extreme observations.

4.1.2 Quartile Deviation

This measure is based on quartiles Q_1 and Q_3 and is denoted by Q.*D*. It is calculated as

$$Q.D = \frac{Q_3 - Q_1}{2} \qquad (4.2)$$

It is also known as semi inter quartile range.

This measure cannot be negative because the upper quartile must be atleast as large as the lower quartile. A small value of quartile deviation indicates a small amount of variability whereas larger values indicate more variability in the data set. It measures half of the difference between the upper and lower quartiles.

Example 4.3: Calculate quartile deviation for the data of the example 3.16.

Solution: $Q_1 = 88.03$ cms, $Q_3 = 94.90$ cms

Therefore, $$Q.D = \frac{94.90 - 88.03}{2}$$

$$= 3.435 \text{ cms}$$

The formula of quartile deviation for grouped data is the same as for ungrouped data. i.e.,

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

Example 4.4: Calculate quartile deviation for the data of the example 3.17

Solution: $Q_1 = 92.3750,\ Q_3 = 102.5833$

$$Q.D = \frac{102.5833 - 92.3750}{2} = 5.104 \text{ cm}$$

Example 4.5: For the data given below:

1030, 1590, 1070, 1670, 1110, 1710, 1190, 1720, 1230, 1740, 1310, 1745, 1332, 1775, 1870, 1350, 1430, 1870, 1950 and 1460,

calculate Quartile Deviation and co-efficient of Quartile Deviation.

Solution: Arraying the data

1030, 1070, 1110, 1190, 1230, 1310, 1332, 1350, 1430, 1460, 1590, 1670, 1710, 1720, 1740, 1745, 1775, 1870, 1870, 1950.

Here, $n = 20$

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

$$= \text{Value of}\left(\frac{20+1}{4}\right)\text{th item}$$

$$= \text{Value of } (5.25) \text{ th item}$$

$$\therefore Q_1 = 5\text{th value} + 0.25\,(6\text{th value} - 5\text{th value})$$

$$= 1230 + 0.25(1310 - 1230)$$

$$= 1250$$

$$Q_3 = \text{Value of } 3\left(\frac{n+1}{4}\right)\text{th item}$$

$$= \text{Value of } 3\left(\frac{20+1}{4}\right)\text{th item}$$

$$= \text{Value of } 15.75 \text{ th item}$$

$$\therefore Q_3 = 15\text{th value } + 0.75\,(16\text{th value } - 15\text{th value})$$

$$= 1740 + 0.75(1745 - 1740)$$

$$= 1743.75$$

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

$$= \frac{1743.75 - 1250}{2} = 246.88$$

$$\text{Co-efficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{1743.75 - 1250}{1743.75 + 1250} = 0.16$$

Example 4.6: For the following frequency distribution, find quartile deviation and co-efficient of quartile deviation.

Marks	**10 – 20**	**20 – 30**	**30 – 40**	**40 – 50**	**50 – 60**
Frequency	3	8	14	7	4

Solution:

Marks	f	Cumulative frequency (C.F.)
10 – 20	3	3
20 – 30	8	3 + 8 = 11
30 – 40	14	11 + 14 = 25
40 – 50	7	25 + 7 = 32
50 – 60	4	32 + 4 = 36
Total	36	

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - c\right)$$

$n = 36, \ \frac{n}{4} = \frac{36}{4} = 9$th observation so, $l = 20, \ h=10, \ f=8, \ c=3$

$$\therefore Q_1 = 20 + \frac{10}{8}(\frac{36}{4} - 3) = 27.5,$$

$n = 36, \ \frac{3n}{4} = \frac{3\times 36}{4} = 27$th observation so, $l = 40, \ h=10, \ f=7, \ c=25$

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - c\right)$$

$$\therefore Q_3 = 40 + \frac{10}{7}\left(3.\frac{36}{4} - 25\right) = 42.8$$

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

$$= \frac{42.8 - 27.5}{2} = 7.65$$

$$\text{Co - efficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{42.8 - 27.5}{42.8 + 27.5} = 0.22$$

Example 4.7: Find the Semi – Interquartile range and co-efficient of Quartile deviation for the data given below about the ages in a locality.

Ages	20	30	40	50	60	70	80
Frequency	3	61	132	158	140	51	2

Solution:

y_i	f	C.B.	c.f.
20	3	15 – 25	3
30	61	25 – 35	64
40	132	35 – 45	196
50	158	45 – 55	354
60	140	55 – 65	494
70	51	65 – 75	545
80	2	75 – 85	547
	547		

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - c\right)$$

$$n = 547, \frac{n}{4} = \frac{547}{4} = 136.7 = 137\text{th observation so,}$$

$$l = 35,\ h = 10,\ f = 132,\ c = 64$$

$$Q_1 = 35 + \frac{10}{132}\left(\frac{547}{4} - 64\right) = 40.51$$

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - c\right)$$

$$n = 547, \frac{3n}{4} = \frac{3 \times 547}{4} = 410.1 = 410\text{th observation so,}$$

$$l = 55,\ h = 10,\ f = 140,\ c = 354$$

$$\therefore Q_3 = 55 + \frac{10}{140}\left(3 \times \frac{547}{4} - 354\right) = 59.02$$

$$\text{Semi - Inter Quartile Range} = \frac{Q_3 - Q_1}{2} = \frac{59.02 - 40.15}{2} = 9.255$$

$$\text{Co - efficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{59.02 - 40.15}{59.02 + 40.15} = \frac{18.87}{99.17} = 0.19$$

Merits of quartile deviation

i) It is easy to calculate.

ii) It is not affected by extreme observations.

Demerits of quartile deviation

i) It is not based on all the observations.

ii) Q.D. will be the same value for all the distributions having the same quartiles.

4.1.3 Mean deviation

It is defined as the mean of the absolute deviations of observations from mean, median or mode. By absolute deviations we mean that we consider all the deviations as positive. It is denoted by M.D. and is calculated as

$$M.D = \frac{\Sigma|Y - M|}{n} \quad \text{(M is mean or median or mode)} \qquad (4.3)$$

Example 4.8: For ungrouped data of the example 3.13, find mean deviation

Solution: Y_i : 88.03, 94.50, 94.90, 95.50, 84.60.

$$\Sigma Y = 88.03 + 94.50 + ... + 84.60 = 457.08$$

$$\overline{Y} = \frac{\Sigma Y_i}{n}$$

$$= \frac{457.08}{5} = 91.416$$

Mean Deviation (M. D) = $\dfrac{\Sigma|Y_i - \overline{Y}|}{n}$

Where $|Y_i - \overline{Y}|$ are

$$|88.03 - 91.416| = 3.386$$

$$|94.50 - 91.416| = 3.084$$

$$|94.90 - 91.416| = 3.484$$

$$|95.05 - 91.416| = 3.634$$

$$|64.60 - 91.416| = 6.816$$

$$\Sigma|Y_i - \overline{Y}| = 20.404$$

$$\text{M.D.} = \frac{20.404}{5} = 4.0808\,\text{cm}$$

Dealing with the grouped data, mean deviation is calculated by multiplying the absolute deviations from mean with the corresponding frequencies and then taking the mean i.e.,

$$\text{M.D.} = \frac{\Sigma f_i \left|Y_i - \bar{Y}\right|}{\Sigma f_i}$$

Example 4.9: For the data of the example 3.2, find mean deviation for grouped data.

