CHAPTER

57

STATISTICAL TECHNIQUE

(MOMENT, MOMENT GENERATING FUNCTION, SKEWNESS, KURTOSIS)

57.1 STATISTICS

Statistics is a branch of science dealing with the collection of data, organising, summarising, presenting and analysing data and drawing valid conclusions and thereafter making reasonable decisions on the basis of such analysis.

57.2 FREQUENCY DISTRIBUTION.

Frequency distribution is the arranged data, summarised by distributing it into classes or categories with their frequencies.

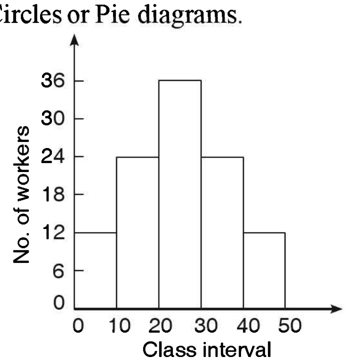
Wages of 100 workers

Wages in Rs.	0-10	10-20	20-30	30-40	40-50
Number of workers	12	23	35	20	10

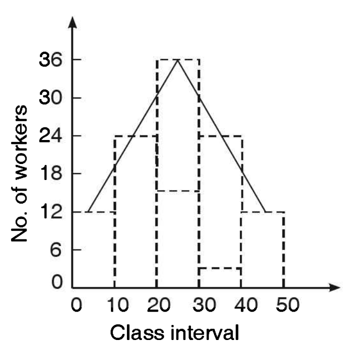
57.3 GRAPHICAL REPRESENTATION

It is often useful to represent frequency distribution by means of a diagram. The different types of diagrams are

1. Histogram
2. Frequency polygon
3. Frequency curve
4. Cumulative frequency curve or Ogive
5. Bar chart
6. Circles or Pie diagrams.



Histogram

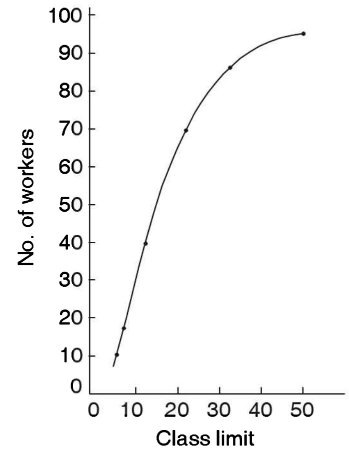


Frequency polygon

1. Histogram. Histogram consists of a set of rectangles having their heights proportional to the class-frequencies, for equal class-intervals. For unequal class-interval, the areas of rectangles are proportional to the frequencies.

2. Frequency Polygon. Frequency Polygon is a line graph of class-frequency plotted against class-mark. It can be obtained by connecting mid-points on the tops of the rectangles in the histogram.

3. Cumulative Frequency curve or the Ogive. If the various points are plotted according to the upper limit of the class as x co-ordinate and the cumulative frequency as y co-ordinate and these points are joined by a free hand smooth curve, the curve obtained is known as cumulative frequency curve or the Ogive.



Ogive

57.4 EXCLUSIVE AND INCLUSIVE CLASS INTERVALS

Exclusive Class intervals: If the upper limit of a class is not included in its class interval, then this type of class interval is called exclusive, *e.g.*,

Income (Rs.)	*No. of workers*
40 – 50	20
50 – 60	30
60 – 70	40
70 – 80	22
80 – 90	18
90 – 100	10

In this method the upper limit of one class is the lower limit of next class. In the above example, in the class 40 – 50 there are twenty persons whose income is from Rs. 40 to 49.99. A person whose income is Rs. 50 is included in the class Rs. 50 – Rs. 60. Not in Rs. 40 – Rs. 50. These type of class-intervals are also known as overlapping class intervals.

Inclusive Class intervals: If the upper limit of a class is included in its class interval, then it is called *inclusive class interval.*

Income (Rs.)	*Frequency*
40 – 49	8
50 – 59	12
60 – 69	15
70 – 79	20
80 – 89	8
90 – 99	7

Adjustment: For the sake of continuity and to get correct class-limits, some adjustment is to be done. Find the difference between the upper limit of the first class and lower limit of the next class and divide it by 2. Then subtract this adjustment from the lower-limit & add it to upper limit to get correct class-limit. Here, in the above example,

$$\text{Adjustment} = \frac{50 - 49}{2} = 0.5$$

Subtract 0.5 from all the lower limit and add 0.5 to all the upper limits.

The adjusted class will be

Income (*Rs.*)	*Frequency*
39.5 – 49.5	8
49.5 – 59.5	12
59.5 – 69.5	15
69.5 – 79.5	20
79.5 – 89.5	8
89.5 – 99.5	7

57.5 AVERAGE OR MEASURES OF CENTRAL TENDENCY

An average is a value which is representative of a set of data. Average value may also be termed as measures of central tendency. There are five types of averages in common.

(*i*) Arithmetic average or mean (*ii*) Median (*iii*) Mode

(*iv*) Geometric Mean (*v*) Harmonic Mean

57.6 ARITHMETIC MEAN

(*a*) If $x_1, x_2, x_3, \ldots\ldots, x_n$ are n numbers, then their arithmetic mean (A.M.) is defined by

$$A.M. = \frac{x_1 + x_2 + x_3 + \ldots. + x_n}{n} = \frac{\Sigma x}{n}$$

If the number x_1 occurs f_1 times x_2 occurs f_2 times and so on, then

$$A.M. = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \ldots. + f_nx_n}{f_1 + f_2 + \ldots.. + f_n} = \frac{\Sigma f\,x}{\Sigma f}$$

This is known as direct method.

Example 1. *Find the mean of 20, 22, 25, 28, 30.*

Solution. $A.M. = \frac{20+22+25+28+30}{5} = \frac{125}{5} = 25$ **Ans.**

Example 2. *Find the mean of the following*

Numbers	*8*	*10*	*15*	*20*
Frequency	*5*	*8*	*8*	*4*

Solution. (*a*) Direct method $\Sigma fx = 8 \times 5 + 10 \times 8 + 15 \times 8 + 20 \times 4$

$= 40 + 80 + 120 + 80 = 320$

$\Sigma f = 5 + 8 + 8 + 4 = 25$

$$A.M. = \frac{\Sigma fx}{\Sigma f} = \frac{320}{25} = 12.8$$ **Ans.**

(*b*) **Short cut method**

Let a be the assumed mean, d the deviation of the variate x from a. Then

$$\frac{\Sigma fd}{\Sigma f} = \frac{\Sigma f(x-a)}{\Sigma f} = \frac{\Sigma fx}{\Sigma f} - \frac{\Sigma fa}{\Sigma f} = A.M. - \frac{a\Sigma f}{\Sigma f} = A.M. - a$$

$$\therefore \quad \boxed{A.M. = a + \frac{\Sigma fd}{\Sigma f}}$$

Example 3. *Find the arithmetic mean for the following distribution:*

Class	*0-10*	*10-20*	*20-30*	*30-40*	*40-50*
Frequency	*7*	*8*	*20*	*10*	*5*

Solution. Let assumed mean $(a) = 25$.

Class	Class-mark x	Frequency f	$x - 25 = d$	$f.d$
0 – 10	5	7	– 20	– 140
10 – 20	15	8	– 10	– 80
20 – 30	25	20	0	0
30 – 40	35	10	+ 10	+ 100
40 – 50	45	5	+ 20	+ 100
Total		$\Sigma f = 50$		$\Sigma fd = -20$

$$A.M. = a + \frac{\Sigma fd}{\Sigma f} = 25 + \frac{-20}{50} = 24.6$$ **Ans.**

(*c*) **Step deviation method**

Let a be the assumed mean, i the class length then

$$D = \frac{x-a}{i}, \quad A.M. = a + \frac{\Sigma f D}{\Sigma f} i$$

Example 4. *Find the arithmetic mean of the data given in example 3 by step deviation method.*

Solution. Let assumed mean $(a) = 25$, $i = 10$ (given)

Class	class-mark x	Frequency f	$D = \frac{x-a}{i}$	$f.D$
0 – 10	5	7	– 2	– 14
10 – 20	15	8	– 1	– 8
20 – 30	25	20	0	0
30 – 40	35	10	+ 1	+ 10
40 – 50	45	5	+ 2	+ 10
Total		$\Sigma f = 50$		$\Sigma fD = -2$

$$A.M. = a + \frac{\Sigma f D}{\Sigma f} . i = 25 + \frac{-2}{50} \times 10 = 24.6$$ **Ans.**

57.7 MEDIAN

Median is defined as the measure of the central item when they are arranged in ascending or descending order of magnitude.

(*a*) When the total number of the items is odd and equal to say n, then the value of $\frac{1}{2}(n+1)$th item gives the median.

(*b*) When the total number of the frequencies is even, say n, then there are two middle items, and so the mean of the values of $\frac{1}{2}n$th and $\left(\frac{1}{2}n+1\right)$th items is the median.

Example 5. *Find the median of 6, 8, 9, 10, 11, 12, 13.*

Solution. Total number of items = $n = 7$

Since, n is an odd number. Hence,

$$\text{The middle item } = \frac{1}{2}(7+1)^{th} = 4th$$

$$\text{Median} = \text{Value of the 4th item} = 10 \qquad \textbf{Ans.}$$

(*c*) For grouped data, $$\boxed{\textbf{Median} = l + \frac{\frac{1}{2}N - C}{f} \cdot i}$$

where l is the lower limit of the median class, f is the frequency of the class, i is the class-length, C is the cumulative frequency of the class preceding the median-class and N is the cumulative frequency of the data.

Example 6. *Find the value of Median from the following data:*

No. of days for which absent (less than)	*5*	*10*	*15*	*20*	*25*	*30*	*35*	*40*	*45*
No. of students	*29*	*224*	*465*	*582*	*634*	*644*	*650*	*653*	*655*

Solution. The given cumulative frequency distribution will first be converted into ordinary frequency as under

Class-Interval	*Cumulative frequency*	*Ordinary frequency*
0 – 5	29	29 = 29
5 – 10	$\boxed{224 = C}$	224 – 29 = 195
$\boxed{10 - 15}$	465	465 – 224 = $\boxed{241}$ f
15 – 20	582	582 – 465 = 117
20 – 25	634	634 – 582 = 52
25 – 30	644	644 – 634 = 10
30 – 35	650	650 – 644 = 6
35 – 40	653	653 – 650 = 3
40 – 45	$\boxed{655} = N$	655 – 653 = 2

Here, $$\frac{N}{2} = \frac{655}{2} = 327.5$$

Hence, Median class = class having *c.f.* just more than $\frac{N}{2}$ *i.e.* 327.5 = 10 – 15

Now, $$\text{Median } = l + \frac{\frac{N}{2} - C}{f} i$$

where l stands for lower limit of median class,

N stands for the total frequency,

C stands for the cumulative frequency of the class just preceding the median class,

i stands for width of class interval

f stands for frequency of the median class.

$$\text{Median} = 10 + \frac{\frac{655}{2} - 224}{241} \times 5 = 10 + \frac{103.5 \times 5}{241} = 10 + 2.15 = 12.15 \qquad \textbf{Ans.}$$

57.8 QUARTILES

Quartiles are the values of the variate which divide the total frequency into four equal parts. When the lower half before the median is divided into two equal parts, the value of the dividing variate is called **lower Quartile** and is denoted by Q_1. The value of the variate dividing the upper half into two equal parts is called the **upper Quartile** and is denoted by Q_3. Q_2 is the medium.

The formulae for computation of Q_1 and Q_3 are,

$$\boxed{Q_1 = l + \frac{i}{f}\left(\frac{N}{4} - C\right)} \Rightarrow \boxed{Q_3 = l + \frac{i}{f}\left(\frac{3N}{4} - C\right)}.$$

57.9 DECILES

Deciles are those values of the variate which divide the total frequency into 10 equal parts.

$$\boxed{D_1 = l + \frac{i}{f}\left(\frac{N}{10} - C\right)} \Rightarrow \boxed{D_2 = l + \frac{i}{f}\left(\frac{2N}{10} - C\right)}$$

$$\boxed{D_3 = l + \frac{i}{f}\left(\frac{3N}{10} - C\right)} \text{ and so on}$$

D_5, the fifth decile is the median.

57.10 PERCENTILES

Percentiles are those values of the variate which divide the total frequency into 100 equal parts.

$$\boxed{P_1 = l + \frac{i}{f}\left(\frac{N}{100} - C\right)} \Rightarrow \boxed{P_2 = l + \frac{i}{f}\left(\frac{2N}{100} - C\right)} \text{ and so on.}$$

P_{50}, the 50$^{\text{th}}$ percentile is the median.

Note. In case of series where frequency is not given,

P_{10} = value of $\frac{10}{100}(n+1)^{th}$ observation, P_{50} = value of $\frac{50}{100}$ $(n + 1)^{\text{th}}$ observations etc.

57.11 MODE

Mode is defined to be the size of the variable which occurs most frequently.

In case of continuous frequency distribution.

$$\boxed{\textbf{Mode} = l + \left(\frac{f - f_{-1}}{2f - f_{-1} - f_{+1}}\right) i}$$

where l is lower limit, i is the class length, f is the frequeny of the modal class, f_{-1} and f_1 are the frequencies of the classes preceding and succeeding the modal class respectively.

The following points must be taken care of while calculating mode :

1. The values (or classes of values) of the variable must be in ascending order of magnitude.

2. If the classes are in inclusive form, then the actual limits of the modal class are to be taken for finding l and i.

3. The classes must be of equal width.

Example 7. *Find the mode of the following items : 0, 1, 6, 7, 2, 3, 7, 6, 6, 2, 6, 0, 5, 6, 0.*

Solution. 6 occurs 5 times and no other item occurs 5 or more than 5 times, hence the mode is 6. **Ans.**

Emperical formula

$$\boxed{\textbf{Mean} - \textbf{Mode} = 3\ [\textbf{Mean} - \textbf{Median}]}$$

Example 8. *Find the mode from the following data:*

Age	0 – 6	6 – 12	12 – 18	18 – 24	24 – 30	30 – 36	36 – 42
Frequency	6	11	25	35	18	12	6

Solution.

Age	*Frequency*	*Cumulative frequency*
0 – 6	6	6
6 – 12	11	17
12 – 18	$25 = f_{-1}$	42
$\boxed{18 - 24}$	$35 = f$	77
24 – 30	$18 = f_1$	95
30 – 36	12	107
36 – 42	6	113

Here, max. frequency of any item is 35.

Hence modal class is 18.24

$$\text{Mode} = l + \frac{f - f_{-1}}{2f - f_{-1} - f_1} \times i = 18 + \frac{35 - 25}{70 - 25 - 18} \times 6$$

$$= 18 + \frac{60}{27} = 18 + 2.22 = 20.22$$ **Ans.**

57.12 GEOMETRIC MEAN

If $x_1, x_2, x_3, \ldots\ldots, x_n$ be n values of variates x, then the geometric mean

$$G = (x_1 \times x_2 \times x_3 \times x_4 \times \ldots \times x_n)^{1/n}$$

Example 9. *Find the geometric mean of 4, 8, 16.*

Solution. $G.M. = (4 \times 8 \times 16)^{1/3} = 8$ **Ans.**

57.13 HARMONIC MEAN

Harmonic mean of a series of values is defined as the reciprocal of the arithmetic mean of their reciprocals. Thus if H be the harmonic mean, then

$$\frac{1}{H} = \frac{1}{n}\left[\frac{1}{x_1} + \frac{1}{x_2} + \ldots. + \frac{1}{x_n}\right]$$

Example 10. *Calculate the harmonic mean of 4, 8, 16.*

Solution. Let, $\frac{1}{H} = \frac{1}{3}\left[\frac{1}{4} + \frac{1}{8} + \frac{1}{16}\right] = \frac{7}{48}$, $H = \frac{48}{7} = 6.857$ **Ans.**

57.14 AVERAGE DEVIATION OR MEAN DEVIATION

It is the mean of the absolute values of the deviations of a given set of numbers from their arithmetic mean.