Solution:

$$\bar{Y} = \frac{\Sigma f_i\, Y_i}{\Sigma f_i} = 97.8333 \text{ cm}$$

C. I.	Y_i	f_i	$(Y_i - \bar{Y})$	$f_i \left|Y_i - \bar{Y}\right|$
86 – 90	88	6	-9.8333	58.9998
91 – 95	93	4	-4.8333	19.3332
96 – 100	98	10	0.1667	1.6670
101 – 105	103	6	5.1667	31.0002
106 – 110	108	3	10.1667	30.5001
111 – 115	113	1	15.1667	15.1667
Total		30		156.6670

$$\text{Mean deviation (M.D)} = \frac{\Sigma f_i \left|Y_i - \bar{Y}\right|}{\Sigma f_i}$$

$$= \frac{156.6670}{30}$$

$$= 5.2222 cm$$

Mean deviation from median is defined in terms of absolute deviations from median as:

$$\text{M.D.} = \frac{\Sigma \left|Y_i - \text{Median}\right|}{n} \qquad (4.4)$$

The mean deviation from median for the data of the example 3.13 is calculated as:

$\|Y_i - \text{Median}\|$	
$\|88.03 - 94.50\|$	$= 6.47$
$\|94.50 - 94.50\|$	$= 0.0$
$\|94.90 - 94.50\|$	$= 0.40$
$\|95.05 - 94.50\|$	$= 0.55$
$\|84.60 - 94.50\|$	$= 9.90$
$\Sigma\|Y_i - \text{median}\|$	$= 17.32$

So $\quad \text{M.D} = \dfrac{17.32}{5}$

$= 3.464$

The mean deviation from median for grouped data is:

$$\text{M.D} = \frac{\Sigma f_i \left|Y_i - \text{Median}\right|}{\Sigma f_i}$$

The calculations for mean deviation from median taking median equals to 98 for the example 4.9 are:

Y_i	f_i	$Y_i - 98$	$\|Y_i - 98\|$	$f_i \| Y_i - 98\|$
88	6	-10	10	60
93	4	-5	5	20
98	10	0	0	0
103	6	-5	5	30
108	3	-10	10	30
113	1	-15	15	15
Total	30	---	---	155

$$\text{M.D} = \frac{155}{30} = 5.167 \text{ cm}$$

Properties of mean deviation

i) M .D from median is less than any other value i.e.,

$$\frac{\Sigma\left|Y_i - \text{median}\right|}{n} \text{ is least}$$

ii) It is always greater than or equal to zero i.e.,

$\text{M.D} \geq 0$

iii) For symmetrical distributions, the following relation holds

$$\text{M.D} = \frac{4}{5}\sigma \qquad (4.5)$$

Where σ is the standard deviation.

Merits of mean deviation

i) It is easy to calculate.

iii) It is based on all the observations.

Demerits of mean deviation

i) It is affected by the extreme values.

ii) It is not readily capable of mathematical development.

iii) It does not take into account the negative signs of the deviations from some average.

Example 4.10 Find mean deviation from median for the following frequency distrib ution.

Ages (Years)	5-10	10-15	15-20	20-25
Frequency	10	20	30	15

Also calculate the co-efficient of mean deviation.

Solution:

Ages	f	$c.f.$	Y_i	$\lvert Y_i - \overline{Y}\rvert$	$f_i \lvert Y_i - \overline{Y}\rvert$
5 – 10	10	10	7.5	8.75	87.5
10 – 15	20	30	12.5	3.75	75
15 – 20	30	60	17.5	1.25	37.5
20 – 25	15	75	22.5	6.25	93.75
Total:	75	-	-	-	293.75

$$\tilde{Y} = l + \frac{h}{f}\left(\frac{n}{2} - c\right)$$

$$\tilde{Y} = 15 + \frac{5(37.5-30)}{30} = 16.25$$

$$\text{Mean Deviation from median} = \frac{\Sigma f\lvert Y - \tilde{Y}\rvert}{\Sigma f} = \frac{293.75}{75} = 3.92$$

$$\text{Co-efficient of mean deviation} = \frac{\text{Mean Deviation from median}}{\text{Median}}$$

$$= \frac{3.92}{16.25} = 0.24$$

4.1.4 The Variance

Variance of the observations is defined as mean of squares of deviations of all the observations from their mean . When it is calculated from the population the variance is called population variance and is denoted by σ^2 and when it is calculated from the sample, based on n values $Y_1, Y_2, \ldots, Y_n$ is called sample variance. The

Population variance σ^2 is defined as $\sigma^2 = \dfrac{\sum (y_i - \mu)^2}{N}$

The sample variance S^2 for un-grouped data is defined as:

$$S^2 = \frac{\Sigma (y - \bar{y})^2}{n} \quad (4.6)$$

Short formula for variance is given by

$$S^2 = \frac{\Sigma y^2}{n} - \left(\frac{\Sigma y}{n}\right)^2 \quad (4.7)$$

For a frequency distribution, the sample variance S^2 is defined as:

$$S^2 = \frac{\Sigma f_i \ (y_i - \bar{y})^2}{\Sigma f_i} \quad (4.8)$$

Short formula for grouped data is given by

$$S^2 = \frac{\Sigma f y^2}{\Sigma f} - \left(\frac{\Sigma f y}{\Sigma f}\right)^2 \quad (4.9)$$

If A is an arbitrary value such that $D = y - A$, S^2 for ungrouped data is given by

$$S^2 = \frac{\Sigma D^2}{n} - \left(\frac{\Sigma D}{n}\right)^2 \quad (4.10)$$

for grouped data

$$S^2 = \frac{\Sigma fD^2}{\Sigma f} - \left(\frac{\Sigma fD}{\Sigma f}\right)^2 \tag{4.11}$$

When data are grouped into a frequency distribution with equal class intervals of size h and $u = \frac{y-A}{h}$, then

$$S^2 = h^2\left[\frac{\Sigma fu^2}{\Sigma f} - \left(\frac{\Sigma fu}{\Sigma f}\right)^2\right] \tag{4.12}$$

Merits and Demerits of Variance.

The variance is based on all the observations of a series. It is easy to calculate and simple to understand. It is affected by extreme values.

4.1.5 Standard Deviation

The standard deviation is defined as the positive square root of the mean of the squares of the deviations of values from their mean. In other words, standard deviation is a positive square root of variance. It is denoted by S and is given by

For ungrouped data

$$S = \sqrt{\frac{\Sigma(Y-\bar{Y})^2}{n}} \tag{4.13}$$

In short cut method

$$S = \sqrt{\frac{\Sigma Y^2}{n} - \left(\frac{\Sigma Y}{n}\right)^2} \tag{4.14}$$

For frequency distribution

$$S = \sqrt{\frac{\Sigma f(Y-\bar{Y})^2}{\Sigma f}} \tag{4.15}$$

In short cut method

$$S = \sqrt{\frac{\Sigma fY^2}{\Sigma f} - \left(\frac{\Sigma fY}{\Sigma f}\right)^2} \tag{4.16}$$

If $D = Y - A$, the deviations of Y from any arbitrary value A then standard deviation is

$$S = \sqrt{\frac{\Sigma D^2}{n} - \left(\frac{\Sigma D}{n}\right)^2} \tag{4.17}$$

For frequency distribution, the formula becomes

$$S = \sqrt{\frac{\Sigma f D^2}{\Sigma f} - \left(\frac{\Sigma fD}{\Sigma f}\right)^2} \tag{4.18}$$

For coding variable $u = \frac{y - A}{h}$, the formula becomes

$$S = h\sqrt{\frac{\Sigma f u^2}{\Sigma f} - \left(\frac{\Sigma f u}{\Sigma f}\right)^2} \tag{4.19}$$

Example 4.11: Calculate variance and standard deviation for the data: 3,6,2,1,7,5.

Solution:

Y	$Y - \bar{Y}$	$(Y - \bar{Y})^2$
3	-1	1
6	2	4
2	-2	4
1	-3	9
7	3	9
5	1	1
24	0	28

$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{24}{6} = 4$$

$$\text{Variance} = S^2 = \frac{\Sigma(Y - \bar{Y})^2}{n} = \frac{28}{6} = 4.67$$

$$\text{Standard deviation} = S = \sqrt{\frac{\Sigma(Y - \bar{Y})^2}{n}}$$

$$= \sqrt{4.67}$$

$$= 2.16$$

Example 4.12: Calculate variance and standard deviation from the following frequency distribution.

Wages	30-35	35-40	40-45	45-50	50-55	55-60
Frequency	12	18	29	32	16	8

Solution:

Wages	f	y	fy	$f(y-\bar{y})^2$
30 – 35	12	32.5	390	1723
35 – 40	18	37.5	675	882
40 – 45	29	42.5	1232.5	116
45 – 50	32	47.5	1520	288
55 – 60	16	52.5	840	1024
60 – 65	8	57.5	460	1352
Total:	115		5117.5	5390

$$\bar{Y} = \frac{\Sigma f Y}{\Sigma f} = \frac{5117.5}{115}$$

$$\text{variance}\,(S)^2 = \frac{\Sigma f\,(y-\bar{y})^2}{\Sigma f}$$

$$= \frac{5390}{115} = 46.87$$

$$S = \sqrt{46.87}$$

$$= 6.846$$

Example 4.13: Calculate variance and standard deviation by using any provisional mean from the data: 3,5,7,13,15,17,23,27.

Solution:

Y	$D = y - 15$	D^2
3	-12	144
5	-10	100
7	-8	64
13	-2	4
15	0	0
17	2	4
23	8	64
27	12	144
Total	-10	524

$$\text{Variance}\quad (S^2) = \frac{\sum D^2}{n} - \left(\frac{\sum D}{n}\right)^2$$

$$= \frac{524}{8} - \left(\frac{-10}{2}\right)^2$$

$$= 65.6 - 1.562$$

$$= 63.938$$

$$\text{Standard Deviation }(S) = \sqrt{\frac{\sum D^2}{n} - \left(\frac{\sum D}{n}\right)^2}$$

$$= \sqrt{63.938} = 7.99$$

PROPERTIES OF THE VARIANCE AND STANDARD DEVIATION

1. The variance and standard deviation of a constant is zero. If a is a constant, then

   var (a) = 0

   S.D. (a) = 0

2. The variance and standard deviation are independent of origin.

   var $(y + a)$ = var (y)

$$\text{var}(y-a) = \text{var}(y)$$

and $$\text{S.D.}(y+a) = \text{S.D.}(y)$$

$$\text{S.D.}(y-a) = \text{S.D.}(y)$$

3. When all the values are multiplied with a constant, the variance of the values is multiplied by square of the constant and their standard deviation is multiplied by the constant i.e.,

$$\text{var}(ay) = a^2, \text{var}(y)$$

$$\text{var}(y/a) = (1/a^2)\,\text{var}(y)$$

and $$\text{S.D}(ay) = |a|\,\text{S.D}(y)$$

$$\text{S.D}(y/a) = |1/a|\,\text{S.D}(y)$$

4. The variance/standard deviation of the sum or difference of two independent variables is the sum of their respective variances/standard deviation for independent variables x and y.

$$\text{var}(x+y) = \text{var}(x) + \text{var}(y)$$

$$\text{var}(x-y) = \text{var}(x) + \text{var}(y)$$

and $$\text{S.D.}(x+y) = \text{S.D.}(x) + \text{S.D.}(y)$$

$$\text{S.D.}(x-y) = \text{S.D.}(x) + \text{S.D.}(y)$$

5. If sets of data consisting $n_1, n_2, \ldots, n_k$ values having corresponding means $\bar{y}_1$, $\bar{y}_2, \ldots, \bar{y}_k$ and variances $S_1^2, S_2^2, \ldots, S_k^2$, the variance of combined set of data is given by

$$S_c^2 = \frac{\sum n_i \left[S_i^2 + (\bar{y}_i - \bar{y}_c)^2\right]}{\sum n_i} \qquad (4.20)$$

where, $\bar{y}_c$[1] $= \dfrac{n_1\bar{y}_1 + n_1\bar{y}_2 + \ldots + n_k\bar{y}_k}{n_1 + n_2 + \ldots + n_k}$

4.2 Co-efficient Of Variation and Other Relative Measures

The most important of all the relative measures of dispersion is co-efficient of variation. Co-efficien of variation is a relative measure of dispersion and independent of units of measurement and expressed in percentage. It is used to compare the variability of different sets of data. The group which has lower value of

[1] $\bar{y}_c$ stands for combined mean notation

co-efficient of variation is comparatively more consistent. The co-efficient of variation is defined as:

$$\text{Co-efficient of variation} = \text{C.V.} = \frac{S}{\bar{y}} \times 100 \text{ (for sample)}$$

$$\text{C.V} = \frac{\sigma}{\mu} \times 100 \text{ (for population)}$$

As it is a ratio of the two quantities with the same units, so is a dimensionless quantity i.e., for the same data whether it is in millimeters, centimeters or meters, etc. The co-efficient of variation remains the same and has no unit.

As the co-efficient of variation expresses variability relative to the mean, it is called a measure of relative variability or relative dispersion.

The co-efficient of variability for the example 4.7 is given by

$$\text{C.V} = \frac{S}{\bar{y}} \times 100$$

$$S = 6.8837$$

$$\bar{y} = 97.8833$$

so $$\text{C.V.} = \frac{6.8837}{97.8833} \times 100 = 7.04\%$$

Large value of C.V indicates that the observations have much spread relative to the size of the mean and vice versa.

This measure can be used to compare the variability of two or more populations. It will take the same value for two or more populations if in each population, the standard deviation is directly proportional to the mean. In such situation, we say that two or more populations are consistent. For example, to compare the consistency of two methods, each method was tried on 16 soil samples and the corresponding results obtained are:

	Method I	Method II
$\bar{y}$	15.0	10.5
S	1.4	1.1
CV(%)	9.3	9.5

The CVs being almost equal indicate that both the methods are equally reliable. We actually do not compare the standard deviations, since the means will apparently be widely different.

Some other relative measures of dispersion are:

a) Co-efficient of Range = $\dfrac{Y_{(n)} - Y_{(1)}}{Y_{(n)} + Y_{(1)}}$ (4.21)

b) Co-efficient of Q.D. = $\dfrac{Q_3 - Q_1}{Q_3 - Q_1}$ (4.22)

c) Mean co-efficient of dispersion = $\dfrac{M.D}{\bar{y}}$ (4.23)

d) Median co-efficient of dispersion = $\dfrac{M.D}{Median}$ (4.24)

e) Co-efficient of standard deviation = $\dfrac{S}{\bar{y}}$ (4.25)

Example 4.14: Calculate co-efficient of variation and co-efficient of standard deviation from the following frequency distribution.

y	0	1	2	3	4
f	17	9	6	5	3

Solution:

y	f	fy	$y-\bar{y}$	$f(y-\bar{y})^2$
0	17	0	1.2	24.48
1	9	9	0.2	0.36
2	6	12	0.8	3.81
3	5	15	1.8	16.2
4	3	12	2.8	23.52
Total	40	48	--	68.4

$$\bar{y} = \frac{\sum fy}{\sum f} = \frac{48}{40} = 1.2$$

$$S = \sqrt{\frac{\sum f(y-\bar{y})^2}{\sum f}}$$

$$= \sqrt{\frac{68.4}{40}} = 1.307$$

$$\text{Co-efficient of S.D} = \frac{\text{Standard deviation}}{\text{Mean}} = \frac{1.307}{1.2} = 1.089$$

$$\text{Co-efficient of variation} = \frac{\text{S.D.}}{\bar{y}} \times 100 = \frac{1.307}{1.2} \times 100 = 108.92\%$$

Example 4.15: Given the following results, find the combined co-efficient of variation.