If $x_1, x_2, x_3, \ldots., x_n$ be a set of numbers with frequencies $f_1, f_2, \ldots. f_n$ respectively. Let $\bar{x}$ be the arithmetic mean of the numbers $x_1, x_2, \ldots., x_n$, then

$$\text{Mean deviation} = \frac{\Sigma f_i \,|\, x_i - \bar{x} \,|}{\Sigma f_i}$$

Example 11. *Find the mean deviation of the following frequency distribution.*

Class	0 – 6	6 – 12	12 – 18	18 – 24	24 – 30
Frequency	8	10	12	9	5

Solution. $a = 15$

Class	Mid-value x	Frequency f	$d = x - a$	fd	$\|x - 14\|$	$f\|x - 14\|$
0 – 6	3	8	– 12	– 96	11	88
6 – 12	9	10	– 6	– 60	5	50
12 – 18	15	12	0	0	1	12
18 – 24	21	9	+ 6	54	7	63
24 – 30	27	5	+ 12	60	13	65
		$\Sigma f = 44$		$\Sigma fd = -42$		$\Sigma f\|x - 14\| = 278$

$$\text{Mean} = \bar{x} = a + \frac{\Sigma f d}{\Sigma f} = 15 - \frac{42}{44} = 14 \text{ (approx.)}$$

$$\text{Mean or Average deviation} = \frac{\Sigma f|x - \bar{x}|}{\Sigma f} = \frac{278}{44} = 6.32 \qquad \textbf{Ans.}$$

57.15 STANDARD DEVIATION

Standard deviation is defined as the square root of the mean of the square of the deviation from the arithmetic mean.

$$S.D. = \sigma = \sqrt{\frac{\Sigma f(x - \bar{x})^2}{\Sigma f}}$$

Note. 1. The square of the standard deviation *i.e.*; σ^2 is called variance.

2. σ^2 is called the second moment about the mean and is denoted by μ_2.

57.16 SHORTEST METHOD FOR CALCULATING STANDARD DEVIATION

We know that $\sigma^2 = \frac{1}{N}\Sigma f(x - \bar{x})^2 = \frac{1}{N}\Sigma f(x - a - \overline{\bar{x} - a})^2$

$$= \frac{1}{N}\Sigma f(d - \overline{\bar{x} - a})^2 \quad \text{where } x - a = d \quad = \frac{1}{N}\Sigma fd^2 - 2(\bar{x} - a)\frac{1}{N}\Sigma fd + (\bar{x} - a)^2\frac{1}{N}\Sigma f$$

$$= \frac{1}{N}\Sigma fd^2 - 2(\bar{x} - a)\frac{1}{N}\Sigma fd + (\bar{x} - a)^2 \qquad [\because \Sigma f = N] \qquad \left[\bar{x} = a + \frac{\Sigma fd}{N} \text{ or } \bar{x} - a = \frac{\Sigma fd}{N}\right]$$

$$\sigma^2 = \frac{1}{N}\Sigma f d^2 - 2\left(\frac{\Sigma f d}{N}\right)\left(\frac{\Sigma f d}{N}\right) + \left(\frac{\Sigma f d}{N}\right)^2 = \frac{1}{N}\Sigma f d^2 - 2\left(\frac{\Sigma f d}{N}\right)^2 + \left(\frac{\Sigma f d}{N}\right)^2$$

$$\sigma^2 = \frac{1}{N}\Sigma fd^2 - \left(\frac{\Sigma fd}{N}\right)^2, \qquad S.D. = \sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

Note. Coefficient of variation $= \frac{\sigma}{x} \times 100$

Example 12. *Calculate the mean and standard deviation for the following table, given the age distribution of 542 members.*

Age in years	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90
No. of members	3	61	132	153	140	51	2

Solution. Assumed mean = 55

Here, we take $d = \frac{x - a}{i} = \frac{x - 55}{10}$

Age grouped	*Mid value* (x)	*Frequency* (f)	$d = \frac{x-55}{10}$	fd	fd^2
20 – 30	25	3	– 3	– 9	27
30 – 40	35	61	– 2	– 122	244
40 – 50	45	132	– 1	– 132	132
50 – 60	55	153	0	0	0
60 – 70	65	140	1	140	140
70 – 80	75	51	2	102	204
80 – 90	85	2	3	6	18
		$\Sigma f = 542$		$\Sigma fd = -15$	$\Sigma fd^2 = 765$

$$\text{Mean} = \bar{x} = a + \frac{\Sigma fd}{\Sigma f}.i = 55 + \frac{(-15)\,10}{542} = 55 - 0.28 = 54.72$$

$$\text{Variance} = \sigma^2 = i^2\left[\frac{1}{N}\Sigma fd^2 - \left(\frac{\Sigma fd}{N}\right)^2\right] = 100\left[\frac{765}{542} - (0.028)^2\right] = 100 \times 1.4107 = 141.07$$

S.D. $= \sigma = 11.9$ years **Ans.**

57.17 SYMMETRY

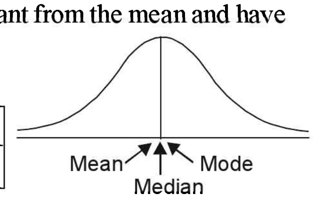
A distribution is said to be symmetrical when its mean, median and mode are identical. *i.e.*;

Mean = Median = Mode.

In other words, a distribution is said to be symmetric when the frequencies are symmetrically distributed about the mean (or when the values of the variable are equidistant from the mean and have the same frequency).

Consider the following frequency distribution:

x	10	20	30	40	50	60	70
f	2	6	10	14	10	6	2



$$\text{Mean} = \bar{x} = \frac{10\times2+20\times6+30\times10+40\times14+50\times10+60\times6+70\times2}{2+6+10+14+10+6+2} = \frac{2000}{50} = 40$$

In this distribution, we observe that the values 20 and 60 are equidistant from the mean, viz. 40 with the same frequency 6.

A symmetrical distribution when plotted on a graph will give a perfectly bell-shaped curve, which is known as normal curve.

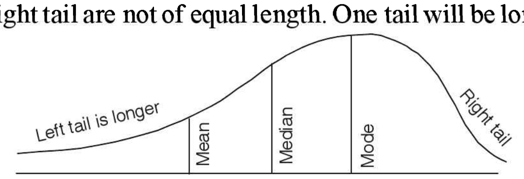
57.18 SKEWNESS

Skewness denotes the opposite of symmetry. It is lack of symmetry,

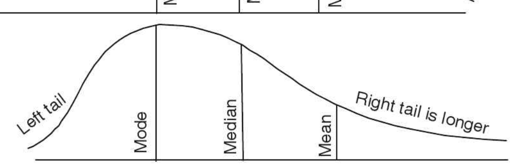
Skew symmetrical Distribution (*U.P. III Semester, 2006*)

A distribution which is not symmetrical is said to be skew symmetrical distribution. In skew symmetrical distribution the left tail and the right tail are not of equal length. One tail will be longer than the other.

(*a*) **Negatively skew distribution.** In negatively skew distribution the left tail is longer than the right tail.



(*b*) **Positively skew distribution.** In positively skew distribution the right tail of the curve will be longer than the left.



In skew distribution mean, median and mode are not equal.

57.19 TEST OF SKEWNESS

1. There is no skewness in the distribution if AM = Mode = Median

2. There is no skewness in the distribution if, Third quartile - Median = Median – First quartile.

3. There is no skewness if

The sum of the frequencies which are less than Mode = Sum of the frequencies which are greater than Mode

4. There is no skewness if quartiles are equidistant from the median.

5. The distribution is negatively skewed if A.M. is less than Mode.

6. The curve is not symmetrical about the median if AM $\neq$ Median $\neq$ Mode.

57.20 USES OF SKEWNESS

1. It gives the nature of the curve.

2. It gives nature and concentration of observations about the mean.

57.21 TYPES OF DISTRIBUTION

1. Fairly symmetrical **2.** Positively skewed **3.** Negatively skewed.

57.22 MEASURE OF SKEWNESS

Measure of skewness is known as the measure of symmetry.

There are two types of measure of skewness.

1. Absolute measure: Absolute measure = Mean – Mode

2. Relative measure : These are four types of relative measure of skewness.

(*i*) Karl Pearson's Coefficient of Skewness

(*ii*) Bowley's Coefficient of Skewness.

(*iii*) Kelly's Coefficient of Skewness and

(*iv*) Measure of skewness based on the moments.

(Mode = 3 Median – 2 Mode)

57.23 KARL PEARSON'S COEFFICIENT OF SKEWNESS:

$$\text{Karl Pearson's Coefficient of Skewness} = \boxed{\frac{\text{Mean} - \text{Mode}}{\text{Standard deviation}}}$$

$$= \frac{\text{Mean} - (3\text{ Median} - 2\text{ Mean})}{\text{Standard deviation}} = \boxed{\frac{3\,(\text{Mean} - \text{Median})}{\text{Standard deviation}}}$$

It generally lies between – 1 and 1.

If its value is zero then there is no skewness.

57.24 TYPES OF SKEWNESS IN TERMS OF MEAN AND MODE

1. There is no skewness in the distribution

$$S_k = 0$$

$$\Rightarrow \frac{\text{Mean} - \text{Mode}}{\text{S.D.}} = 0 \quad \Rightarrow \quad \text{Mean} - \text{Mode} = 0 \quad \Rightarrow \quad \text{Mean} = \text{Mode}.$$

2. The distribution is positively skewed if $S_k > 0$.

$$\frac{\text{Mean} - \text{Mode}}{\text{S.D.}} > 0 \quad \Rightarrow \quad \text{Mean} - \text{Mode} > 0 \quad \Rightarrow \quad \text{Mean} > \text{Mode}.$$

3. The distribution is negatively skewed if $S_k < 0$

$$\Rightarrow \quad \frac{\text{Mean} - \text{Mode}}{\text{S.D.}} < 0 \quad \Rightarrow \quad \text{Mean} - \text{Mode} < 0 \quad \Rightarrow \quad \text{Mean} < \text{Mode}$$

Example 13. *Compute the coefficient of Skewness from the following data:*

x	6	7	8	9	10	11	12
f	3	6	9	13	8	5	4

(*U.P. III Semester, 2009-2010*)

Solution. Let $a = 9$

x	f	$d = x - 9$	fd	fd^2	c.f.
6	3	– 3	– 9	27	3
7	6	– 2	– 12	24	9
8	9	– 1	– 9	9	18
9	13	0	0	0	31
10	8	1	8	8	39
11	5	2	10	20	44
12	4	3	12	36	48
	$\Sigma f = 48$		$\Sigma fd = 0$	$\Sigma fd^2 = 124$	

$$\text{Mean} = a + \frac{\Sigma fd}{\Sigma f} = 9 + \frac{0}{48} = 9$$

Mode = Item of maximum frequency (13) = 9

$$\text{S.D.} = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2} = \sqrt{\frac{124}{48} - \left(\frac{0}{48}\right)^2} = \sqrt{\frac{124}{48}} = 1.61$$

$$\text{Karl Pearson's Coefficient of Skewness} = \frac{\text{Mean} - \text{Mode}}{\text{S.D.}} = \frac{9-9}{1.61} = \frac{0}{1.61} = 0$$ **Ans.**

Example 14. *Calculate Karl Pearson's Coefficient of Skewness from the given data :*

Life time in months	*30-40*	*40-50*	*50-60*	*60-70*	*70-80*	*80-90*	*90-100*	*100-110*	*110-120*
No. of mobile	*4*	*6*	*8*	*26*	*28*	*12*	*8*	*5*	*3*

Solution. Let $a = 75$

Life time in months	No. of mobile (f)	Mid-value (x)	$d = x - 75$	fd	fd^2	*Cumulative frequency*
30 – 40	4	35	– 40	– 160	6400	4
40 – 50	6	45	– 30	– 180	5400	10
50 – 60	8	55	– 20	– 160	3200	18
60 – 70	26	65	– 10	– 260	2600	$\boxed{44}$
$\boxed{70 - 80}$	$\boxed{28}$	75	0	0	0	72
80 – 90	12	85	10	120	1200	84
90 – 100	8	95	20	160	3200	92
100 – 110	5	105	30	150	4500	97
110 – 120	3	115	40	120	4800	100
	$\Sigma f = 100$			$\Sigma fd = -210$	$\Sigma fd^2 = 31300$	

$l = 70$
$N = 100$
$cf = 40$
$\Sigma f = 100$
$i = 10$

Mean $= a + \dfrac{\Sigma fd}{\Sigma f} = 75 + \dfrac{-210}{100} = 72.9$

Median $= l + \dfrac{\frac{N}{2} - c.f.}{f}\, i = 70 + \dfrac{\frac{100}{2} - 44}{28}\,(10) = 70 + 2.143 = 72.143$

S.D. $= \sqrt{\dfrac{\Sigma fd^2}{\Sigma f} - \left(\dfrac{\Sigma fd}{\Sigma f}\right)^2} = \sqrt{\dfrac{31300}{100} - \left(\dfrac{-210}{100}\right)^2} = \sqrt{313 - 4.41}$

$= \sqrt{308.59} = 17.57$

Karl Pearson's Coefficient of Skewness $= \dfrac{3\,(\text{Mean} - \text{Median})}{\text{S.D.}} = \dfrac{3\,(72.9 - 72.143)}{17.57}$

$= \dfrac{3\,(0.757)}{17.57} = \dfrac{2.271}{17.57} = 0.1293$ **Ans.**

EXERCISE 57.1

Calculate Karl Pearson's Coefficient of Skewness from the data given below:

1. S.D. = 6.5, AM = 29.6, mode = 27.52 **Ans.** $S_k = 0.32$

2. Mean = 100, Variance = 35, Median = 99.61. **Ans.** $S_k = 0.2$

3. AM = 45, Median = 48, S.D. = 22.5 **Ans.** $S_k = -0.4$

4. The sum of the 20 observation is 300 and sum of the squares of the observation is 5000, Median = 15. **Ans.** $S_k = 0$

5. Find the Karl Pearson's Coefficient of Skewness for the following

Years under	10	20	30	40	50	60
No. of persons	15	32	51	78	97	109

Ans. – 0.32

6. Calculate Karl Pearson's Coefficient of Skewness from the following data :

Cost per item (in Rs.)	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5
No. of items	35	40	48	100	125	87	43	22

Ans. – 0.2445

7. From the following data calculate Karl Pearson's Coefficient of Skewness.

Scores	0	10	20	30	40	50	60	70	80
No. of players	150	140	100	80	80	70	30	14	0

Ans. – 0.462

8. The weekly wages in Rs. of the workers in a shoe factory are given below :

Weekly wages (in Rs.)	500 - 600	600 - 700	700 - 800	800 - 900	900 - 1000	1000-1100
No. of workers	8	12	4	2	1	1

Calculate Karl Pearson's Coefficient of Skewness. **Ans.** 0.34

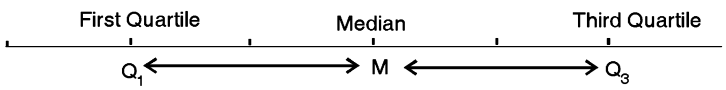
9. Which of the following two series is symmetrical :
Series (a) : Mean = 32, Median = 34, S.D. = 20
Series (b) : Mean = 32, Median = 36, S.D. = 25 **Ans.** series (*a*) is more symmetrical than series (*b*)

10. Karl Pearson's Coefficient of Skewness of a distribution = 0.32, Standard deviation = 6.5
A.M. = 29.6.
From the above data find the mode and the median for the distribution.
Ans. Mode = 27.52, Median = 28.91

57.25 BOWLEY'S COEFFICIENT OF SKEWNESS

Bowley's Coefficient of Skewness is based on the quartiles and Median. A distribution is symmetrical if the distance between the first quartile and median is equal to the distance between the median and third quartile i.e.

$$\text{Median} - Q_1 = Q_3 - \text{Median}$$



It is not so *i.e.*

$$\text{Median} - Q_1 \neq Q_3 - \text{Median}$$

then the distribution is skewed.

57.26 MEASURE OF BOWLEY'S COEFFICIENT OF SKEWNESS

There are two ways to measure Bowley's Coefficient of Skewness.

1. Bowley's absolute measure of Skewness = $Q_3 + Q_1 - 2$ Median
2. Bowley's Relative Measure of Skewness

   Bowley's Coefficient of Skewness $= \dfrac{Q_3 + Q_1 - 2\text{ Median}}{Q_3 - Q_1}$

   This formula is also known as Quartile Coefficient of Skewness.

57.27 CHARACTERISTICS OF BOWLEY'S COEFFICIENT OF SKEWNESS

1. If the distribution is open of unequal class interval then Pearson's Coefficient of Skewness cannot be calculated and Bowley's Coefficient of Skewness can be calculated.
2. Bowley's Coefficient of Skewness lies between – 1 and + 1.
3. Bowley's measure is calculated only from the continuous distribution with exclusive classes.

57.28 LIMITATIONS OF BOWLEY'S COEFFICIENT OF SKEWNESS

1. It is based on the central 50% of the data and ignores the remaining 50% of the data on the extremes.
2. Bowley's formulae and Pearson's formulae cannot be compared. However if the distribution is symmetrical thus both coefficients are zero.