$n_1 = 100$	$S_1 = 2.4$	$\bar{y}_1 = 12.6$
$n_2 = 120$	$S_2 = 4.2$	$\bar{y}_2 = 15.8$
$n_3 = 150$	$S_3 = 3.7$	$\bar{y}_3 = 10.5$

Solution: Combined mean is given by

$$\bar{y}_c = \frac{n_1\bar{y}_1 + n_2\bar{y}_2 + n_3\bar{y}_3}{n_1 + n_2 + n_3} = \frac{100(12.5) + 120(15.8) + 150(10.5)}{100+120+150}$$

$$= 12.76$$

Combined variance is given by

$$S_c^2 = \frac{n_1\left[S_1^2 + (\bar{y}_1 - \bar{y}_c)^2\right] + n_2\left[S_2^2 + (\bar{y}_2 - \bar{y}_c)^2\right] + n_3\left[S_3^2 + (\bar{y}_3 - \bar{y}_c)^2\right]}{n_1 + n_2 + n_3}$$

$$= \frac{100\left[5.76 + (12.5 - 12.76)^2\right] + 120\left[17.64 + (15.8 - 12.76)^2\right] + 150\left[13.69 + (10.5 - 12.76)^2\right]}{n_1 + n_2 + n_3}$$

$$= \frac{6628.19}{370}$$

$$S^2_c = 17.9140, \quad S_c = 4.23255$$

$$\therefore \; C.Vc = \frac{S_c}{\bar{Y}_c} \times 100$$

$$= \frac{4.2325}{12.76} \times 100 = 33.17\%$$

4.3 Moments

The measures of location alongwtih measures of variability are useful to describe a data set but fail to tell anything about the shape of the distribution. For this purpose, we need to define certain other measures. Some important measures about the shape of the distribution depend upon what we call moments. These measures are discussed under skewness and kurtosis.

4.3.1 Moments about mean

The moments about mean are the mean of deviations from the mean after raising them to integer powers. The rth population moment about the mean is denoted by μ_r is defined as:

For ungrouped data

$$\mu_r = \frac{\sum_{i=1}^{N}(y_i - \mu)^r}{N} \qquad (4.26)$$

Where, $r = 1,2, \ldots.$

and the corresponding sample moments about mean $\bar{y}$, denoted by m_r is given by

$$m_r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^r}{n} \qquad (4.27)$$

So, the first moment m_1 is given by

$$m_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^1}{n}$$

m_1 is always zero as the numerator $\sum_{i=1}^{n}(y_i - \bar{y}) = 0$

Second moment m_2 is given by: $m_2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}$

This is the same as variance S^2.

Third moment m_3 is given by: $m_3 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^3}{n}$

and fourth moment m_4 is given by: $m_4 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^4}{n}$

If the data are grouped then the rth sample moment about the mean $\bar{y}$ is

defined as: $$m_r = \frac{\sum_{i=1}^{n} f_i (y_i - \bar{y})^r}{n}$$

$$\text{where } n = \sum_{i=1}^{n} f_i \qquad (4.28)$$

$$\text{so } m_1 = \frac{\sum_{i=1}^{n} f_i (y_i - \bar{y})^1}{n} = \frac{\sum f (y - \bar{y})}{\sum f} = 0$$

$$m_2 = \frac{\sum_{i=1}^{n} f_i (y_i - \bar{y})^2}{n} = \frac{\sum f (y - \bar{y})^2}{\sum f}$$

$$m_3 = \frac{\sum_{i=1}^{n} f_i (y_i - \bar{y})^3}{n} = \frac{\sum f (y - \bar{y})^3}{\sum f}$$

$$m_4 = \frac{\sum_{i=1}^{n} f_i (y_i - \bar{y})^4}{n} = \frac{\sum f (y - \bar{y})^4}{\sum f}$$

Example 4.16: Calculate first four moments about the mean for the following set of marks obtained in the examination.

45, 32, 37, 46, 39, 36, 41, 48 and 36.

Solution:

y	$y-\bar{y}$	$(y-\bar{y})^2$	$(y-\bar{y})^3$	$(y-\bar{y})^4$
32	-8	64	512	4096
36	-4	16	64	256
36	-4	16	64	256
37	-3	9	27	81
39	-1	1	1	1
41	1	1	1	1
45	5	25	125	625
46	6	36	216	1296
48	8	64	512	4096
360	0	232	186	10706

$$\bar{y} = \frac{\sum y}{n} = \frac{360}{9} = 40$$

$$m_1 = \frac{\sum (y-\bar{y})^1}{n} = \frac{0}{9} = 0$$

$$m_2 = \frac{\sum (y-\bar{y})^2}{n} = \frac{232}{9} = 25.78$$

$$m_3 = \frac{\sum (y-\bar{y})^3}{n} = \frac{186}{9} = 20.67$$

$$m_4 = \frac{\sum (y-\bar{y})^4}{n} = \frac{10706}{9} = 1189.56$$

Example 4.17: Find first four moments about the mean for the following data.

y	22	27	32	37	42	47	52
f	1	4	8	11	15	9	2

Solution:

y	*f*	*fy*	$y-\bar{y}$	$f(y-\bar{y})$	$f(y-\bar{y})^2$	$f(y-\bar{y})^3$	$f(y-\bar{y})^4$
22	1	22	-17	-17	289	4913	83521
27	4	108	-12	-48	576	6912	82944
32	8	256	-7	-56	392	2744	19208
37	11	407	-2	-22	44	88	176
42	15	630	3	45	135	405	1215
47	9	423	8	72	576	4608	36864
52	2	104	13	26	338	4394	57122
Total	50	1950	–	0	2350	5250	281050

$$\bar{y} = \frac{\Sigma fy}{\Sigma f} = \frac{1950}{50} = 39$$

$$m_1 = \frac{\Sigma f(y-\bar{y})}{\Sigma f} = 0$$

$$m_2 = \frac{\Sigma f(y-\bar{y})^2}{\Sigma f} = \frac{2350}{50} = 47$$

$$m_3 = \frac{\Sigma f(y-\bar{y})^3}{\Sigma f} = \frac{-5250}{50} = -105$$

$$m_4 = \frac{\Sigma f(y-\bar{y})^4}{\Sigma f} = \frac{281050}{50} = 5621$$

4.3.2 Moment about an arbitrary value

The rth sample moment about any arbitrary origin a denoted by m'_r is defined as:

$$m'_r = \frac{\sum_{i=1}^{n}(y_i - a)^r}{n} = \frac{\sum_{i=1}^{n} D_i^r}{n} \qquad (4.29)$$

where, $D_i = (y_i - a)$

so

$$m'_1 = \frac{\sum_{i=1}^{n}(y_i - a)}{n} = \frac{\sum_{i=1}^{n} D_i}{n} \qquad (4.30)$$

$$m'_2 = \frac{\sum_{i=1}^{n}(y_i - a)^2}{n} = \frac{\sum_{1=1}^{n} D_1^2}{n} \qquad (4.31)$$

$$m'_3 = \frac{\sum_{i=1}^{n}(y_i - a)^3}{n} = \frac{\sum_{i=1}^{n} D_i^3}{n} \qquad (4.32)$$

$$m'_4 = \frac{\sum_{i=1}^{n}(y_i - a)^4}{n} = \frac{\sum_{i=1}^{n} D_i^4}{n} \qquad (4.33)$$

The moments about the mean are usually called central moments and the moments about any arbitrary origin a are called non-central moments or raw moments.

The rth sample moment for grouped data about any arbitrary origin a denoted by m'_r is defined as:

$$m'_r = \frac{\sum_{i=1}^{n} f_i (y_i - a)^r}{\sum f} = \frac{\sum_{i=1}^{n} f_i D_i^r}{\sum f} \quad (4.34)$$

$$m'_1 = \frac{\sum_{i=1}^{n} f_i (y_i - a)}{\sum f} = \frac{\sum_{i=1}^{n} f_i D_i}{\sum f} \quad (4.35)$$

$$m'_2 = \frac{\sum_{i=1}^{n} f_i (y_i - a)^2}{\sum f} = \frac{\sum_{i=1}^{n} f_i D_i^2}{\sum f} \quad (4.36)$$

$$m'_3 = \frac{\sum_{i=1}^{n} f_i (y_i - a)^3}{\sum f} = \frac{\sum_{i=1}^{n} f_i D_i^3}{\sum f} \quad (4.37)$$

$$m'_4 = \frac{\sum_{i=1}^{n} f_i (y_i - a)^4}{\sum f} = \frac{\sum_{i=1}^{n} f_i D_i^4}{\sum f} \quad (4.38)$$

$$m_1 = m'_{1_1} - m'_1 = 0 \quad \ldots\ldots\ldots\ldots \quad (4.38.a)$$

$$m_2 = m'_2 - (m'_1)^2 \quad \ldots\ldots\ldots\ldots \quad (4.38.b)$$

$$m_3 = m'_3 - 3m'_2\, m'_1 + 2(m'_1)^3 \ldots\ldots\ldots\ldots (4.38.c)$$

$$m_4 = m'_4 - 4m'_3\, m'_1 + 6m'_2(m'_1)^2 - 3(m'_1)^4 \ldots \quad (4.38.d)$$

Example 4.18: Find first four moments about the mean for ungrouped data of the example 2.1.

Example 4.18: The moments can be calculated directly by using relation (4.27) or relation 4.29 by selecting an arbitrary origin a. We calculate moments about origin a taking $a = 98$ and then calculate moments about mean by using the relations 4.40 to 4.43.

Y_i	$D_i = Y_i = 98$	D_i^2	D_i^3	D_i^4
87	-11	121	-1331	14641
91	-7	49	-343	2401
89	-9	81	-729	6561
88	-10	100	-1000	10000
89	-9	81	-729	6561
91	-7	49	-343	2401
87	-11	121	-1331	14641

92	-6	36	-216	1296
90	-8	64	-512	4096
98	0	0	0	0
95	-3	9	-27	81
97	-1	1	-1	1
96	-2	4	-8	16
100	2	4	8	16
101	3	9	27	81
96	-2	4	-8	16
98	0	0	0	0
99	1	1	1	1
98	0	0	0	0
100	2	4	8	16
102	4	16	64	256
99	1	1	1	1
101	3	9	27	81
105	7	49	343	2401
103	5	25	125	625
107	9	81	729	6561
105	7	49	343	2401
106	8	64	512	4096
107	9	81	729	6561
112	14	196	2744	38416
12929	-11	1309	-917	124225

$$\because m'_r = \frac{\sum D_i^r}{n}$$

$$\Rightarrow \quad m'_1 = \frac{\sum D_i}{n}$$

$$m'_1 = -11/30 = -0.3667$$

$$m'_2 = \sum D_i^2/n$$
$$= 1309/30 = 43.6333$$

$$m'_3 = \sum D_i^3/n$$
$$= -917/30 = -30.5667$$

$$m'_4 = \sum D_i^4/n$$
$$= 124225/30 = 4140.8333$$

The moments about mean are given by

$$m_1 = m'_1 - m'_1$$
$$= 0$$
$$m_2 = m'_2 - (m'_1)^2$$
$$= 43.6333-(-0.3667)^2$$
$$= 43.6333-0.1345$$
$$= 43.4988$$
$$m_3 = m'_3 - 3m'_2\ m'_1 + 2(m'_1)^3$$
$$= -30.5667-3(43.6333)(-0.3667)+2(-0.3667)^3$$
$$= -30.5667+48.0010-0.0989$$
$$= 17.3354$$
$$m_4 = m'_4 - 4m'_3\ m'_1 + 6m'_2\ (m'_1)^2 - 3(m'_1)^4$$
$$= 4140.8333 - 4(-30.5667)(-0.3667)$$
$$+\ 6(43.6333)(-0.3667)^2 - 3(-0.3667)^4$$
$$= 4139.5-44.8352+35.2039 - 0.0542$$
$$= 4131.1478$$

4.3.3 Moments for grouped data

The moments for grouped data about an arbitrary origin with equal class interval h may be written as:

$$m'_r = h^r \frac{\sum_{i=1}^{k} f_i \left(\frac{y_i - a}{h}\right)^r}{n} = h^r \frac{\sum_{i=1}^{k} f_i u_i^r}{n} \qquad (4.39)$$

where $u_i = \dfrac{y_i - a}{h}$

The moments about an arbitrary origin and moments about mean have the following relationships.

$$m_1 = m'_1 - m'_1 = 0 \qquad (4.40)$$
$$m_2 = m'_2 - (m'_1)^2 \qquad (4.41)$$

$$\text{Co-efficient of skewness} = \frac{\text{Mean - Mode}}{\text{Standard deviation}}$$

$$= \frac{67.45 - 67.35}{2.92} = 0.034$$

Example 4.21: The weight of 38 male students at a university are given in the following frequency table:

Weight	118-126	127-135	136-144	145-153	154-162	163-171
f	3	5	9	12	5	4

Calculate Bowley's co-efficient of skewness.

Solution:

Weights	f	C.B.	c.f.
118 – 126	3	117.5 – 126.5	3
127 – 135	5	126.5 – 135.5	8
136 – 144	9	135.5 – 144.5	17
145 – 153	12	144.5 – 153.5	29
154 – 162	5	153.5 – 162.5	34
163 – 171	4	162.5 – 171.5	38
Total:	38	---	-

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - c\right)$$

$$= 135.5 + \frac{9}{12}\left(\frac{38}{4} - 8\right)$$

$$= 136.62$$

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - c\right)$$

$$= 144.5 + \frac{9}{12}\left(3 \times \frac{38}{4} - 17\right)$$

$$= 153.13$$

$$\text{Median} = Q_2 = l + \frac{h}{f}\left(\frac{2n}{4} - c\right)$$

$$= 144.5 + \frac{9}{12}\left(2 \times \frac{38}{4} - 17\right)$$

$$= 146$$

$$\therefore \text{Bowly's co-efficient of skewness} = \frac{Q_3 + Q_1 - 2\,\text{Median}}{Q_3 - Q_1}$$

$$= \frac{153.13 + 136.62 - 2(146)}{153.13 - 137}$$

$$= \frac{-2.25}{16.13}$$

$$= -0.139$$

4.5.1 Kurtosis

The word kurtosis is used to indicate the length of the tails and peakedness of symmetrical distributions. Symmetrical distributions may be platykurtic, mesokurtic (normal) or leptokurtic.

The mesokurtic is the usual normal distribution. Leptokurtic is more peaked and has many values around the mean and in the tails away from the mean whereas platykurtic is bit flat and has more values between the mean and tails as compared to the mesokurtic (normal) distribution. Figure 4.3 shows these shapes of distributions.