Example 15. *From the following data find Bowley's Coefficient of Skewness:*
Difference of quartiles = 80 Mode = 60
Sum of the quartiles = 120 and Mean = 45

Solution. Here, we have

$Q_3 + Q_1 = 120$
$Q_3 - Q_1 = 80$
Mode = 60
Mean = 45

We know that

Mode = 3 Median – 2 Mean

$\Rightarrow$ 60 = 3 Median – 2 (45) $\Rightarrow$ Median = 50

Bowley's Coefficient of Skewness $= \dfrac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} = \dfrac{120 - 2(50)}{80} = \dfrac{1}{4} = 0.25$ **Ans.**

Example 16. *The sum of the upper quartile and lower quartile is 100 and the median is 55. The Bowley's Coefficient of Skewness is –0.6. Find the upper and lower quartiles.*

Solution. We know that

$$\text{Bowley's Coefficient of Skewness} = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} \quad \ldots (1)$$

Here, $Q_1 + Q_3 = 100$, $M = 55$, $S_k = -0.6$,

Putting the values of $Q_1 + Q_3$, M and Coefficient of Skewness in (1), we get

$$-0.6 = \frac{100 - 2\,(55)}{Q_3 - Q_1} \Rightarrow -0.6 = \frac{100 - 110}{Q_3 - Q_1}$$

$$\Rightarrow \quad -0.6 = \frac{-10}{Q_3 - Q_1} \Rightarrow Q_3 - Q_1 = \frac{50}{3}$$

But $\quad Q_3 + Q_1 = 100 \quad \ldots (2) \quad$ And $\quad Q_3 - Q_1 = \dfrac{50}{3} \quad \ldots (3)$

On adding, we get $2\,Q_3 = \dfrac{500}{3} \Rightarrow Q_3 = \dfrac{250}{3}$

Putting the value of Q_3 in (2), we get

$$\frac{250}{3} + Q_1 = 100 \Rightarrow Q_1 = 100 - = \frac{250}{3} = \frac{50}{3}$$

Hence, $\quad Q_1 = \dfrac{50}{3}$ and $Q_3 = \dfrac{250}{3}$ **Ans.**

Example 17. *Calculate Bowley's Coefficient of Skewness from the data given below.*

No. of houses	0	1	2	3	4	5	6
No. of Airconditioners	15	20	14	25	13	8	4

Solution.

No. of Houses (x)	*No. of Air conditioners* (f)	*Cumulative frequency*
0	15	15
$\boxed{1}$	20	35
2	14	49
3	25	74
$\boxed{4}$	13	87
5	8	95
6	4	99

Q_1 = size of $\dfrac{N+1}{4}$ = 25th item, Hence $Q_1 = 1$

Q_3 = size of $\dfrac{3\,(N+1)}{4}$ = 75th item. Hence $Q_3 = 4$

Median = size of $\dfrac{N+1}{2}$ = 50th item. Hence Mean = 3

$$\text{Bowley's Coefficient of Skewness} = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} = \frac{4 + 1 - 2\,(3)}{4 - 1} = -\frac{1}{3} = -0.33$$ **Ans.**

Example 18. *The following table shows the distances between the worker's residence and their office situated at Connaught Place, New Delhi.*

Distances	*0 – 10*	*10 – 20*	*20 – 30*	*30 – 40*	*40 – 50*	*50 – 60*	*60 – 70*
No. of workers	*2*	*5*	*10*	*15*	*10*	*4*	*1*

Calculate the Bowley's Coefficient of Skewness.

Solution.

Distance (x)	*No. of workers* (f)	*c.f*
0 – 10	2	2
10 – 20	5	7
20 – 30	10	$\boxed{17}$
$\boxed{30-40}$	$\boxed{15}$	$\boxed{32}$
40 – 50	10	42
50 – 60	4	46
60 – 70	1	47

$$N = 47, \quad \frac{N+1}{4} = 12, \quad \frac{N+1}{2} = 24, \quad \frac{3(N+1)}{4} = 36$$

The class 30 – 40 is the median class

$$l = 30, \quad i = 10, \; f = 15, \; cf = 17, \; N = 47$$

$$\text{Median} = l + \frac{\frac{N+1}{2} - c.f.}{f}(i) = 30 + \frac{\frac{47+1}{2} - 17}{15}(10) = 30 + \frac{14}{3} = \frac{104}{3}$$

$$Q_1 = \mathbf{l} + \frac{\frac{N+1}{4} - c.f.}{f}\, i = 20 + \frac{\frac{47+1}{4} - 7}{10}(10) = 20 + 5 = 25$$

$$Q_3 = l + \frac{3\left(\frac{N+1}{4}\right) - c.f.}{f}\, i = 40 + \frac{36-32}{10}(10) = 40 + 4 = 44$$

$$\text{Bowley's Coefficient of Skewness} = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} = \frac{44 + 25 - 2\left(\frac{104}{3}\right)}{44 - 25}$$

$$= \frac{69 - \frac{208}{3}}{19} = \frac{-\frac{1}{3}}{19} = -\frac{1}{57} = -0.0175 \qquad \textbf{Ans.}$$

EXERCISE 57.2

1. The data for a distribution is given below :

$Q_1 = 8.6$, Median = 12.3, $Q_3 = 14.04$

Calculate Bowley's Coefficient of Skewness. **Ans.** – 0.36

2. Calculate (*a*) Karl Pearson's Coefficient of Skewness

(*b*) Bowley's Coefficient of Skewness from the following data :

	City A	City B
A. M.	150	140
Median	142	155
S.D.	30	55
Q_3	195	260
Q_1	62	80

Ans.

	City A	City B
Karl Pearson's Coeff.	0.8	– 0.82
Bowley's Coeff.	– 0.203	0.167

3. Compute the quartile Coefficient of Skewness for the following distribution :

x	3 – 7	8 – 12	13 – 17	18 – 22	23 – 27	28 – 32	33 – 37	38 – 42
f	2	108	580	175	80	32	18	5

Ans. 0.119

4. Calculate Bowley's Coefficient of Skewness for the following data :

x	5	10	15	20	25	30	35	40	45
f	9	10	12	15	11	7	6	5	2

Ans. 0.33

5. A blood donation camp was held at Janakpuri. The distribution of the blood donar is given below :

Age in years	20 – 25	25 – 30	30 – 35	35 – 40	40 – 45	45 – 50	50 – 55	55 – 60
No. of donars	50	70	80	180	150	120	70	50

Calculate the Bowley's Coefficient of Skewness. **Ans.** 0.232

6. Calculate Bowley's Coefficient of Skewness for the following distribution.

Classes	1 – 5	6 – 10	11 – 15	16 – 20	21 – 25	26 – 30	31 – 35
Frequency	3	4	68	30	10	6	2

Hint: First change to exclusive distribution for taking real class limit as first class 0.5 – 5.5

Ans. 0.262

7. The participant of different ages took part in "Marathan" race at INDIA GATE on 2nd October 2008 as follows :

Age (in years)	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
No. of participants	358	2417	976	129	62	18	10

Calculate Bowley's Coefficient of Skewness. **Ans.** 0.131

57.29 KELLY'S COEFFICIENT OF SKEWNESS

Bowley's measure neglects the extreme data to measure skewness. The entire data should be taken into account in measuring skewness.

Kelly modified Bowley's measure of skewness by taking any two deciles equidistant from the median or any two percentiles equidistant from the median.

$$\text{Kelly's Coefficient of Skewness} = \frac{P_{10} + P_{90} - 2P_{50}}{P_{90} - P_{10}}$$

$$\text{Kelly's Coefficient of Skewness} = \frac{D_1 + D_9 - 2\,Median}{D_9 - D_1}$$

P denotes percentile and D denotes decile.

We know that, Median = $P_{50} = D_5$

Note : This method is only theoretical generally Karl Pearson's method is widely used.

Example 19. *Calculate percentile Coefficient of Skewness from the following :*

$P_{90} = 110,\quad P_{10} = 30,\quad P_{50} = 80$

Solution. Here, we have

$P_{90} = 110,\quad P_{10} = 30,\quad P_{50} = 80$

$$\text{Kelly's Coefficient of Skewness} = \frac{P_{90} + P_{10} - 2\,Median}{P_{90} - P_{10}}$$

$$= \frac{110 + 30 - 2\,(80)}{110 - 30} = \frac{140 - 160}{80} = \frac{-20}{80} = \frac{-1}{4} = -0.25$$ **Ans.**

Example 20. *Calculate Kelly's Coefficient of Skewness for the following data :*

$D_1 = 60$, $D_9 = 290$, Median = 165

Solution. Here, we have

$D_1 = 60$, $D_9 = 290$, Median = 165

$$\text{Kelly's Coefficient of Skewness} = \frac{D_1 + D_9 - 2Median}{D_9 - D_1} \quad \text{... (1)}$$

Putting the values of D_1, D_9 and the median in (1), we get

$$\text{Kelly's Coefficient of Skewness} = \frac{60 + 290 - 2\,(165)}{290 - 60}$$

$$= \frac{350 - 330}{230} = \frac{20}{230} = \frac{2}{23} = 0.087 \quad \textbf{Ans.}$$

Example 21. *The weights in kg of 9 boys in a class are : 40, 42, 45, 48, 50, 52, 55, 56, 57. Calculate Kelly's Coefficient of Skewness based on percentiles.*

Solution. Here, we have 40, 42, 45, 48, 50, 52, 55, 56, 57

$n = 9$

$$P_{10} = \text{Value of } \frac{10}{100}\,(n+1)^{\text{th}} \text{ observation} = \text{Value of } \frac{10}{100}\,(9+1)^{\text{th}} \text{ observation}$$

= Value of first observation = 40

$$P_{50} = \text{Value of } \frac{50}{100}\,(n+1)^{\text{th}} \text{ observation} = \text{Value of } \frac{50}{100}\,(9+1)\text{th observation}$$

= Value of 5th observation = 50

$$P_{90} = \text{Value of } \frac{90}{100}\,(n+1)\text{th observation} = \text{Value of } \frac{90}{100}\,(9+1)\text{th observation}$$

= Value of 9th observation = 57

$$\text{Now, Kelly's Coefficient of Skewness} = \frac{P_{10} + P_{90} - 2 \text{ Median}}{P_{90} - P_{10}} = \frac{40 + 57 - 2\,(50)}{57 - 40}$$

$$= \frac{97 - 100}{17} = -\frac{3}{17} = -0.176 \quad \textbf{Ans.}$$

Example 22. *Calculate the Kelly's Coefficient of Skewness on the basis of percentiles for the following data.*

Marks obtained	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100
No. of students	3	8	9	14	16	18	8	4

Solution.

Marks	*No. of students*	*c.f*
20 – 30	3	3
30 – 40	8	11
40 – 50	9	20
50 – 60	14	34
60 – 70	16	50
70 – 80	18	68
80 – 90	8	76
90 – 100	4	80

$$P_{10} = \text{Marks of } 10\left(\frac{N}{100}\right)\text{th student} = \text{Marks of } 10\left(\frac{80}{100}\right)\text{th student}$$

$$= \text{Marks of 8 th student.}$$

P_{10} lies in the class 30 – 40

$$P_{10} = l + \frac{10\left(\frac{N}{100}\right) - c.f}{f}(i) = 30 + \frac{10\left(\frac{80}{100}\right) - 3}{8}(10) = 30 + \frac{8-3}{8}(10)$$

$$= 30 + \frac{50}{8} = 36.25$$

$$\text{Median} = \text{Marks of } \frac{80}{2} \text{ th student} = \text{Marks obtained by 40th student}$$

Median Class = 60 – 70.

$$\text{Median} = l + \frac{\frac{N}{2} - c.f.}{f}(i) = 60 + \frac{\frac{80}{2} - 34}{16}(10) = 60 + \frac{6}{16}(10) = 63.75$$

$$P_{90} = \text{Marks of } 90\left(\frac{N}{100}\right)\text{th student} = \text{Marks of } 90\left(\frac{80}{100}\right)\text{th student} = \text{Marks of 72nd student}$$

P_{90} class = 80 – 90

$$P_{90} = l + \frac{\frac{90}{100}(N) - c.f}{f}(i) = 80 + \frac{\frac{90}{100}(80) - 68}{8}(10) = 80 + \frac{72-68}{8}(10) = 80 + 5 = 85$$

$$\text{Kelly's Coefficient of Skewness} = \frac{P_{10} + P_{90} - 2\text{ Median}}{P_{90} - P_{10}}$$

$$= \frac{36.25 + 85 - 2\,(63.75)}{85 - 36.25} = \frac{-6.25}{48.75} = -0.128$$ **Ans.**

EXERCISE 57.3

Find Kelly's Coefficient of Skewness from the following table :

1. $P_{10} = 25$, $P_{90} = 200$, Median = 100 **Ans.** 0.143
2. $D_1 = 15.5$, $D_9 = 120.5$, Median = 70 **Ans.** – 0.038
3.

x	2.5	3.5	12.5	17.5
$c.f.$	7	18	25	30

On the basis of deciles. **Ans.** 0.866

4.

Rate (in Rs.)	Below 10	10 – 20	20 – 40	40 – 60	60 – 80	above 80
No. of workers	8	10	22	35	20	5

on the basis of deciles. **Ans.** – 0.07

5.

Wages (in Rs.)	800-900	900-1000	1000-1100	1100-1200	1200-1300	1300-1400	1400-1500
No. of workers	10	33	47	110	160	80	60

on the basis of percentiles. **Ans.** – 0.07

57.30 MOMENTS

The rth moment of a variable x about the mean $\bar{x}$ is usually denoted by μ_r is given by

$$\mu_r = \frac{1}{N}\Sigma f_i (x_i - \bar{x})^r, \qquad \Sigma f_i = N$$

The rth moment of a variable x about any point a is defined by $\mu'_r = \frac{1}{N}\Sigma f_i (x_i - a)^r$

57.31 MOMENT ABOUT MEAN

Let $\bar{x}$ be the arithmetic mean, then

$$\mu_r = \frac{1}{N}\sum_{i=1}^{a} f_i\,(x_i - \bar{x})^r, \quad r = 0,\ 1,\ 3 \ldots\ldots \qquad \text{where, } N = \sum_{i=1}^{n} f_i$$

If $r = 0$, $$\mu_0 = \frac{1}{N}\sum_{i=1}^{n} f_i\,(x_i - \bar{x})^0 = 1$$

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

If $r = 2$, $$\mu_2 = \frac{1}{N}\sum_{i=1}^{n} f_i\,(x_i - \bar{x})^2 \qquad [\mu_2 = \text{variance}]$$

If $r = 3$, $$\mu_3 = \frac{1}{N}\sum_{i=1}^{n} f_i\,(x_i - \bar{x})^3$$

If $r = 4$, $$\mu_4 = \frac{1}{N}\sum_{i=1}^{n} f_i\,(x_i - \bar{x})^4$$

57.32 MOMENTS ABOUT ANY NUMBER (RAW MOMENTS)

Let a be a arbitrary number then

$$\mu'_r = \frac{1}{N}\sum_{i=1}^{n} f_i\,(x_i - a)^r, \quad r = 0, 1, 2, \ldots\ldots \qquad \text{where } N = \sum_{i=1}^{n} f_i$$

If $r = 0$, $$\mu'_0 = \frac{1}{N}\sum_{i=1}^{n} f_i\,(x_i - a)^0 = 1$$

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

$$= \frac{1}{N}\sum_{i=1}^{n} f_i\,x_i - \frac{a}{N}\sum_{i=1}^{n} f_i$$

$$= x - a \qquad \left[\because \sum_{i=1}^{n} f_i = N\right]$$

If $r = 2$, $$\mu'_2 = \frac{1}{N}\sum_{i=1}^{n} f_i\,(x_i - a)^2$$

If $r = 3$, $$\mu'_3 = \frac{1}{N}\sum_{i=1}^{n} f_i\,(x_i - a)^3$$

If $r = 4$, $$\mu'_4 = \frac{1}{N}\sum_{i=1}^{n} f_i\,(x_i - a)^4$$

....

....

57.33 MOMENT ABOUT THE ORIGIN

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

If $r = 0$, $$v_0 = \frac{1}{N} \sum_{i=1}^{n} f_i x_i^0 = 1$$

If $r = 1$, $$v_1 = \frac{1}{N} \sum_{i=1}^{n} f_i x_i = \bar{x}$$

If $r = 2$, $$v_2 = \frac{1}{N} \sum_{i=1}^{n} f_i x_i^2$$

If $r = 3$, $$v_3 = \frac{1}{N} \sum_{i=1}^{n} f_i x_i^3$$

If $r = 4$, $$v_4 = \frac{1}{N} \sum_{i=1}^{n} f_i x_i^4$$

....

....

57.34 RELATION BETWEEN μ_r AND μ'_r :

We have,

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

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

$$= \frac{1}{N} \sum_{i=1}^{n} f_i [(x_i - a) - \mu'_1]^r \qquad [\mu'_1 = \bar{x} - a]$$

On expanding by Binomial theorem,

$$= \frac{1}{N} \sum_{i=1}^{n} f_i [(x_i - a)^r - {}^rC_1 (x_i - a)^{r-1} \mu'_1 + {}^nC_2 (x_i - a)^{r-2} \mu'^2_1 - \ldots\ldots + (-1)^r \mu'^r_1]$$

$$\mu_r = \mu'_r - {}^rC_1 \mu'_{r-1} \mu'_1 + {}^rC_2 \mu'_{r-2} \mu'^2_1 - \ldots\ldots + (-1)^r \mu'^r_1$$

Putting $r = 2, 3, 4, \ldots\ldots$ we get

$$\mu_2 = \mu'_2 - 2\mu'^2_1 + \mu'^2_1 = \mu_2^1 - \mu'^2_1 \qquad [\because \mu_0^1 = 1]$$

$$\mu_3 = \mu'_3 - 3\mu'_2 \mu'_1 + 3\mu'^3_1 - \mu'^3_1$$

$$= \mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_2$$

$$\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2\mu'^2_1 - 3\mu'^4_1$$

Thus, we have the following relations:

$$\boxed{\mu_1 = 0} \qquad \boxed{\mu_2 = \mu'_2 - \mu'^2_1}$$

$$\boxed{\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1} \qquad \boxed{\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2\mu'^2_1 - 3\mu'^4_1}$$

Note. 1. The sum of the coefficients of the various terms on the R.H.S. is zero.

2. The dimension of each term on R.H.S. is the same as that of terms on the L.H.S.

3. The number of terms on R.H.S. = order of the moment on L.H.S.

Conversely

$$\mu'_r = \frac{1}{N} \sum f_i (x_i - a)^r$$

$$= \frac{1}{N} \sum f_i (x_i - \bar{x} + \bar{x} - a)^r$$

$$= \frac{1}{N}\Big[\sum f_i (x_i - \bar{x})^r + {}^rC_1 \sum f_i (x_i - \bar{x})^{r-1} (\bar{x} - a) + {}^rC_2 \sum f_i (x_i - \bar{x})^{r-2} (\bar{x} - a)^2$$

$$\ldots + {}^rC_{r-1} \sum f_i (x_i - \bar{x}) (\bar{x} - a)^{r-1} + \sum f_i (\bar{x} - a)^r \Big]$$

$$= \frac{1}{N} \sum f_i (x_i - \bar{x})^r + {}^rC_1 \frac{1}{N} \sum f_i (x_i - \bar{x})^{r-1} (\bar{x} - a)$$

$$+ {}^rC_2 \frac{1}{N} \sum f_i (x_i - \bar{x})^{r-2} (\bar{x} - a)^2 + \ldots + {}^rC_{r-1} \frac{1}{N} \sum f_i (x_i - \bar{x}) (\bar{x} - a)^{r-1}$$

$$+ \frac{1}{N} \sum f_i (\bar{x} - a)^r$$

$\Rightarrow \qquad \mu'_r = \mu_r + r\,\mu_{r-1}\,\mu'_1 + \dfrac{r(r-1)}{2}\,\mu_{r-2}\,\mu'^2_1 + \ldots + r\,\mu_1\,\mu'^{r-1}_1 + \mu'^r_1$

If $r = 1$ $\qquad \mu'_1 = x - a = \mu_1 - a$

If $r = 2$, $\qquad \mu'_2 = \mu_2 + 2\mu_1\,\mu'_1 + \mu'^2_1 =$

$\qquad = \mu_2 + 0 + \mu'^2_1 \qquad [\mu_1 = 0]$

$\Rightarrow \qquad \mu'_2 = \mu_2 + \mu'^2_1$

If $r = 3$ $\qquad \mu'_3 = \mu_3 + 3\mu_2\,\mu'_1 + 3\mu_1\,\mu'^2_1 + \mu'^3_1$

$\qquad = \mu_3 + 3\mu_2\,\mu'_1 + 0 + \mu'^3_1 \qquad [\mu_1 = 0]$

$\Rightarrow \qquad \mu'_3 = \mu_3 + 3\mu_2\,\mu'_1 + \mu'^3_1$

If $r = 4$, $\qquad \mu'_4 = \mu_4 + 4\mu_3\,\mu'_1 + 6\mu_2\,\mu'^2_1 + 4\mu_1\,\mu'^3_1 + \mu'^4_1$

$\qquad = \mu_4 + 4\mu_3\,\mu'_1 + 6\mu_2\,\mu'^2_1 + 0 + \mu'^4_1$

$\Rightarrow \qquad \mu'_4 = \mu_4 + 4\mu_3\,\mu'_1 + 6\mu_2\,\mu'^2_1 + \mu'^4_1$

$$\boxed{\mu'_1 = \mu_1 - a} \qquad \boxed{\mu'_2 = \mu_2 + \mu'^2_1}$$

$$\boxed{\mu'_3 = \mu_3 + 3\mu_2\,\mu'_1 + \mu'^3_1} \qquad \boxed{\mu'_4 = \mu_4 + 4\mu_3\,\mu'_1 + 6\mu_2\,\mu'^2_1 + \mu'^4_1}$$

57.35 RELATION BETWEEN v_r AND μ_r

$$v_r = \frac{1}{N} \sum_{i=1}^{n} f_i\, x_i^r\ ; \qquad r = 0, 1, 2, \ldots\ldots$$

$$= \frac{1}{N} \sum_{i=1}^{n} f_i\, (x_i - a + a)^r$$

On expanding by binomial theorem, we have

$$\frac{1}{N}\sum_{i=1}^{n} f_i\,[(x_i-a)^r + {}^rC_1\,(x_i-a)^{r-1}\,a + \ldots.. + a^r]$$

$$= \mu'_r + {}^rC_1\,\mu'_{r-1}\,a + \ldots.. + a^r$$

On putting $a = \bar{x}$, we get $\qquad v_r = \mu_r + {}^rC_1\,\mu_{r-1}\bar{x} + {}^rC_2\,\mu_{r-2}\,\bar{x}^2 + \ldots.. + \bar{x}^r \qquad \ldots(1)$

On taking r = 1, 2, 3, 4 in (1), we get

$$v_1 = \mu_1 + \mu_0\bar{x} = \bar{x} \qquad [\mu_1 = 0,\ \mu_0 = 1]$$

$$v_2 = \mu_2 + {}^2C_1\,\mu_1\bar{x} + {}^2C_2\mu_0\,\bar{x}^2 = \mu_2 + \bar{x}^2$$

$$v_3 = \mu_3 + {}^3C_1\,\mu_2\,\bar{x} + {}^3C_2\,\mu_1\,\bar{x}^2 + {}^3C_3\,\mu_0\,\bar{x}^3 = \mu_3 + 3\mu_2\,\bar{x} + \bar{x}^3$$

$$v_4 = \mu_4 + {}^4C_1\,\mu_3\,\bar{x} + {}^4C_2\,\mu_2\,\bar{x}^2 + {}^4C_3\,\mu_1\,\bar{x}^3 + {}^4C_4\,\mu_0\,\bar{x}^4$$

$$= \mu_4 + 4\mu_3\,\bar{x} + 6\mu_2\,\bar{x}^2 + \bar{x}^4$$

$$\boxed{v_1 = \bar{x}} \qquad \boxed{v_2 = \mu_2 + \bar{x}^2}$$

$$\boxed{v_3 = \mu_3 + 3\mu_2\bar{x} + \bar{x}^3} \qquad \boxed{v_4 = \mu_4 + 4\mu_3\bar{x} + 6\mu_2\bar{x}^2 + \bar{x}^4}$$

57.36 MEASURE OF SKEWNESS BASED ON MOMENT

1. Measure of skewness is given by β_1 where

$$\boxed{\beta_1 = \frac{\mu_3^2}{\mu_2^3}}$$

Karl Pearson's coefficient of skewness $= \pm\dfrac{\mu_3}{\mu_2^{3/2}}$

The sign of Karl Pearson's coefficient of skewness is determined from the sign of μ_3.

2. Measure of Kurtosis is given by β_2, where

$$\boxed{\beta_2 = \frac{\mu_4}{\mu_2^2}}$$

3. Gamma Coefficients

$$\boxed{\gamma_1 = \pm\sqrt{\beta_1}}$$

$$\boxed{\gamma_2 = \beta_2 - 3}$$

Example 23. *Find the relation between moment about the mean and moment about any arbitary point. The first four moments of a distribution about the value 4 of the variate are – 1.5, 17, – 30 and 108. Calculate the first four moments about the mean and find β_1 and β_2.* (*Uttarakhand, III Semester, 2008*)

Solution. We have,

$a = 4,\quad \mu_1' = -1.5,\quad \mu_2' = 17,\quad \mu_3' = -30$ and $\mu_4' = 108$

Moment about the mean

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - (\mu_1')^2 = 17 - (-1.5)^2 = 17 - 2.25 = 14.75$$

$$\mu_3 = \mu_3' - 3\mu_2'\,\mu_1' + 2\,(\mu_1')^3$$
$$= -30 - 3\,(17)\,(-1.5) + 2\,(-1.5)^3 = -30 + 76.5 - 6.75$$
$$= 39.75$$

$$\mu_4 = \mu_4' - 4\mu_3'\,\mu_1' + 6\,\mu_2'\,\mu_1'^2 - 3\,\mu_1'^4$$
$$= 108 - 4\,(-30)\,(-1.5) + 6\,(17)\,(-1.5)^2 - 3\,(-1.5)^4$$
$$= 108 - 180 + 229.5 - 15.19 = 142.31$$

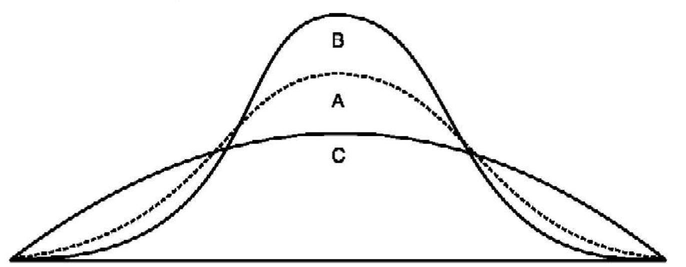
$$\beta_1 = \frac{(\mu_3)^2}{(\mu_2)^3} = \frac{(39.75)^2}{(14.75)^3} = \frac{1580.06}{3209.05} = 0.4924$$

$$\beta_2 = \frac{\mu_4}{(\mu_2)^2} = \frac{142.31}{(14.75)^2} = \frac{142.31}{217.56} = 0.6541$$ **Ans.**

57.37 KURTOSIS (*U.P. III, Semester Dec. 2006*)

It measures the degree of peakedness of a distribution and is given by Measure of kurtosis

$$\beta_2 = \frac{\mu_4}{\mu_2^2}, \quad \mu_2 = \frac{\Sigma\,(x-\bar{x})^2}{N}, \quad \mu_4 = \frac{\Sigma\,(x-\bar{x})^4}{N}$$



A Mesokurtic B Leptokurtic C Platykurtic

If $\beta_2 = 3$, the curve is normal or mesokurtic.

If $\beta_2 > 3$, the curve is peaked or leptokurtic.

If $\beta_2 < 3$, the curve is flat topped or platykurtic.

$$\gamma_2 = \beta_2 - 3$$

57.38 GROUPING ERROR AND ITS SHEPPARD'S CORRECTIONS (FOR MOMENTS)

If the distribution is not symmetrical and the number of class intervals is greater than $\frac{1}{20}$th of the range, then the computation of moments will have an error known as grouping error.

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

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

where h is the width of the class-interval while μ_1 and μ_3 require no correction.

These formulae are known as **Sheppard's corrections.**

Example 24. *Find the corrected values of the following moments using Sheppard's correction. The width of classes in the distribution is 10 :*

$$\mu_2 = 210, \quad \mu_3 = 460, \quad \mu_4 = 96700, \quad h = 12$$

Solution. We have,

$$\mu_2 = 210, \quad \mu_3 = 460, \quad \mu_4 = 96700, \quad h = 12$$

$$\text{corrected } \mu_2 = \mu_2 - \frac{h^2}{12} = 210 - \frac{(12)^2}{12} = 210 - 12 = 198$$

$$\text{corrected } \mu_3 = 460$$

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

$$= 96700 - \frac{(12)^2}{2}(210) + \frac{7}{240}(12)^4$$

$$= 96700 - 15120 + 604.8 = 82184.8 \qquad \textbf{Ans.}$$

Example 25. *Calculate the variance and third central moment from the following data :*

x_i	0	1	2	3	4	5	6	7	8
f_i	1	9	26	59	72	52	29	7	1

(*U.P. III Semester Dec. 2005*)

Solution. Let mean = 4

x_i	f_i	$x_i - 4$	$f_i(x_i - 4)$	$f_i(x_i - 4)^2$	$f_i(x_i - 4)^3$
0	1	-4	-4	16	-64
1	9	-3	-27	81	-243
2	26	-2	-52	104	-208
3	59	1	59	59	59
4	72	0	0	0	0
5	52	1	52	52	52
6	29	2	58	116	232
7	7	3	21	63	189
8	1	4	4	16	64
	$\Sigma f_i = 256$		$\Sigma fi(x_i - 4) = -7$	$\Sigma fi(x_i - 4)^2 = 507$	$\Sigma fi(x_i - 4)^3 = -37$

$$\mu'_1 = \frac{\Sigma f_i (x_i - 4)}{\Sigma f_i} = \frac{-7}{256}$$

$$\mu'_2 = \frac{\Sigma f_i (x_i - 4)^2}{\Sigma f_i} = \frac{507}{256}$$

$$\mu'_3 = \frac{\Sigma f_i (x_i - 4)^3}{\Sigma f_i} = \frac{-37}{256}$$

$$\mu_2 = \mu'_2 - \mu'^2_1 = \frac{507}{256} - \left(\frac{-7}{256}\right)^2 = 1.98047 - 0.00075 = 1.97972$$

$$\mu_2 = \mu'_2 \quad \mu'^2_1 = \frac{507}{256} \quad \left(\frac{7}{256}\right)^2 = 1.98047 \quad 0.00075 = 1.97972$$

$$\mu_3 = \mu'_3 - 3\mu'_2 \cdot \mu'_1 + 2(\mu'_1)^3$$

$$= \frac{-37}{256} - 3\left(\frac{507}{256}\right)\left(-\frac{7}{256}\right) + 2\left(-\frac{7}{256}\right)^3$$

$$= -0.14453 + 0.16246 - 0.00004$$

$$= 0.01789$$ **Ans.**

Example 26. *Calculate* $\mu_1, \mu_2, \mu_3, \mu_4$ *for the following frequency distribution :*

Marks	*0-10*	*10-20*	*20-30*	*30-40*	*40-50*	*50-60*
No. of students	*1*	*6*	*10*	*15*	*11*	*7*

Solution.