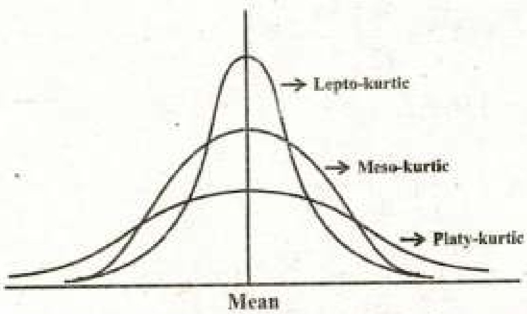

Figure 4.3: Mesokurtic, platykurtic and leptokurtic distributions

One of the numerical measures used to know about the symmetry of a distribution is $\sqrt{\beta_1}$ (β_1 is a Greek letter read as beta one) and is defined as:

$$\sqrt{\beta_1} = \frac{\mu_3}{\sqrt{(\mu_2)^3}} \tag{4.46}$$

It is dimensionless quantity. If $\sqrt{\beta_1} = 0$, the distribution is symmetrical. If $\sqrt{\beta_1} < 0$, the distribution is negatively skewed and for $\sqrt{\beta_1} > 0$, the distribution is positively skewed. The population parameters are estimated from the corresponding sample statistic. The estimate of $\sqrt{\beta_1}$ is denoted by $\sqrt{b_1}$ and is defined as:

$$\sqrt{b_1} = \frac{m_3}{\sqrt{m_2^3}} \tag{4.47}$$

It should be noted that these sample statistics tell us only about the particular data set under consideration and not for the whole population.

Some other measures of skewness are:

a) **Karl Pearson's first co-efficient of skewness**

$$\frac{\text{Mean - Mode}}{S} \tag{4.48}$$

b) **Karl Pearson's second co-efficient of skewness**

$$\frac{3(\text{Mean - Median})}{S} \tag{4.49}$$

These coefficients are pure numbers and these are zero for symmetrical distributions, negative for negatively skewed distributions and positive for positively skewed distributions.

c) **Bowley's coefficient of skewness based on quartiles**

$$S_k = \frac{Q_1 + Q_3 - 2\,\text{median}}{Q_3 - Q_1} \tag{4.50}$$

It is a pure number and lies between −1 and +1. For symmetrical distributions its value is zero.

The skewness of the distribution of a data set can easily be seen by drawing histogram or frequency curve.

Example 4.20: The heights of 100 college students measured to nearest inch are given in the following table:

Height	60-62	63-65	66-68	69-71	72-74
f	5	18	42	27	8

Calculate co-efficient of skewness.

Solution:

Heights	f	y	$D = y - 67$	fD	fD^2	*C.B.*	*c.f.*
60 – 62	5	61	-6	-30	180	59.5 – 62.5	5
63 – 65	18	64	-3	-54	162	62.5 – 65.5	23
66 – 68	42	67	0	0	0	65.5 – 68.5	65
69 – 71	27	70	3	81	243	68.5 – 71.5	92
72 – 74	8	73	6	48	288	71.5 – 74.5	100
Total	**100**			**45**	**873**		

$$\text{Mean} = A + \frac{\sum f D}{\sum f}$$

$$= 67 + \frac{45}{100} = 67.45 \text{ inches}$$

$$\text{Standard deviation} = \sqrt{\frac{\sum f D^2}{\sum f} - \left(\frac{\sum f D}{\sum f}\right)^2}$$

$$= \sqrt{\frac{873}{100} - \left(\frac{45}{100}\right)^2} = 2.92 \text{ inches}$$

$$\text{Mode} = \hat{X} = l + \frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h$$

$$= 65.5 + \frac{(42 - 18)}{(42 - 18) + (42 - 27)} \times 3$$

$$= 67.35 \text{ inches}$$

$$m_1' = \frac{\sum f\, y}{\sum f} \qquad m_2' = \frac{\sum f\, y^2}{\sum f}$$

$$m_3' = \frac{\sum f\, y^3}{\sum f} \qquad m_4' = \frac{\sum f\, y^4}{\sum f}$$

4.4 Sheppad's Correction for Grouping Error

In case of grouped data, we proceed to calculate first four moments by replacing all the members of a class by the mid value of the respective class. The choice of class boundaries and the mid values, affects the values of our approximation to the first four moments. The first and third moments are not affected that much as the second and fourth moments because, in case of second and fourth moments all the deviations become positive and the grouping error accumulates. Sheppard has suggested the following corrections for second and fourth moments in case of grouped data where the frequency curve of the grouped data approaches to the base line gradually and slowly at each end of the distribution.

$$\text{corrected } m_2 = m_2 - \frac{h^2}{12} \tag{4.44}$$

$$\text{corrected } m_4 = m_4 - \frac{h^2}{2} m_2 + \frac{7}{240} h^4 \tag{4.45}$$

For the example 4.19 corrected m_2 and m_4 are:

$$\text{corrected } m_2 = 45.8055 - \frac{5^2}{12} = 43.722$$

$$\text{corrected } m_4 = 4917.3692 - \frac{5^2}{2}(45.8055) + \frac{7}{240}(5^4) = 4363.0296$$

4.5 Skewness

The word skewness means lack of symmetry of a distribution. A symmetrical distribution is one in which mean, median and mode are identical and the portion of frequency polygon to the left of the mean is the mirror image of the portion to the right of the mean. If a distribution is not symmetrical, it is called skewed or asymmetrical.

The figure 4.1 shows the three types of distributions, i.e., the symmetrical distribution in figure 4.1(a), positively skewed distribution in figure 4.1(b) and negatively skewed distribution in figure 4.1(c). A positively skewed distribution is one whose tail extends to the right hand side and a negatively skewed distribution has longer tail towards left hand side. The positions of mean, median and mode are shown in figure 4.1

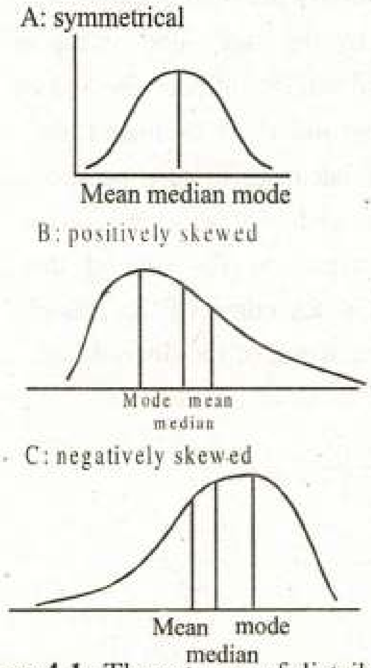

Figure 4.1: Three types of distributions.

The skewness may be very extreme and in such a case these are called J-shaped distributions. This is shown in figure 4.2.

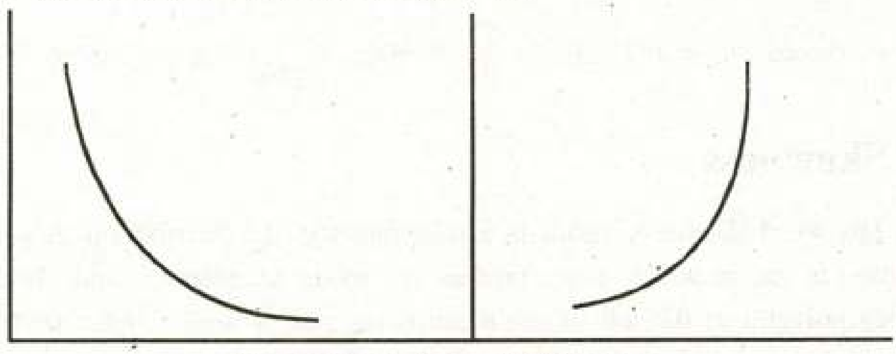

Figure 4.2: J-shaped distributions.

$$m_3 = m'_3 - 3m'_2\ m'_1 + 2(m'_1)^3 \qquad (4.42)$$
$$m_4 = m'_4 - 4m'_3\ m'_1 + 6m'_2\ (m'_1)^2 - 3(m'_1)^4 \qquad (4.43)$$

Moments about mean are calculated by using the above relations in which one needs to calculate moments about mean the arbitrary origin first or these can be directly calculated by using formula moment about mean.