Marks	*No. of students* f	*Mid value* x	fx	$x \quad \bar{x}$	$f(x-\bar{x})$	$f(x-\bar{x})^2$	$f(x-\bar{x})^3$	$f(x-\bar{x})^4$
0-10	1	5	5	– 30	– 30	900	– 27000	810000
10-20	6	15	90	– 20	– 120	2400	– 48000	960000
20-30	10	25	250	– 10	– 100	1000	– 10000	100000
30-40	15	35	525	0	0	0	0	0
40-50	11	45	495	10	110	1100	11000	110000
50-60	7	55	385	20	140	2800	56000	1120000
	$\Sigma f - 50$		$\Sigma fx - 1750$		$\Sigma f(x \quad \bar{x}) = 0$	$\Sigma f(x \quad \bar{x})^2 = 8200$	$\Sigma f(x \quad \bar{x})^3 = -18000$	$\Sigma f(x \quad \bar{x})^4 = 3100000$

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{1750}{50} = 35$$

$$\mu_1 - \frac{\Sigma f(x-x)}{\Sigma f} - \frac{0}{50} - 0$$

$$\mu_2 - \frac{\Sigma f(x-\bar{x})^2}{\Sigma f} - \frac{8200}{50} - 164$$

$$\mu_3 = \frac{\Sigma f(x \quad \bar{x})^3}{\Sigma f} = \frac{-18000}{50} = -360$$

$$\mu_4 - \frac{\Sigma f(x-x)^4}{\Sigma f} - \frac{3100000}{50} - 62000$$ **Ans.**

Example 27. *Find out the kurtosis of the data given below :*

Class-interval	*0-10*	*10-20*	*20-30*	*30-40*
Frequency	*1*	*3*	*4*	*2*

Solution. Let assumed mean be 25

Class	*Frequency* f_i	*Mid value* x_i	$x_i - 25$	$f_i\,(x_i - 25)$	$f_i(x_i - 25)^2$	$f_i\,(x_i - 25)^3$	$f_i(x_i - 25)^4$
0-10	1	5	−20	−20	400	−8000	160000
10-20	3	15	−10	−30	300	−3000	30000
20-30	4	25	0	0	0	0	0
30-40	2	35	10	20	200	2000	20000
	$\Sigma f - 10$			$\Sigma f_i(x_i - 25) = -30$	$\Sigma f_i\,(x_i - 25)^2 = 900$	$\Sigma f_i\,(x_i - 25)^3 = -9000$	$\Sigma f_i\,(x_i - 25)^4 = 210000$

$$\mu'_1 = \frac{\Sigma f_i\,(x_i \quad 25)}{\Sigma f_i} = \frac{-30}{10} - -3$$

$$\mu'_2 = \frac{\Sigma f_i\,(x_i - 25)^2}{\Sigma f_i} = \frac{900}{10} = 90$$

$$\mu'_3 - \frac{\Sigma f_i\,(x_i - 25)^3}{\Sigma f_i} - \frac{-9000}{10} = -900$$

$$\mu'_4 = \frac{\Sigma f_i\,(x_i - 25)^4}{\Sigma f_i} = \frac{210000}{10} = 21000$$

$$\mu_2 = \mu'_2 - \mu'^2_1 = 90 - (-3)^2 - 90 - 9 = 81 \quad \text{[Article 57.34 on page 1560]}$$

$$\mu_4 = \mu'_4 - 4\,\mu'_3\,\mu'_1 + 6\mu'_2\,\mu'^2_1 - 3\mu'^4_1$$

$$= 21000 - 4\,(-900)\,(-3) + 6\,(90)\,(-3)^2 - 3\,(-3)^4$$

$$= 21000 - 10800 + 4860 - 243 = 14{,}817$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{14{,}817}{(81)^2} = \frac{14{,}817}{6561} = 2.258$$

$$\gamma_2 = \beta_2 - 3 = 2.258 - 3 = -0.742$$ **Ans.**

Example 28. *Calculate the first four moments of the following data. Also make sheppard's correction.*

Values	*10-20*	*30-30*	*30-30*	*40-50*	*50-60*	*60-70*	*70-80*
Frequency	*1*	*20*	*69*	*108*	*78*	*22*	*2*

Solution. **Calculation of Moments**

Values	f	x	$d = x - 45$	$\frac{x-45}{10}$	$f_i\left(\frac{x-45}{10}\right)$	$f_i\left(\frac{x-45}{10}\right)^2$	$f_i\left(\frac{x-45}{10}\right)^3$	$f_i\left(\frac{x-45}{10}\right)^4$
10–20	1	15	– 30	– 3	– 3	9	– 27	81
20–30	20	25	– 20	– 2	– 40	80	– 160	320
30–40	69	35	– 10	– 1	– 69	69	– 69	69
40–50	108	45	0	0	0	0	0	0
50–60	78	55	10	1	78	78	78	78
60–70	22	65	20	2	44	88	176	352
70–80	2	75	30	3	6	18	54	162
	N = 300				$\Sigma f_i\left(\frac{x-45}{10}\right) = 16$	$\Sigma f_i\left(\frac{x-45}{10}\right)^2 = 342$	$\Sigma f_i\left(\frac{x-45}{10}\right)^3 = 52$	$\Sigma f_i\left(\frac{x-45}{10}\right)^4 = 1062$

Now
$$\mu'_1 = \frac{1}{N}\Sigma f_i\left(\frac{x-45}{10}\right) \cdot 10 = \left(\frac{16}{300}\right) 10 = 0.53$$
$$\mu'_2 = \frac{1}{N}\Sigma f_i\left(\frac{x-45}{10}\right)^2 \cdot 10^2 = \left(\frac{342}{300}\right)(10)^2 = 114$$
$$\mu'_3 = \frac{1}{N}\Sigma f_i\left(\frac{x-45}{10}\right)^3 \cdot 10^3 = \left(\frac{52}{300}\right)(10)^3 = 173.33$$
$$\mu'_4 = \frac{1}{N}\Sigma f_i\left(\frac{x-45}{10}\right)^4 \cdot 10^4 = \left(\frac{1062}{300}\right)(10)^4 = 35400.$$

Moments about mean

$$\mu_1 = 0 \text{ (always)}$$
$$\mu_2 = \mu'_2 - \mu'^2_1 = 114 - (0.53)^2 = 113.72$$
$$\mu_3 = \mu'_3 - 3\mu'_2\,\mu'_1 + 2\mu'^3_1 = 173.33 - 3(114)(0.53) + 2(0.53)^3 = -7.63$$
$$\mu_4 = \mu'_4 - 4\mu'_3\,\mu'_1 + 6\mu'_2\,\mu'^2_1 - 3\mu'^4_1$$ [Article 57.34 on page 1560]
$$= 35400 - 4(173.33)(0.53) + 6(114)(0.53)^2 - 3(0.53)^4 = 35224.44.$$

Sheppard's Correction, h = 10

$$\mu_1 \text{ (Corrected)} = \mu_1 = 0$$
$$\mu_2 \text{ (Corrected)} = \mu_2 - \frac{h^2}{12} = 113.72 - \frac{(10)^2}{12} = 105.39$$
$$\mu_3 \text{ (Corrected)} = \mu_3 = -7.63$$
$$\mu_4 \text{ (Corrected)} = \mu_4 - \frac{1}{2}h^2\,\mu_2 + \frac{7}{240}h^4$$
$$= 35224.44 - \frac{1}{2}(10)^2(113.72) + \frac{7}{240}(10)^4 = 29830.11$$ **Ans.**

Example 29. *The first four moments of a distribution about the value 4 of the variable arc – 1.5, 17, – 30 and 108. Find the moments about mean,* β_1 *and* β_2.

Find also the moments about (i) the origin, and (ii) the point x = 2.

(U.P. III, Semester Dec. 2006)

Solution. In the usual notation, we are given assumed mean a = 4 and

$$\mu'_1 = \quad 1.5, \quad \mu'_2 = 17, \quad \mu'_3 = \quad 30 \quad \text{and} \quad \mu'_4 = 108.$$

(*a*) Moments about mean:

$$\mu_2 = \mu'_2 - \mu'^2_1 = 17 - (-1.5)^2 = 17 - 2.25 = 14.75$$

$$\mu_3 - \mu'_3 - 3\mu'_2\,\mu'_1 + 2\mu'^3_1 \qquad \text{[Article 57.34 on page 1560]}$$

$$- -30 - 3 \times (17) \times (-1.5) + 2\,(-1.5)^3$$

$$= -30 + 76.5 - 6.75 = 39.75$$

$$\mu_4 = \mu'_4 - 4\mu'_3\,\mu'_1 + 6\mu'_2\,\mu'^2_1 - 3\mu'^4_1$$

$$= 108 - 4(-30)(-1.5) + 6(17)(-1.5)^2 - 3(-1.5)^4$$

$$= 108 \quad 180 \mid 229.5 \quad 15.1875 = 142.3125$$

Hence $$\beta_1 - \frac{\mu_3^2}{\mu_2^3} = \frac{(39.75)^2}{(14.75)^3} = 0.4924$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{142.3125}{(14.75)^2} = 0.6541$$

Also $$\bar{x} = a + \mu'_1 = 4 + (-1.5) = 2.5$$

(*b*) *Moments about origin.*

We have moments about x

$\bar{x} = 2.5$, $\mu_2 = 14.75$, $\mu_3 = 39.75$ and $\mu_4 = 142.31$ (approx.)

We know that $\bar{x} = a + \mu'_1$, where μ'_1 is the first moment about the point $x = a$.

Taking a = 0, we get the first moment about origin as $\mu'_1 = \bar{x}$ = mean = 2.5.

[Using Article 57.35 of page 1561]

$$v_1 = \bar{x} = 2.5$$

$$v_2 = \mu_2 + (\bar{x})^2 = 14.75 + (2.5)^2 = 14.75 + 6.25 = 21$$

$$v_3 - \mu_3 + 3\mu_2\,x + (x)^3 - 39.75 + 3(14.75)(2.5) + (2.5)^3$$

$$- 39.75 + 110.625 + 15.625 - 166$$

$$v_4 - \mu_4 + 4\mu_3\,\bar{x} + 6\mu_2\,(\bar{x})^2 + (\bar{x})^4$$

$$- 142.3125 + 4(39.75)(2.5) + 6(14.75)(2.5)^2 + (2.5)^4$$

$$= 142.3125 + 397.5 + 553.125 + 39.0625$$

$$= 1132.$$

Moments about the point x = 2.

We have $\bar{x} = a + \mu'_1$. Taking a = 2, the first moment about the point x = 2 is

$$\mu'_1 = \bar{x} - 2 = 2.5 - 2 = 0.5 \qquad \text{[Converse of Art. 57.34 on page 1560]}$$

Hence, $$\mu'_2 - \mu_2 + \mu'^2_1 - 14.75 + 0.25 - 15$$

$$\mu'_3 = \mu_3 \mid 3\mu_2\,\mu'_1 \mid \mu'^3_1 = 39.75 \mid 3(14.75)(0.5) \mid (0.5)^3$$

$$- 39.75 + 22.125 + 0.125 - 62$$

$$\mu'_4 = \mu_4 + 4\mu_3\,\mu'_1 + 6\mu_2\,\mu'^2_1 + \mu'^4_1$$
$$= 142.3125 + 4(39.75)(0.5) + 6(14.75)(0.5)^2 + (0.5)^4$$
$$= 142.3125 + 79.5 + 22.125 + 0.0625 = 244$$

Ans.

Example 30. *The first three moments about the origin are given by*

$$\mu'_1 = \frac{n+1}{2},\quad \mu'_2 = \frac{(n+1)(2n+1)}{6} \quad and \quad \mu'_3 = \frac{n(n+1)^2}{4}$$

Examine the skewness of the data.

Solution.

$$\mu'_1 = \frac{n+1}{2},\quad \mu'_2 = \frac{(n+1)(2n+1)}{6},\quad \mu'_3 = \frac{n(n+1)^2}{4}$$

$$\mu_2 = \mu'_2 - \mu'^2_1$$
$$= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}$$
$$= \frac{2(2n^2+3n+1) - 3(n^2+2n+1)}{12} = \frac{n^2-1}{12}$$

$$\mu_3 = \mu'_3 - 3\mu'_2\,\mu'_1 + 2\mu'^3_1$$
$$= \frac{n(n+1)^2}{4} - \frac{3(n+1)(2n+1)}{6} \times \frac{(n+1)}{2} + 2\left(\frac{n+1}{2}\right)^3$$
$$= \frac{n^3+2n^2+n}{4} - \frac{2n^3+5n^2+4n+1}{4} + \frac{n^3+3n^2+3n+1}{4}$$
$$= 0$$

$$\text{Coefficient of skewness} = \gamma_1 = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{0}{\sqrt{\mu_2^3}} = 0$$

The data is symmetrical. **Ans.**

Example 31. *The first three moments of a distribution, about the value '2' of the variable are 1, 16 and − 40. Show that the mean is 3, variance is 15 and* $\mu_3 = -86$.

Solution. We have

$$a = 2,\ \mu'_1 = 1,\ \mu'_2 = 16 \text{ and } \mu'_3 = -40$$

We know that $\mu'_1 = \bar{x} - a \Rightarrow \bar{x} = \mu'_1 + a = 1 + 2 = 3$

$$\text{Variance} = \mu_2 = \mu'_2 - \mu'^2_1 = 16 - (1)^2 = 15$$

$$\mu_3 = \mu'_3 - 3\mu'_2\,\mu'_1 + 2\mu'^3_1 = -40 - 3(16)(1) + 2(1)^3 = -40 - 48 + 2 = -86.$$

Example 32. *The first four moments of a distribution, about the value '35' are − 1.8, 240, − 1020 and 144000. Find the values of* $\mu_1, \mu_2, \mu_3, \mu_4$.

Solution. We have,

$$\mu'_1 = -1.8,\quad \mu'_2 = 240,\quad \mu'_3 = -1020,\quad \mu'_4 = 144000$$
$$\mu_1 = 0$$
$$\mu_2 = \mu'_2 - \mu'^2_1 = 240 - (-1.8)^2 = 236.76 \qquad \text{(Article 57.34 on page 1560]}$$
$$\mu_3 = \mu'_3 - 3\mu'_2\,\mu'_1 + 2\mu'^3_1 = -1020 - 3(240)(-1.8) + 2(-1.8)^3 = 264.36$$
$$\mu_4 = \mu'_4 - 4\mu'_3\,\mu'_1 + 6\mu'_2\,\mu'^2_1 - 3\mu'^4_1$$
$$= 144000 - 4(-1020)(-1.8) + 6(240)(-1.8)^2 - 3(-1.8)^4 = 141290.11.$$

Ans.

Example 33. *For a distribution, the mean is 10, variance is 16, γ_1 is 1, and β_2 is 4. Find the first four moments about the origin.*

Solution. We have,

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

Now, $\gamma_1 = 1 \Rightarrow \sqrt{\beta_1} = 1 \Rightarrow \beta_1 = 1$

$\Rightarrow$ $\beta_1 = 1 \Rightarrow \dfrac{\mu_3^2}{\mu_2^3} = 1 \Rightarrow \mu_3^2 = \mu_2^3 = (16)^3 = (64)^2$

$\Rightarrow$ $\mu_3^2 = (64)^2 \Rightarrow \mu_3 = 64$

and $\mu_2^3 = (16)^3 \Rightarrow \mu_2 = 16$

$\beta_2 = 4$

$\Rightarrow$ $\dfrac{\mu_4}{\mu_2{}^2} = 4 \Rightarrow \mu_4 = 4(16)^2 = 1024$

Moments about the origin

$$v_1 = \bar{x} = 10$$ (See Art. 57.33 on page 1560]

$$v_2 = \mu_2 + \bar{x}^2 = 16 + 100 = 116$$

$$v_3 = \mu_3 + 3\mu_2\bar{x} + \bar{x}^3 = 64 + 3(16)(10) + (10)^3 = 64 + 480 + 1000 = 1544$$

$$v_4 = \mu_4 + 4\mu_3\bar{x} + 6\mu_2\bar{x}^2 + \bar{x}^4 = 1024 + 4(64)(10) + 6(16)(100) + (10)^4$$

$$= 1024 + 2560 + 9600 + 10000 = 23184$$ **Ans.**

EXERCISE 57.4

1. Find the mean and s.d. of the following series:

Expenditure	No. of students
Below Rs. 5	6
" 10	16
" 15	28
" 20	38
" 25	46

Ans. Mean = Rs. 1293, S.D. = Rs. 6.41

2. Calculate the arithmetic mean and the standard deviation of the following values of the world's annual gold output (in millions of pounds) for 10 different years:

Year	94	95	96	93	87	79	73	69	68	67
Gold output	78	82	83	89	95	103	108	117	130	97

Ans. Mean = 80.50, standard deviation = 9.27

3. For a frequency distribution of marks in History of 200 candidates (grouped in intervals 0–5, 5–10, ...) the mean and standard deviation (s.d.) were found to be 40 and 15. Later it was discovered that the score 43 was misread as 53 in obtaining the frequency distribution. Find the corrected mean and s.d. corresponding to the corrected frequency distribution.

Ans. Mean = 39.95, Standard deviation = 14.975 approx.

4. A student while calculating the mean and the standard deviation on 25 observations, obtained the following values:

mean = 56 cms. : standard deviation = 2 cms.

It was later discovered at the time of checking that he had wrongly copied down an observation as 64. What is the mean and s.d. if correct value is omitted ?

Ans. Mean = 55.67 cms., S.D. = 1.18 cms. approx

5. Calculate first four moments about mean, for the following individual series:

5	5	5	5	5	5

Ans. $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$

6. Calculate the first four moments about the mean for the following data. Also calculate β_1 and β_2.

x	1	2	3	4	5	6	7	8	9
f	1	6	13	25	30	22	9	5	2

Ans. $\mu'_1 = -0.09$, $\mu_2 = 2.4873$, $\mu_3 = 0.6789$, $\mu_4 = 18.3358$ approx.,
$\beta_1 - 0.0299$ approx. $\beta_2 - 2.9627$

7. Find the first four moments about mean for the following frequency distribution:

Marks	0–10	10–20	20–30	30–40	40–50
No. of Students	5	10	40	20	25

Ans. $\mu_1 = 0$, $\mu_2 = 125$, $\mu_3 = -300$, $\mu_4 = 37625$

8. The following table gives the monthly wages of workers in a factory. Compute the standard deviation, and skewness

Monthly wages (in Rs.)	*No. of workers*	*Monthly wages* (in Rs.)	*No. of workers*
125–175	2	375–425	4
175–225	22	425–475	6
225–275	19	475–525	1
275–325	14	525–575	1
325 375	3		

Ans. S.D. = Rs. 88.52, Skewness = 0.7

9. The first four moments of a distribution about x = 4 are 1, 4, 10, 45. Show that the mean is 5 and the variance is 3 and μ_3 and μ_4 are 0 and 26 respectively.

10. Calculate $\mu_1, \mu_2, \mu_3, \mu_4$ for the series : 4, 7, 10, 13, 16, 19, 22.

Ans. $\mu_1 = 0$, $\mu_2 = 36$, $\mu_3 = 0$, $\mu_4 = 2268$

11. If the first four moments of a distribution about the value 5 are equal to -4, 22, -117 and 560. Determine the corresponding moments:
(*i*) about the mean, and (*ii*) about zero **Ans.** (*i*) 0, 6, 83, 992 (*ii*) 1, 7, 102, 1361

12. In a certain distribution, the first four moments about the point $x = 4$ are -1.5, 17, -30 and 308. Calculate β_1 and β_2. **Ans.** $\beta_1 = 0.4923$, $\beta_2 = 1.573$

13. Compute first four moments of the data 3, 5, 7, 9 about the mean. Also, compute the first four moments about the point 4. **Ans.** $\mu_1 = 0, \mu_2 = 5, \mu_3 = 0, \mu_4 = 41$
$\mu'_1 = 2, \mu'_2 = 9, \mu'_3 = 38, \mu'_4 = 177$

14. The first four moments of distribution about the value 5 of the variable are 2, 20, 40 and 50. Calculate mean, μ_2, μ_3, μ_4, β_1 and β_2. **Ans.** Mean = 7, $\mu_2 = 16, \mu_3 = -64, \mu_4 = 162, \beta_1 = 1$ and $\beta_2 = 0.63$

15. Show that, if the variable takes the values 0, 1, 2, ..., n with frequencies proportional to the binomial coefficients, 1, ${}^nC_1, {}^nC_2, \ldots {}^nC_n$ respectively then the mean of the distribution is $\frac{1}{2}n$, the mean square deviation about origin is $\frac{1}{4}n(n+1)$ and the variance is $\frac{1}{4}n$.

16. Show that, if the variable takes the values 0, 1, 2, ..., n with frequencies given by the terms of the binomial series $q^n, {}^nC_1 q^{n-1}p, {}^nC_2 q^{n-2}p^2, \ldots, p^n$ where p + q = 1, then the mean square deviation is $n^2p^2 + npq$ and the variance is npq.

57.39 MOMENT GENERATING FUNCTION

The moment generating function of the variate x about $x = a$ is defined as the expected value of $e^{t(x-a)}$ and is denoted by $M_a(t)$.

$$M_a(t) = \Sigma p_i e^{t(x_i - a)}$$
$$= \Sigma p_i \left[1 + t(x_i - a) + \frac{t^2}{2}(x_i - a)^2 + \ldots + \frac{t^r}{r!}(x_i - a)^r + \ldots\right]$$
$$= \Sigma p_i + t\Sigma p_i (x_i - a) + \frac{t^2}{2}\Sigma p_i (x_i - a)^2 + \ldots \frac{t^r}{r!}\Sigma p_i (x_i - a)^r + \ldots$$
$$= 1 + t\Sigma f_i (x_i - a) + \frac{t^2}{2}\Sigma f_i (x_i - a)^2 + \ldots + \frac{t^r}{r!}\Sigma f_i (x_i - a)^r + \ldots \qquad [p = f]$$
$$= \mu_0 + t\mu'_1 + \frac{t^2}{2!}\mu'_2 + \ldots + \frac{t^r}{r!}\mu'_r + \ldots \qquad [\Sigma p_i = 1 = \mu_0]$$

where μ'_r is the moment of order r about a.

Hence μ'_r = coefficient of $\frac{t^r}{r!}$ or $\mu'_r = \left[\frac{d^r}{dt^r} M_a(t)\right]_{t=0}$

Again $M_a(t) = \Sigma p_i e^{t(x_i - a)}$
$$= e^{-at}\Sigma p_i e^{tx_i}$$
$$= e^{-at} M_0(t)$$

Thus the moment generating function about the point $a = e^{at}$ moment generating funtion about the origin.

Example 34. *Find the moment generating function of the discrete binomial distribution given by*

$$f(x) = {}^nC_x p^x q^{n-x}$$

Also, find the first and second moment about the mean and standard deviation.

Solution. Here, we have

$$f(x) = {}^nC_x\, p^x\, q^{n-x}$$

Moment generating function about the origin

$$= M_0(t) = \sum e^{tx} \,.\, {}^nC_x\, p^x\, q^{n-x}$$

$$= \sum {}^nC_x\, (pe^t)^x \,.\, q^{n-x}$$

$$= q^n + {}^nC_1\, q^{n-1}\, pe^t + {}^nC_2\, q^{n-2}\, (pe^t)^2 + \ldots\ldots \qquad \text{[By Binomial Theorem]}$$

$$= [q + pe^t]^n$$

$$v_1 = \left[\frac{d}{dt} M_0(t)\right]_{t=0} = [n\,(q + pe^t)^{n-1} \,.\, (pe^t)]_{t=0}$$

$$= n\,(q + p)^{n-1}\, p \qquad [q + p = 1]$$

$$= np$$

$$v_2 = \left[\frac{d^2}{dt^2} M_0(t)\right] = \frac{d}{dt}\, [n\,(q + pe^t)^{n-1}\,(pe^t)]_{t=0}$$

$$= [n\,(n-1)\,(q + pe^t)^{n-2}\,(pe^t)^2 + n\,(q + pe^t)^{n-1}\,(pe^t)]_{t=0}$$

$$= [n\,(n-1)\,(q + p)^{n-2}\,p^2 + n\,(q + p)^{n-1} \,.\, p]$$

$$= n\,(n-1)\,p^2 + np \qquad [q + p = 1]$$

$$= n\,p\,[(n-1)\,p + 1] = np\,[np + (1 - p)] = np\,[np + q]$$

$$= n^2p^2 + npq$$

$$\mu_1 = \bar{x} = v_1 = np$$

$$\mu_2 = \mu'_2 - \bar{x}^2 = v_2 - v_1^2 = (n^2p^2 + npq) - (np)^2$$

$$\Rightarrow \qquad \mu_2 = npq$$

Standard deviation $= \sqrt{npq}$

Mean $= np$ **Ans.**

Example 35. *Find the moment generating function of the discrete Poisson distribution given by*

$$f(x) = \frac{e^{-m} \,.\, m^x}{x!}$$

Also, find the first and second moments about mean and variance.

Solution. Here, we have

$$f(x) = \frac{e^{-m}\, m^x}{x!}$$

Moment generating function about the origin

$$= M_0(t) = \sum e^{tx}\, \frac{e^{-m}\, m^x}{x!}$$

$$= e^{-m} \sum \frac{(me^t)^x}{x!}$$

$$= e^{-m}\left[1 + me^t + \frac{(me^t)^2}{2!} + \frac{(me^t)^3}{3!} + \ldots\right]$$

$$= e^{-m} . e^{me^t} = e^{m(e^t - 1)}$$

$$v_1 = \left[\frac{d}{dt} M_0(t)\right]_{t=0} = \left[\frac{d}{dt} e^{m(e^t-1)}\right]_{t=0} = \left[e^{m(e^t-1)}\, me^t\right]_{t=0} = e^{m(1-1)}.m = m$$

$$v_2 = \left[\frac{d^2}{dt^2} M_0(t)\right]_{t=0} = \frac{d}{dt}\left[e^{m(e^t-1)} me^t\right]_{t=0} = \left[e^{m\,(e^t-1)}(me^t)^2 + e^{m(e^t-1)}\,.\,me^t\right]_{t=0}$$

$$= \left[e^{m\,(1-1)}m^2 + e^{m(1-1)}\,.\,m\right] = m^2 + m$$

$$\mu_2 = v_2 - \bar{x}^2 = v_2 - v_1^2 = (m^2 + m) - m^2 = m$$

Hence, mean $= m$

Variance $= m$ **Ans.**

Example 36. *Find the moment generating function of the random variable whose moments are*

$$\mu_r' = (r+1)!\, 2^r$$

Solution. $M_x(t) = E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx} P(X=x) = \sum_{r=0}^{\infty} \frac{t^r}{r!}\mu'_r = \sum_{r=0}^{\infty} \frac{t^r}{r!}(r+1)!\,.\,2^r = \sum_{r=0}^{\infty} (r+1)\,(2t)^r$

$$= 1 + 2.2t + 3.\,(2t)^2 + \ldots = (1-2t)^{-2}$$ **Ans.**

57.40 MOMENT GENERATING FUNCTION OF A FUNCTION OF CONTINUOUS VARIATE

The moment generating function of the continuous probability distribution about $x = a$ is given by

$$M_0(t) = \int_{-\infty}^{\infty} e^{t\,(x-a)} f(x)\, dx$$

57.41 PROPERTY OF MOMENT GENERATING FUNCTION

The moment generating function of the sum of two independent chance variables is the product of their respective moment generating functions.

Symbolically, $M_{x+y}(t) = M_x(t) \times M_y(t)$ provided that x and y are independent random variables.

Proof. Let x and y be two independent random variables so that $x + y$ is also a random variable.

The m.g.f. of the sum $x + y$ w.r.t. origin is

$$M_{x+y}(t) = \Sigma\, p_i\, \{e^{t\,(x+y)}\} = \Sigma\, p_i\, \{e^{tx}\,.\,e^{ty}\} = \Sigma\, p_i\, (e^{tx}).\Sigma\, p_i\, (e^{ty})$$

Since x and y are independent variables and so are e^{tx} and e^{ty},

Hence, $M_{x+y}(t) = M_x(t)\,.\,M_y(t)$ **Proved.**

Example 37. *Find the moment generating function of the exponential distribution.*

$$f(x) = \frac{1}{c} e^{-x/c},\ 0 \le x \le \infty,\ c > 0$$ (*U.P. III Semester, Dec. 2005*)

Hence, find its mean and standard deviation.

Solution. The moment generating function about the origin is

$$M_0(t) = \int_0^{\infty} e^{tx}\,.\,f(x)\,dx = \int_0^{\infty} e^{tx}\,.\,\frac{1}{c} e^{-\frac{x}{c}}\,dx = \frac{1}{c}\int_0^{\infty} e^{\left(t-\frac{1}{c}\right)x}\,dx$$

$$= \frac{1}{c}\,\frac{1}{t-\frac{1}{c}}\left[e^{\left(t-\frac{1}{c}\right)x}\right]_0^{\infty}$$

$$= \frac{1}{c} \cdot \frac{1}{t - \frac{1}{c}} [0 - 1] = \frac{1}{ct - 1}(-1) = \frac{1}{1 - ct} = (1 - ct)^{-t}$$

$$= 1 + ct + c^2 t^2 + c^3 t^3 + \ldots\ldots \qquad \text{[Binomial Theorem]}$$

Moment about origin $= \left[\frac{d}{dt} M_0(t)\right]_{t=0} = \frac{d}{dt}[1 + ct + c^2t^2 + c^3t^3 + \ldots\ldots]_{t=0}$

$$= [c + 2c^2 t + 3c^3 t^2 + \ldots\ldots]_{t=0} = c$$

$$\mu_1 = \bar{x} = c$$

$$\mu'_2 = \left[\frac{d^2}{dt^2} M_0(t)\right] = \frac{d^2}{dt^2}[1 + ct + c^2t^2 + c^3t^3 + \ldots\ldots]_{t=0}$$

$$= \frac{d}{dt}[c + 2c^2 t + 3c^3t^2 + \ldots\ldots]_{t=0} = [2c^2 + 6c^3 t + \ldots]_{t=0}$$

$$\mu'_2 = 2c^2$$

$$\mu_2 = \mu'_2 - \mu_1^2 = 2c^2 - c^2 = c^2$$

Standard deviation $= \sqrt{\mu_2} = \sqrt{c^2} = c$ **Ans.**

Example 38. *Obtain the moment generating function of the random variable x having probability distribution*

$$f(x) = \begin{cases} x & \text{for } 0 < x < 1 \\ 2 - x & \text{for } 1 \le x < 2 \\ 0 & \text{elsewhere} \end{cases}$$ *Also determine* μ_1', μ_2' *and* μ_2.

Solution. $M_x(t) = \int e^{tx} f(x)\, dx$

$$= \int_0^1 x \cdot e^{tx} dx + \int_1^2 (2 - x) e^{tx} dx + \int_2^\infty 0 . e^{tx} dx = \left(\frac{xe^{tx}}{t} - \frac{e^{tx}}{t^2}\right)_0^1 + \left(\frac{2e^{tx}}{t} - \frac{xe^{tx}}{t} + \frac{e^{tx}}{t^2}\right)_1^2$$

$$= \frac{e^t}{t} - \frac{e^t}{t^2} + \frac{1}{t^2} + \left[\left(\frac{2e^{2t}}{t} - \frac{2e^{2t}}{t} + \frac{e^{2t}}{t^2}\right) - \left(\frac{2e^t}{t} - \frac{e^t}{t} + \frac{e^t}{t^2}\right)\right] = \frac{e^{2t} - 2e^t + 1}{t^2}$$

$$= \left(\frac{e^t - 1}{t}\right)^2 = \frac{1}{t^2}(e^{2t} - 2e^t + 1) = \frac{1}{t^2}\left[1 + \frac{2t}{1} + \frac{4t^2}{2} + \frac{8t^3}{6} + \frac{16t^4}{24} + \ldots - 2\left(1 + t + \frac{t^2}{2} + \frac{t^3}{6} + \frac{t^4}{24} + \ldots\right) + 1\right]$$

$$= \frac{1}{t^2} + \frac{2}{t} + 2 + \frac{4t}{3} + \frac{2t^2}{3} + \ldots\ldots - \frac{2}{t^2} - \frac{2}{t} - 1 - \frac{t}{3} - \frac{t^2}{12} + \ldots. + 1$$

μ_1' = coefficient of $\frac{t}{1} = \frac{4}{3} - \frac{1}{3} = \frac{3}{3} = 1$

μ_2' = coefficient of $\frac{t^2}{2!} = \frac{4}{3} - \frac{1}{6} = \frac{8-1}{6} = \frac{7}{6}$ $\qquad \mu_2 = \mu_2' - (\mu_1')^2 = \frac{7}{8} - (1)^2 = \frac{1}{6}$

Hence $\mu_1' = 1$, $\mu_2' = \frac{7}{6}$, $\mu_2 = \frac{1}{6}$ **Ans.**

Example 39. *Find the moment generating function of the continuous normal distribution given by*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; -\infty < x < \infty$$

Solution. Here, we have

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Moment generating function about the origin

$$M_0(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \qquad \ldots (1)$$

Putting $\frac{x-\mu}{\sigma} = z$ so that $dx = \sigma\, dz$ in (1), we get

$$M_0(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(\sigma z+\mu)} e^{-\frac{z^2}{2}} (\sigma\, dz) = \frac{e^{\mu t}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t\sigma z - \frac{z^2}{2}} dz$$

$$= \frac{e^{\mu t + \frac{1}{2}t^2\sigma^2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(e^{t\sigma z - \frac{1}{2}t^2\sigma^2 - \frac{z^2}{2}} \right) dz = \frac{1}{\sqrt{2\pi}} e^{\mu t + \frac{1}{2}t^2\sigma^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2 - 2t\sigma z + t^2\sigma^2)} dz$$

$$= \frac{1}{\sqrt{2\pi}} e^{\mu t + \frac{1}{2}t^2\sigma^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t\sigma)^2} dz = e^{\mu t + \frac{1}{2}t^2\sigma^2} (1) \qquad \left[\int_0^{\infty} e^{-z^2} dz = \frac{\sqrt{\pi}}{2} \right]$$

$$= e^{\mu t + \frac{1}{2}t^2\sigma^2}$$ **Ans.**

Example 40. *The random variable X assuming only non-negative values has a gamma probability distribution if its probability distribution is given by*

$$f(x) = \begin{cases} \frac{\alpha^\beta}{\Gamma\beta} x^{\beta-1} e^{-\alpha x}; & x > 0,\ \alpha > 0,\ \beta > 1 \\ 0, & elsewhere \end{cases}$$

Find the moment generating function of Gamma probability distribution.