Example 4.19: Find first four moments about an arbitrary origin $a = 98$ for grouped data of the example 3.2

Also find moments about mean by using the moments about origin.

Solution: As width of classes are equal, we use the formula involving h. First the moments about origin $a = 98$ are calculated as follows:

y_i	f_i	$u_i = (y_i\text{-}98)/5$	$f_i u_i$	$f_i u_i^2$	$f_i u_i^3$	$f_i u_i^4$
88	6	-2	-12	24	-48	96
93	4	-1	-4	4	-4	4
98	10	0	0	0	0	0
103	6	1	6	6	6	6
108	3	2	6	12	24	48
113	1	3	3	9	27	81
Total	30	---	-1	55	5	235

We know that:

$$m'_r = \frac{h^r\ \Sigma f_i\ u_i^r}{n}$$

here $h = 5,\ n = \sum f_i = 30$

$$m'_1 = h(\sum f_i u_i)/n$$
$$= 5(-1)/30 = -0.1667$$
$$m'_2 = h^2(\sum f_i u_i^2)/n$$
$$= 5^2\,(55)/30 = 45.8333$$
$$m'_3 = h^3(\sum f_i u_i^3)/n$$
$$= 5^3\,(5)/30 = 20.8333$$
$$m'_4 = h^4(\sum f_i u_i^4)/n$$
$$= 5^4\,(235)/30 = 4895.8333$$

The moments about the mean are

$$m_1 = m'_1 - m'_1 = 0$$

$$m_2 = m'_2 - (m'_1)^2$$

$$= 45.8333 - (-0.1667)^2$$

$$= 45.8055$$

$$m_3 = m'_3 - 3m'_2 \; m'_1 + 2(m'_1)^3$$

$$= 20.8333 - 3(45.8333)(-0.1667) + 2(-0.1667)^3$$

$$= 20.8333 + 22.9167 - 0.0093$$

$$= 43.7407$$

$$m_4 = m'_4 - 4m'_3 \; m'_1 + 6m'_2 \; (m'_1)^2 - 3(m'_1)^4$$

$$= 4895.8333 - 4(-20.8333)(-0.1667) + 6(45.8333)(-0.1667)^2 - 3(-0.1667)^4$$

$$= 4895.8333 + 13.8916 + 7.6419 + 0.0023$$

$$= 4917.3692$$

The first four moments calculated from the same data in ungrouped form and grouped form are slightly different. This is because of the assumption that each observation in a class is equal to mid point of that class while grouping the data.

4.3.4 Moment about zero

If the variable y assumes n values $y_1, y_2, y_3, \ldots, y_n$ then rth moment about zero can be obtained by taking a = 0, so, for relation 4.29

$$m'_r = \frac{\sum y^r}{n}$$

Putting $r = 1,2,3$ and 4 we get

$$m'_1 = \frac{\sum y}{n} \qquad m'_2 = \frac{\sum y^2}{n}$$

$$m'_3 = \frac{\sum y^3}{n} \qquad m'_4 = \frac{\sum y^4}{n}.$$

m'_1, m'_2, m'_3 and m'_4 are the first four moments about zero.

For frequency Distribution, the raw moment about zero are given by

4.29 The daily income of employees range from Rs.0 to Rs.18. They are grouped in intervals of Rs.2 and class frequencies from the lowest to the highest class are 5, 39, 69, 41, 29, 22, 16, 7 and 5. Find the co-efficient of skewness.

4.30 First four moments of distribution about $x = 2$ are 1, 2.5, 5.5 and 16, calculate mean and co-efficient of variation.

4.31 Find moments about mean β_1 and β_2. Given the first 4 moments about $y = 20$ as –2, 15, -25 and 80 respectively.

4.32 What is meant by skewness and kurtosis. What aspects of the frequency curve are measured by them.

4.33 Second moment about mean of two distributions are 9 and 16 while fourth moment about mean are 230 and 780 respectively, which of the distribution is

i) Leptokurtic ii) Platykurtic

4.34 What can you say about skewness in each of the following cases?

i) Median is 26.01 while two Quartiles are 13.73 and 28.29.

ii) Mean = 140 and Mode = 148.7.

iii) First three moments about 16 are 0.35, 2.9 and 1.93 respectively

4.35 i) The second moment about mean of two distributions are 13.76 and 63.0 while the fourth moments about the mean are 528.06 and 9500 respectively. Which of the distributions is

a) Leptokurtic b) Mesokurtic c) Platykurtic.

ii) The fourth central moment of a symmetrical distribution is 243. What would be the value of standard deviation for which distribution is mesokurtic?

4.36 The second moment about mean of a distribution is 25, what would be the value of fourth moment about mean if the distribution is

i) LeptoKurtic ii) MesoKurtic iii) PlayKurtic.

4.37 Which of the following is correct for a negatively skewed distribution;

i) A. M. is greater than mode. ii) A. M. is less than mode.

iii) A. M. is greater than median.

4.38 What would be the shape and the name of the distribution if

i) Mean = Median = Mode ii) Mean > Median > Mode.

iii) Mean < Median < Mode iv) $\beta_1 = 0$ and $\beta_2 = 3$

v) $\beta_1 = 0$ and $\beta_2 = 5$

4.39 Against each statement, write T for true and F for false statement.

i) The sum of squares of the deviations for a data set from the median is minimum.

ii) The sum of absolute deviations for a data set from the mean is minimum.

iii) If each of the observations in a data set is multiplied by a constant, the variance of the resulting data set increases.

iv) If a constant is added in each of the observations in a data set, then the variance of the resulting data set increases.

v) Standard deviation is a positive square root of variance.

vi) Range is a measure of absolute dispersion.

vii) A relative measure is independent of unit.

viii) Mean deviation is not based on all the observations.

ix) Semi-inter quartile Range is also called inter quartile range.

x) The standard deviation is dependent upon origin.

4.15 Calculate mean deviation (about median) for the distribution given below:

Groups	Frequencies
100 – 110	56
110 – 120	59
120 – 130	61
130 – 140	68
140 – 150	77
150 – 160	59
160 – 170	51
170 – 180	42
180 – 190	36
190 – 200	25

4.16 Calculate standard deviation by using arithmetic mean and also by using any provisional mean and compare the results for the data given below:

3, 5, 7, 13, 15, 17, 23, 27

4.17 A manufacturer of television tubes produces two types *A* and *B* of tubes. The tubes have respective mean life times as $\overline{X}_A$ = 1496 hours and $\overline{X}_B$ = 1895 hours and standard deviations S_A = 280 hours and

S_B = 310 hours. Which tube has the greatest:

i) absolute dispersion ii) relative dispersion.

4.18 i) What are moments about mean and about an arbitrary value. Give the relation between them.

ii) Define the moment ratios b_1 and b_2.

4.19 Computer calculated mean and standard deviation from 20 observation as 42 and 5 respectively. It was later discovered at the time of checking that it had copied down two values as 45 and 38 whereas the correct values were 35 and 58 respectively. Find the correct value of co-efficient of variation.