Solution.
$$M_x(t) = \int_0^{\infty} e^{tx} f(x)\, dx = \int_0^{\infty} e^{tx} \cdot \frac{\alpha^\beta}{\Gamma\beta} \cdot x^{\beta-1} e^{\alpha} dx = \frac{\alpha^\beta}{\Gamma\beta} \int_0^{\infty} x^{\beta-1} e^{-x(\alpha-t)} dx$$

$$= \frac{\alpha^\beta}{(\alpha-t)^\beta \Gamma\beta} \int_0^{\infty} y^{\beta-1} e^{-y} dy \quad \text{| where } y = x(\alpha - t) \text{ so that } dy = (\alpha - t)\, dx$$

$$= \frac{1}{\left(1-\frac{t}{\alpha}\right)^\beta} \cdot \frac{1}{\Gamma\beta} \Gamma\beta = \left(1-\frac{t}{\alpha}\right)^{-\beta}; |t| < \alpha.$$ **Ans.**

Note: $M_x(t) = \begin{cases} \sum_x e^{tx} p(x), \text{ if } x \text{ is discrete} \\ \int_{-\infty}^{\infty} f_x(x)\, dx, \text{ if } x \text{ is continuous} \end{cases}$

EXERCISE 57.5

Find the moment generating function of the following functions:

1. $f(x) = me^{-mx}; x, m > 0$ **Ans.** $\sum_{r=0}^{\infty} \left(\frac{t}{m}\right)^r,\ \mu'_r = \frac{r!}{m^r}$

2. $f(x) = e^{-x}(1 + e^{-x})^{-2}, -\infty < x < \infty$ **Ans.** $\beta(1-t,\ 1+t),\ 1-t > 0$
$\pi t \operatorname{cosec} \pi t,\ t < 1$

3. If x is a discrete random variable with probability function $f(x) = \frac{1}{k^x},\ x = 1, 2 \ldots\ldots, (k = \text{constant})$

Find its moment generating function. **Ans.** $\dfrac{e^t}{k - e^t}$

4. Find the moment generating function for the distribution where,

$$f(x) = \begin{cases} \dfrac{2}{3}, & \text{at } x = 1 \\ \dfrac{1}{3}, & \text{at } x = 2 \\ 0, & \text{otherwise} \end{cases}$$

Ans. $\dfrac{2}{3} \cdot e^t + \dfrac{1}{3} e^{2t}$

5. Find the moment generating function of the random variable x having the probability density function:

$$f(x) = \begin{cases} \dfrac{1}{k}, & \text{for } -1 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

Ans. $\dfrac{1}{kt}\left(e^{2t} - e^{-t}\right), \; t \neq 0$; $\dfrac{3}{k}$, for $t = 0$

6. A random variable x has density function given by

$$f(x) = \begin{cases} 2e^{-2x}, & x \geq 0 \\ 0 \quad , & x < 0 \end{cases}$$

Obtain the moment generating function. **Ans.** $\dfrac{2}{2-t}$, if $t < 2$

7. Find the moment generating function, if it exists, given the probability distribution frequency :

$$f(x) = \begin{cases} xe^{-x}, & x > 0 \\ 0 \quad , & \text{otherwise} \end{cases}$$

Ans. $\dfrac{1}{1-t}$

8. Find the moment generating function for the given distribution

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{for } a < x < b \\ 0 \quad , & \text{otherwise} \end{cases}$$

Ans. $\dfrac{e^{bt} - e^{at}}{t(b-a)}$

9. A random variable x has the probability distribution function

$$f(x) = \begin{cases} \dfrac{1}{2x}, & \text{for } x = 1, 2, 3........ \\ 0, & \text{otherwise} \end{cases}$$

Ans. $\dfrac{e^t}{2 - e^t}$

10. The first four moments of a distribution about the value '0' are –0.20, 1.76, –2.36 and 10.88. Find the moments about the mean and measure the kurtosis. (U.P., III Semester, Dec. 2009)

Hint: In the usual notation, we are given assumed mean $a = 0$ and

$$\mu_1' = -0.20, \quad \mu_2' = 1.76, \quad \mu_3' = -2.36, \quad \mu_4' = 10.88$$

Moment about mean

$$\mu_1 = \mu_1' = -0.20$$

$$\mu_2 = \mu_2' - \mu_1'^2 = 1.76 - (0.20)^2 = 1.76 - 0.04$$

$$\Rightarrow \quad \mu_2 = 1.72$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 22\mu_1'^3$$

$$= -2.36 - 3(1.76)(-0.20) + 2(-0.20)^3 = -2.36 + 1.056 - 0.016 = -1.32$$

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

$$= 10.88 - 4(-2.36)(-0.20) + 6(1.76)(-0.20)^2 - 3(-0.20)^4$$

$$= 10.88 - 1.888 + 0.4224 - 0.0048$$

$$= 9.4096$$

$$\text{Kurtosis} = \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{9.4096}{(1.72)^2} = \frac{9.4096}{2.9584} = 3.180638188$$

Here, $\beta_2 > 3$, so the curve is peaked or Leptokurtic. **Ans.**

CHAPTER

58

METHOD OF LEAST SQUARES

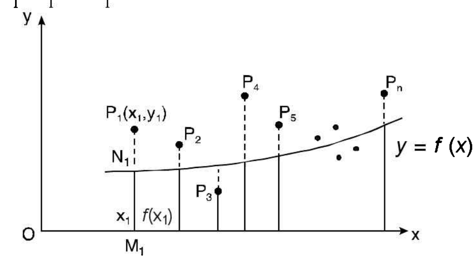
58.1 PRINCIPLE OF LEAST SQUARE

The method of least squares is probably the most systematic procedure to fit a unique curve through the given points.

Let $y = f(x)$ be the equation of curve to be fitted to the given data (observed or experimental) points (x_1, y_1), (x_2, y_2).... (x_n, y_n) . At $x = x_i$, the observed (or experimental) value of the ordinate is y_i and the corresponding value on the fitting curve is $N_1 M_1$, *i.e.*, $[f(x_1)]$. The difference of the observed and the expected (theoretical) value is

$$= P_1 M_1 - N_1 M_1 = P_1 N_1 = e_1.$$

This difference is called the error.



$$e_1 = y_1 - f(x_1)$$

Similarly, $$e_2 = y_2 - f(x_2)$$

$$e_3 = y_3 - f(x_3)$$

....

$$e_n = y_n - f(x_n)$$

Some of the errors $e_1, e_2, e_3, ... e_n$ will be positive and others negative.

In finding the total errors, errors are added. In addition, some negative and some positive errors may cancel and in some cases sum of all the errors may be zero, which leads to false result. To avoid such situation, we may make all the errors positive by squaring.

$$\text{Sum} = S = e_1^{\,2} + e_2^{\,2} + e_3^{\,2} + ... + e_n^{\,2}$$

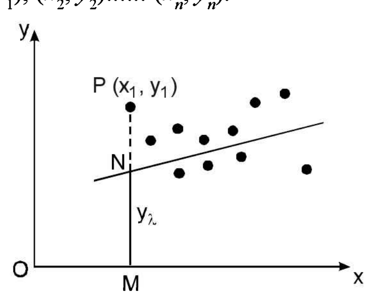
The curve of the best fit is that for which the sum of the squares of errors (S) is minimum. This is called the principle of least squares.

58.2 METHOD OF LEAST SQUARES

Let $$y = a + bx \qquad ...(1)$$

be the straight line to be fitted to the given data points (x_1, y_1), (x_2, y_2)...... (x_n, y_n).

Let y_{t_1} be the theoretical ordinate for x_1.



$$PM = y_1 \qquad NM = y_{t_1}$$

$$PN = PM - NM$$

Then $$e_1 = y_1 - y_{t_1} \qquad (PN = e_1)$$

$$e_1 = y_1 - (a + bx_1) \qquad (y_{t_1} = a + bx_1)$$

On squaring, we get $$e_1^2 = (y_1 - a - bx_1)^2$$

$$S = e_1^2 + e_2^2 + e_3^2 + \ldots\ldots e_n^2 = \Sigma e^2{}_i$$

$$S = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

For S to be minimum

$$\frac{\partial S}{\partial a} = \sum_{i=1}^{n} 2(y_i - a - bx_i)(-1) = 0 \text{ or } \Sigma (y - a - bx) = 0 \qquad \ldots (2)$$

[To generalise y_i, y_i is written as y]

$$\frac{\partial S}{\partial b} = \sum_{i=1}^{n} 2\ (y_i - a - bx_i)(-x_i) \text{ or } \Sigma (xy - ax - bx^2) = 0 \qquad \ldots (3)$$

On simplification equations (2) and (3) become

$$\Sigma y = na + b\,\Sigma x \qquad \ldots (4)$$

$$\Sigma xy = a\,\Sigma x + b\,\Sigma x^2 \qquad \ldots (5)$$

The equations (4) and (5) are known as Normal equations.

On solving equations (4) and (5), we get the values of a and b.

On putting the values of a and b in (1), we get the equation of required line.

To Remember : The normal equations (4) and (5) are for

$$y = a + bx$$

(i) Equation (4) is obtained by putting Σ before all the terms on both sides of (1).

i.e., $\Sigma y = \Sigma a + \Sigma bx \Rightarrow \Sigma y = na + b\,\Sigma x$

(ii) Equation (5) is obtained on multiplying equation (1) by x and putting Σ before each obtained term on both the sides.

i.e.,

$$\Sigma xy = \Sigma ax + \Sigma bx^2$$
$$\Sigma xy = a\,\Sigma x + b\,\Sigma x^2$$

Example 1. *Find the best values of a and b so that y = a + bx fits the data given in the table.*

x	0	1	2	3	4
y	1.0	2.9	4.8	6.7	8.6

Solution. $y = a + bx \qquad \ldots (1)$

x	y	xy	x^2
0	1.0	0	0
1	2.9	2.9	1
2	4.8	9.6	4
3	6.7	20.1	9
4	8.6	34.4	16
$\Sigma x = 10$	$\Sigma y = 24.0$	$\Sigma xy = 67.0$	$\Sigma x^2 = 30$

Normal equations are

$$\Sigma y = na + b\,\Sigma x \qquad \ldots (2)$$

$$\Sigma xy = a\,\Sigma x + b\,\Sigma x^2 \qquad \ldots (3)$$

On putting the values of Σx, Σy, Σxy, Σx^2 in (2) and (3), we have

$$24 = 5a + 10\,b \qquad \ldots (4)$$

$$67 = 10a + 30b \qquad \ldots (5)$$

On solving (4) and (5), we get

$$a = 1, \quad b = 1.9$$

On substituting the values of a and b in (1), we get

$$y = 1 + 1.9\,x$$ **Ans.**

Example 2. *By the method of least squares, find the straight line that best fits the following data :*

x	1	2	3	4	5
y	14	27	40	55	68

Solution. Let the equation of the straight line best fit be $y = a + bx$ ... (1)

x	y	xy	x^2
1	14	14	1
2	27	54	4
3	40	120	9
4	55	220	16
5	68	340	25
$\Sigma x = 15$	$\Sigma y = 204$	$\Sigma xy = 748$	$\Sigma x^2 = 55$

Here, $n = 5$

Normal equations are $\Sigma y = na + b\,\Sigma x$ ... (2)

$$\Sigma xy = a\,\Sigma x + b\,\Sigma x^2 \quad \text{... (3)}$$

On putting the values of Σx, Σy, Σxy and Σx^2 in (2) and (3), we have

$$204 = 5a + 15\,b \quad \text{... (4)}$$

$$748 = 15a + 55b \quad \text{... (5)}$$

On solving equations (4) and (5), we get

$$a = 0, \quad b = 13.6$$

On substituting the values of a and b in (1), we get

$$y = 13.6\,x$$ **Ans..**

Example 3. *Use least-squares method to fit a curve of the form $y = ae^{bx}$ to the data :*

x	1	2	3	4	5	6
y	7.209	5.265	3.846	2.809	2.052	1.499

Solution. $y = ae^{bx}$ ... (1)

On taking log of both sides, we get

$$\log_e y = \log_e a + bx \quad \text{... (2)}$$

On putting $\log_e y = Y$, $\log_e a = c$ in (2), we get

$$Y = c + bx \quad \text{... (3)}$$

x	y	$Y = \log_e y$	xY	x^2
1	7.209	1.97533	1.97533	1
2	5.265	1.66108	3.32216	4
3	3.846	1.34703	4.04109	9
4	2.809	1.03283	4.13132	16
5	2.052	0.71881	3.59405	25
6	1.499	0.40480	2.4288	36
$\Sigma x = 21$		$\Sigma Y = 7.13988$	$\Sigma xY = 19.49275$	$\Sigma x^2 = 91$

Normal equations are $\Sigma Y = nc + b\,\Sigma x$ ... (4)

$$\Sigma xY = c\,\Sigma x + b\,\Sigma x^2 \quad \text{... (5)}$$

On putting the values of n, Σx, ΣY, ΣxY and Σx^2 in equations (4) and (5), we get

$$7.13988 = 6c + 21b \quad \text{... (6)}$$

$$19.49275 = 21c + 91b \quad \text{... (7)}$$

On solving (6) and (7), we obtain $b = -0.3141$, $c = 2.28933$

$$c = \log_e a = 2.28933 \quad \Rightarrow \quad a = 9.86832$$

On substituting the values of a and b in (1), we get

$$y = 9.86832\, e^{-0.3141x}$$ **Ans.**

58.3 CHANGE OF ORIGIN AND SCALE

In some problems the magnitude of the variables in the given data is so large that the calculation becomes very tedious. The size of the data can be reduced by assuming some origin for x, y series.

The problem is further simplified by taking suitable scale for the values of x and y. If these values are equally spaced.

Let z be the width of the interval and (x_0, y_0) be taken as origin. Then putting

$$u = \frac{x - x_0}{h} \text{ and } v = y - y_0$$

Example 4. *Show that the line of fit to the following data is given by $y = 0.7x + 11.285$:*

x	0	5	10	15	20	25
y	12	15	17	22	24	30

Solution. Let $x_0 = 12.5$, $h = 2.5$, $y_0 = 20$ $\left[x_0 = \frac{\Sigma x}{N} \quad y_0 = \frac{\Sigma y}{N} \right]$

$$u = \frac{x - 12.5}{2.5} \text{ and } v = y - 20$$

$\Rightarrow$ The transformed equation is $v = a + bu$ ... (1)

x	y	$u = \frac{x-12.5}{2.5}$	$v = y - 20$	uv	u^2
0	12	– 5	– 8	40	25
5	15	– 3	– 5	15	9
10	17	– 1	– 3	3	1
15	22	1	2	2	1
20	24	3	4	12	9
25	30	5	10	50	25
		$\Sigma u = 0$	$\Sigma v = 0$	$\Sigma uv = 122$	$\Sigma u^2 = 70$

Normal equations are $\Sigma v = na + b\,\Sigma u$ ... (2)

$$\Sigma uv = a\,\Sigma u + b\,\Sigma u^2 \quad \text{... (3)}$$

On putting the values of Σu, Σv, Σuv, Σu^2 in (2) and (3), we get

$$0 = 6a + 0 \quad \Rightarrow \quad a = 0$$

$$122 = a \times 0 + b \times 70 \quad \Rightarrow \quad b = \frac{122}{70} = 1.743$$

Putting the values of a and b in (1), we get

$$v = 1.743\,u \quad \text{... (4)}$$

Putting $u = \dfrac{x-12.5}{2.5}$ and $v = y - 20$ in (4), we get

$$y - 20 = 1.743\left(\frac{x-12.5}{2.5}\right)$$

$$\Rightarrow \quad 2.5y - 50 = 1.743x - 1.743 \times 12.5$$

$$\Rightarrow \quad 2.5y = 1.743x - 21.7875 + 50$$

$$\Rightarrow \quad 2.5y = 1.743x + 28.2125$$

$$\Rightarrow \quad y = 0.7x + 11.285$$

Ans.

Example 5. *Fit a straight line to the following data :*

x	71	68	73	69	67	65	66	67
y	69	72	70	70	68	67	68	64

Solution. $$y = a + bx \quad \text{... (1)}$$

$$u = x - 69 \text{ and } v = y - 68$$

Transformed Equation is $$v = a + bu \quad \text{... (2)}$$

x	y	$u = x - 69$	$v = y - 68$	uv	u^2
71	69	2	1	2	4
68	72	– 1	4	– 4	1
73	70	4	2	8	16
69	70	0	2	0	0
67	68	– 2	0	0	4
65	67	– 4	– 1	4	16
66	68	– 3	0	0	9
67	64	– 2	– 4	8	4
		$\Sigma u = -6$	$\Sigma v = 4$	$\Sigma uv = 18$	$\Sigma u^2 = 54$

Normal equations are $$\Sigma v = na + b\,\Sigma u \quad \text{... (3)}$$

$$\Sigma uv = a\,\Sigma u + b\,\Sigma u^2 \quad \text{... (4)}$$

On putting the values of Σu, Σv, Σuv, Σu^2 in (3) and (4), we get

$$4 = 8a + b(-6) \quad \text{... (5)}$$

$$18 = -6a + 54b \quad \text{... (6)}$$

On solving (5) and (6), we get

$$a = \frac{9}{11}, \quad b = \frac{14}{33}$$

On putting the values of a and b in (2), we get

$$v = \frac{9}{11} + \frac{14}{33}u \quad \text{... (7)}$$

On putting $u = x - 69$ and $v = y - 68$ in (7), we get

$$y - 68 = \frac{9}{11} + \frac{14}{33}(x-69)$$

$$y - 68 = 0.8182 + 0.4242\,x - 29.2727$$

$$y = 68.8182 - 29.2727 + 0.4242\,x$$

$$y = 39.5455 + 0.4242\,x$$

Ans.