4.20 A distribution consists of 3 components with frequency 100, 120 and 150 having means: 5.5, 15.8 and 10.5 and standard deviations: 2.4, 4.2, and 3.7 respectively. Find the co-efficient of variation for the combined distribution.

4.21 Calculate first four moments about mean for the following set of examination marks.

45, 32, 37, 46, 39, 36, 41, 48 and 36.

4.22 Calculate first four moments for the following distribution of wages about $y = 10$. Find moments about mean

Earning	5	6	7	8	9	10	11	12	13	14	15
f	1	2	5	10	20	51	22	11	5	3	1

4.23 First three moments of a distribution about $y = 4$ are 1, 4 and 10 respectively. Find co-efficient of variation. Is the distribution symmetrical or positively skewed or negatively skewed.

4.24 First four moments of a certain distribution about $y = 17.5$ are 0.3, 74, 45 and 12125 respectively. Find out whether the distribution is lepto kurtic or platyckurtic.

4.25 What can you say of skewness in each of the following cases:

i) $Q_2 = 26.01$, $Q_3 = 38.29$, $Q_1 = 13.73$

ii) Mean =1403 and Mode = 1487

4.26 Given that $\Sigma f = 76$, $\Sigma fy = 572$, $\Sigma fy^2 = 4848$, $\Sigma fy^3 = 44240$ and $\Sigma fy^4 = 42580$. Find first four moments about mean and test the distribution for symmetry and kurtosis.

4.27 If distribution has mean 1403 and mode 1487, what can you say about the skewness.

4.28 Lower and upper quartiles of a distribution are 142.36 and 167.73 respectively while median is 153.50. Find co-efficient of skewness.

4.6 Calculate Mean Deviation from Mean, Mean co-efficient of dispersion and variance from the data given below:

Weights (kg)	No. of students	Weights (kg)	No. of students
50 – 53	23	65 – 68	66
53 – 56	24	68 – 71	49
56 – 59	39	71 – 74	38
59 – 62	46	74 – 77	21
62 – 65	54	77 – 80	12

also calculate Range, Quartile Deviation, and co-efficient of Quartile Deviation.

4.7 Calculate quartile deviation for the data given below:

Groups	25-50	50-75	75-100	100-125	125-150	150-175
Frequency	10	12	16	17	20	18

also calculate co-efficient of Quartile Deviation.

4.8 Calculate standard deviation, variance and co-efficient of variation from the following data:

y	525	500	475	450	425	400	375
f	24	35	46	37	47	34	22

4.9 The mean of a set of 10 values is 25.2 and its standard deviation is 3.72 for another set of 15 values mean and standard deviation are 25.2 and 4.05 respectively. Find the combined standard deviation of the 25 values taken together.

4.10 For a group of 50 boys, the mean score and the standard deviation of scores on a test are 59.5 and 8.38 respectively, for a group of 40 girls, the mean and standard deviation are 54.0 and 8.23 respectively on the same test. Find the mean and standard deviation for the combined group of 90 children.

4.11 By multiplying each number 3, 6, 1, 7, 2, 5 by 2 and then adding 5, we obtained 11, 17, 7, 19, 9, 15. What is the relationship between standard deviation and means for the two sets of numbers?

4.12 The scores obtained by 5 students on a set of examination papers were 70, 50, 60, 70, 50. Their scores are changed by

i) adding 10 point to scores   ii) increasing all scores by 10%.

What effect will these changes have on mean and on standard deviation?

4.13 Compute the mean wages and the co-efficient of variation for the employees working in two factories are given in the following table:

Wages	Factories	
	Factory *A*	**Factory *B***
30 – 35	12	4
35 – 40	18	10
40 – 45	29	31
45 – 50	32	67
50 – 55	16	35
55 – 60	8	15

4.14 Compute median and mean deviation from median for the data given below:

Daily Wages	No. of Domestic Servants
6	5
8	10
10	18
12	20
14	22
16	14
18	7
20	3
22	1

The leptokurtic distribution may be composite of two normal distributions with the same mean but different variances. The platykurtic distribution may be the composite of two normal populations with the same variance but different means.

The dimensionless measure of kurtosis based on the moments is β_2 and is defined as:

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \tag{4.51}$$

If $\beta_2 = 3$, the distribution is mesokurtic (normal). If $\beta_2 < 3$, the distribution is platykurtic and if $\beta_2 > 3$, distribution is leptokurtic. If this measure is calculated by using sample moments then

$$b_2 = \frac{m_4}{m_2^2} \tag{4.52}$$

Here, b_2 is the estimate of β_2

Another measure of kurtosis not widely used is Percentile co-efficient of Kurtosis, and is denoted by K.

$$K = \frac{Q.D}{P_{90} - P_{10}} \tag{4.53}$$

Where, Q.D. stands for quartile deviation and P_{90} and P_{10} are the 90th and 10th percentiles respectively. K is 0.263 for a normal distribution and lies between 0 and 0.50.

Example 4.22: Calculate $\sqrt{b_1}$, b_2 for ungrouped data of the example 4.18 and for grouped data of the example 4.19.

Solution:

For ungrouped data: For ungrouped data of the example 4.18 we have,

$$m_2 = 43.4988, m_3 = 17.3354, m_4 = 4131.1478$$

so,
$$\sqrt{b_1} = \frac{m_3}{\sqrt{m_2^3}}$$

$$= \frac{17.3354}{\sqrt{(43.4988)^3}}$$

$$= 0.0604$$

$$b_2 = \frac{m_4}{m_2^2}$$

$$= \frac{4131.1478}{(43.4988)^2}$$

$$= 2.1833$$

For grouped data: For grouped data of the example 4.19, we have

$m_2 = 43.7220$, $m_3 = 43.7407$, $m_4 = 4363.0296$

$$\text{so, } \sqrt{b_1} = \frac{m_3}{\sqrt{m_2^3}} = \frac{43.7407}{\sqrt{(43.7220)^3}} = 0.1513$$

$$b_2 = \frac{m_4}{m_2^2} = \frac{4363.0296}{(43.7220)^3} = 2.2824$$

Exercise 4

Ans on Page 251

4.1 What do you understand by dispersion? What are the most usual methods of measuring dispersion, indicate the advantages and disadvantages of these methods?

4.2 Define mean deviation and its co-efficient. Discuss its advantages and uses.

4.3 i) What is semi-inter quartile Range.

ii) Define range and discuss its uses.

4.4 Explain the difference between absolute dispersion and relative dispersion. Describe the properties of the standard deviation.

4.5 i) Define various measures of dispersion and given their formulae.

ii) The following table gives the marks of students:

Marks	30-39	40-49	50-59	60-69	70-79
f	8	87	190	86	20

Calculate:

a) Quartile deviation b) Co-efficient of skewness

xi) Co-efficient of variation is a Relative measure of dispersion.

xii) The first moment about mean is one.

xiii) If b_1=3,the distribution will be symmetrical.

xiv) Mean Deviation is always less than the standard deviation.

xv) The normal Distribution is also known as Mesokurtic Distribution.

4.40 Fill in the blanks.

i) A measure of dispersion is ________.

ii) A measure of dispersion expressed as a co-efficient is called ________ measures of dispersion.

iii) Sum of absolute deviations are minimum if computed from ________.

iv) The value of standard deviation does not ________ if a constant is added or subtracted from all observations.

v) Co-efficient of variation is always ________ from unit of ________.

vi) A data having least C.V. is considered more ________.

vii) The lack of symmetry is called ________.

viii) In a symmetrical distribution, the quartile deviation is ________ from the median on both sides.

ix) Shepherd correction is applicable when the frequency distribution tends to ________ in both directions.

x) A relative measure of dispersion is the ________ between absolute dispersion and the average.