EXERCISE 58.1

1. Find the linear least square polynomials based on data

x	-2	-1	0	1
y	6	3	2	2

is given. Find the least square straight line approximation to the data. **Ans.** $10y = 26 - 13x$

2. The following table shows the number of salesmen working for a certain concern :

Year	1998	1999	2000	2001	2002	2003
Number	28	38	46	40	56	60

Use the method of least squares to fit a straight line trend. **Ans.** $y = 5.9428x - 11843.9047$

3. If F is pull required to lift a load W by means of a pulley, fit a linear law $F = a - bw$ connecting F and W against the following data:

W	50	70	100	120
F	12	15	21	25

Ans. $F = 2.2785 - 0.1879\text{W}$

4. Determine the constants a and b by the least-squares method such that $y = ae^{bx}$ fits the following data :

x	1.0	1.2	1.4	1.6
y	40.170	73.196	133.372	243.02

Ans. $a = 2,\ b = 3$

5. Fit a least-square geometric curve $y = ax^b$ to the following data :

x	1	2	3	4	5
y	0.5	2	4.5	8	12.5

Ans. $a = 0.5012,\ b = 1.9977$

6. The pressure and volume of a gas are related by the equation $PV^r = k$, r and k being constants. Fit this equation to the following set of observations :

P (kg/cm^2)	0.5	1.0	1.5	2.0	2.5	3.0
V (litres)	1.62	1.00	0.75	0.62	0.52	0.46

Ans. $PV^{1.276} = 1.039$

7. The pressure of the gas corresponding to various volumes V is measured, given by the following data :

V (cm^3)	50	60	70	90	100
P ($kg\ cm^{-2}$)	64.7	51.3	40.5	25.9	78

Fit the data to the equation $PV^r = C$ **Ans.** $PV^{0.28997} = 167.78765$

58.4 TO FIT UP THE PARABOLA

Let $$y = a + bx + cx^2 \quad \text{... (1)}$$

be the equation of a parabola.

The following normal equations are obtained as in Art. 58.2

The normal equations are
$$\Sigma y = na + b\,\Sigma x + c\,\Sigma x^2 \quad \text{... (2)}$$
$$\Sigma xy = a\,\Sigma x + b\,\Sigma x^2 + c\,\Sigma x^3 \quad \text{... (3)}$$
$$\Sigma x^2 y = a\,\Sigma x^2 + b\,\Sigma x^3 + c\,\Sigma x^4 \quad \text{... (4)}$$

On solving these three normal equations, we get the values of a, b and c.

On putting the values of a, b and c in (1), we get the required equation of parabola.

To remember the normal equations (2), (3) and (4) for $y = a + bx + cx^2$.

(*i*) Equation (2) is obtained by putting Σ before each term on both sides of (1).

(*ii*) Equation (3) is obtained on multiplying (1) by x and putting Σ before each term on both sides of obtained equation.

(*iii*) Equation (4) is obtained on multiplying (1) by x^2 and putting Σ before each term on both sides of obtained equation.

Example 6. *Find least squares polynomial approximation of degree two to the data:*

x	0	1	2	3	4
y	– 4	– 1	4	11	20

Also compute the least error.

Solution. Let the equation of the polynomial be

$$y = a + bx + cx^2 \qquad \text{... (1)}$$

x	y	xy	x^2	x^2y	x^3	x^4
0	– 4	0	0	0	0	0
1	–1	–1	1	–1	1	1
2	4	8	4	16	8	16
3	11	33	9	99	27	81
4	20	80	16	320	64	256
$\Sigma x = 10$	$\Sigma y = 30$	$\Sigma xy = 120$	$\Sigma x^2 = 30$	$\Sigma x^2y = 434$	$\Sigma x^3 = 100$	$\Sigma x^4 = 354$

Normal equations are $\Sigma y = na + b\,\Sigma x + c\,\Sigma x^2$ ... (2)

$$\Sigma xy = a\,\Sigma x + b\,\Sigma x^2 + c\,\Sigma x^3 \qquad \text{... (3)}$$

$$\Sigma x^2 y = a\,\Sigma x^2 + b\,\Sigma x^3 + c\,\Sigma x^4 \qquad \text{... (4)}$$

On putting the values of $\Sigma x, \Sigma y, \Sigma xy, \Sigma x^2, \Sigma x^2y, \Sigma x^3, \Sigma x^4$ in equations (2), (3), (4), we obtain

$$30 = 5a + 10b + 30c \qquad \text{... (5)}$$

$$120 = 10a + 30b + 100c \qquad \text{... (6)}$$

$$434 = 30a + 100b + 354c \qquad \text{... (7)}$$

On solving these equations, we get $a = -4$, $b = 2$, $c = 1$.

The required polynomial is

$$y = -4 + 2x + x^2, \quad \text{Error} = 0$$ **Ans.**

Example 7. *Employ the method of least squares to fit a parabola $y = a + bx + cx^2$ in the following data :*

(x, y) : (–1, 2), (0, 0), (0, 1), (1, 2)

Solution. Let the equation of the parabola be

$$y = a + bx + cx^2 \qquad \text{... (1)}$$

Here, $n = 4$

x	y	x^2	x^3	x^4	xy	x^2y
–1	2	1	– 1	1	– 2	2
0	0	0	0	0	0	0
0	1	0	0	0	0	0
1	2	1	1	1	2	2
$\Sigma x = 0$	$\Sigma y = 5$	$\Sigma x^2 = 2$	$\Sigma x^3 = 0$	$\Sigma x^4 = 2$	$\Sigma xy = 0$	$\Sigma x^2y = 4$

Normal equations are

$$\Sigma y = na + b\,\Sigma x + c\,\Sigma x^2 \qquad \text{... (2)}$$

$$\Sigma xy = a\,\Sigma x + b\,\Sigma x^2 + c\,\Sigma x^3 \qquad \text{... (3)}$$

$$\Sigma x^2y = a\,\Sigma x^2 + b\,\Sigma x^3 + c\,\Sigma x^4 \qquad \text{... (4)}$$

On putting the values of $\Sigma y, \Sigma xy, \Sigma x^2 y$ etc. in (2), (3) and (4), we get

$$5 = 4a + 0\,b + 2\,c \quad \Rightarrow \quad 5 = 4a + 2c \qquad \text{... (5)}$$

$$0 = 0a + 2\,b + 0\,c \quad \Rightarrow \quad 0 = 2b \qquad \text{... (6)}$$

$$4 = 2a + 0\,b + 2\,c \quad \Rightarrow \quad 4 = 2a + 2c \qquad \text{... (7)}$$

On solving (5), (6) and (7), we get

$$a = 0.5, \quad b = 0, \quad c = 1.5$$

On putting these values in (1), we get

$$y = 0.5 + 1.5x^2$$ **Ans.**

Example 8. *Fit a parabola $y = a\,x^2 + b\,x + c$ to the following data taking x as independent variable.*

x	1	2	3	5	7	11	13	17	19	23
y	2	3	5	7	11	13	17	19	23	29

(*U.P. III Semester, 2009-2010*)

Solution. Here, we have

$$y = ax^2 + bx + c \qquad \text{... (1)}$$

x	y	$x\,y$	x^2	$x^2\,y$	x^3	x^4
1	2	2	1	2	1	1
2	3	6	4	12	8	16
3	5	15	9	45	27	81
5	7	35	25	175	125	625
7	11	77	49	539	343	2401
11	13	143	121	1573	1331	14641
13	17	221	169	2873	2197	28561
17	19	323	289	5491	4913	83521
19	23	437	361	8303	6859	130321
23	29	667	529	15341	12167	279841
$\Sigma x = 101$	$\Sigma y = 129$	$\Sigma xy = 1926$	$\Sigma x^2 = 1557$	$\Sigma x^2y = 34354$	$\Sigma x^3 = 27971$	$\Sigma x^4 = 540009$

Normal equations are

$$\Sigma y = na + b\,\Sigma x + c\,\Sigma x^2 \qquad \text{... (2)}$$

$$\Sigma xy = a\,\Sigma x + b\,\Sigma x^2 + c\,\Sigma x^3 \qquad \text{... (3)}$$

$$\Sigma x^2y = a\,\Sigma x^2 + b\,\Sigma x^3 + c\,\Sigma x^4 \qquad \text{... (4)}$$

On putting the values of $\Sigma\, x$, $\Sigma\, y$, $\Sigma\, xy$, $\Sigma\, x^2$, $\Sigma\, x^2y$, $\Sigma\, x^3$, $\Sigma\, x^4$, in equations (2), (3), (4), we get

$$129 = 10a + 101\, b + 1557c \qquad \text{... (5)}$$
$$1926 = 101a + 1557\, b + 27971c \qquad \text{... (6)}$$
$$34354 = 1557a + 27971b + 540009c \qquad \text{... (7)}$$

On solving (5), (6), (7), we get

$$a = 1.41259297$$
$$b = 1.089013957$$
$$c = 0.003136583595$$

Hence, the equation of the required parabola is

$$y = 1.41259297\, x^2 + 1.089013957x + 0.003136583595 \quad \textbf{Ans.}$$

Example 9. *Fit a second degree parabola to the following :*

x	*1*	*2*	*3*	*4*	*5*
y	*1090*	*1220*	*1390*	*1625*	*1915*

(*R.G.P.V., Bhopal, III Semester, Dec. 2003*)

Solution. Let the equation of the parabola be

$$y = a + bx + cx^2 \qquad \text{... (1)}$$

x	y	xy	x^2	x^2y	x^3	x^4
1	1090	1090	1	1090	1	1
2	1220	2440	4	4880	8	16
3	1390	4170	9	12510	27	81
4	1625	6500	16	26000	64	256
5	1915	9575	25	47875	125	625
$\Sigma x = 15$	$\Sigma y = 7240$	$\Sigma xy = 23775$	$\Sigma x^2 = 55$	$\Sigma x^2y = 92355$	$\Sigma x^3 = 225$	$\Sigma x^4 = 979$

Normal equations are

$$\Sigma\, y = na + b\, \Sigma\, x + c\, \Sigma\, x^2 \qquad \text{... (2)}$$
$$\Sigma\, xy = a\, \Sigma\, x + b\, \Sigma\, x^2 + c\, \Sigma\, x^3 \qquad \text{... (3)}$$
$$\Sigma\, x^2y = a\, \Sigma\, x^2 + b\, \Sigma\, x^3 + c\, \Sigma\, x^4 \qquad \text{... (4)}$$

On putting the values of n, $\Sigma\, x$, $\Sigma\, x^2$, $\Sigma\, x^3$, $\Sigma\, x^4$, Σy, $\Sigma\, xy$, $\Sigma\, x^2y$, in (3), (4) and (5), we get

$$7240 = 5a + 15b + 55c \qquad \text{... (5)}$$
$$23775 = 15a + 55b + 225c \qquad \text{... (6)}$$
$$92355 = 55a + 225b + 979c \qquad \text{... (7)}$$

Steps for solution of (5), (6) and (7) are the following :-

3 (5), $\quad 21720 = 15a + 45b + 165c \qquad \text{... (8)}$

(6) – (8), $\quad 2055 = 10b + 60c \qquad \text{... (9)}$

11 (5), $\quad 79640 = 55a + 165b + 605c \qquad \text{... (10)}$

(7) – (10), $\quad 12715 = 60b + 374c \qquad \text{... (11)}$

6 (9), $\quad 12330 = 60b + 360c \qquad \text{... (12)}$

(11) – (12), $\quad 385 = 14c \quad \Rightarrow \quad c = \dfrac{55}{2}$

From (9), $\quad 2055 = 10b + 60\left(\dfrac{55}{2}\right) \quad \Rightarrow \quad b = \dfrac{81}{2}$

From (5), $7240 = 5a + 15\left(\frac{81}{2}\right) + 55\left(\frac{55}{2}\right) \quad \Rightarrow \quad a = 1024$

On putting the values of a, b, c in (1), we get

$$y = 1024 + \frac{81}{2}x + \frac{55}{2}x^2$$

The equation of the required parabola is

$$2y = 2048 + 81x + 55x^2$$ **Ans.**

Example 10. *Fit a second degree parabola to the following data :*

x	*10*	*15*	*20*	*25*	*30*	*35*	*40*
y	*11*	*13*	*16*	*20*	*27*	*34*	*41*

(*R.G.P.V., Bhopal, III Semester, June 2005*)

Solution. Let the equation of the parabola be

$$y = a + bx + cx^2 \quad \text{... (1)}$$

x	y	xy	x^2	x^2y	x^3	x^4
10	11	110	100	1100	1000	10000
15	13	195	225	2925	3375	50625
20	16	320	400	6400	8000	160000
25	20	500	625	12500	15625	390625
30	27	810	900	24300	27000	810000
35	34	1190	1225	41650	42875	1500625
40	41	1640	1600	65600	64000	2560000
$\Sigma x = 175$	$\Sigma y = 162$	$\Sigma xy = 4765$	$\Sigma x^2 = 5075$	$\Sigma x^2y = 154475$	$\Sigma x^3 = 161875$	$\Sigma x^4 = 5481875$

Normal equations are
$$\Sigma y = na + b\,\Sigma x + c\,\Sigma x^2 \quad \text{... (2)}$$
$$\Sigma xy = a\,\Sigma x + b\,\Sigma x^2 + c\,\Sigma x^3 \quad \text{... (3)}$$
$$\Sigma x^2y = a\,\Sigma x^2 + b\,\Sigma x^3 + c\,\Sigma x^4 \quad \text{... (4)}$$

On putting the values of $n, \Sigma x, \Sigma x^2, \Sigma x^3, \Sigma x^4, \Sigma y, \Sigma xy, \Sigma x^2y$, in the equation (2), (3) and (4), we get

$$162 = 7a + 175b + 5075c \quad \text{... (5)}$$
$$4765 = 175a + 5075b + 161875c \quad \text{... (6)}$$
$$154475 = 5075a + 161875b + 5481875c \quad \text{... (7)}$$

$25 \times (5)$, $4050 = 175a + 4375b + 126875c$ ... (8)

$(6) - (8)$, $715 = 700b + 35000c$ ... (9)

$29 \times (6)$, $138185 = 5075a + 147175b + 4694375c$ ... (10)

$(7) - (10)$, $16290 = 14700b + 787500c$ ... (11)

$21 \times (9)$, $15015 = 14700b + 735000c$ ... (12)

$(11) - (12)$, $1275 = 52500c \quad \Rightarrow \quad c = \frac{17}{700}$

From (9), $715 = 700b + 35000\left(\frac{17}{700}\right) \quad \Rightarrow \quad b = -\frac{135}{700}$

From (5), $162 = 7a + 175\left(\frac{-135}{700}\right) + 5075\left(\frac{17}{700}\right)$

$$7a = 162 + \frac{135}{4} - \frac{493}{4} \Rightarrow a = \frac{145}{14}$$

On putting the values of a, b and c in (1), we get

$$y = \frac{145}{14} - \frac{135x}{700} + \frac{17x^2}{700}$$

Hence, the required parabola is

$$700y = 7250 - 135x + 17x^2$$ **Ans.**

58.5 CHANGE OF SCALE IN SECOND DEGREE EQUATIONS

If the data is of equal interval in large numbers then we change the scale as

$$u = \frac{x - x_0}{h} \text{ and } v = y - y_0.$$

Example 11. *Fit a second degree parabola to the following data by least squares method :*

x	1929	1930	1931	1932	1933	1934	1935	1936	1937
y	352	356	357	358	360	361	361	360	359

(*U.P., II Semester, Summer 2001*)

Solution. Taking $x_0 = 1933$, $y_0 = 357$

Again taking $u = x - x_0$, $v = y - y_0$

$u = x - 1933$, $v = y - 357$

The equation $y = a + bx + cx^2$ is transformed to $v = A + Bu + Cu^2$

x	$u = x - 1933$	y	$v = y - 357$	uv	u^2	u^2v	u^3	u^4
1929	– 4	352	– 5	20	16	– 80	– 64	256
1930	– 3	356	– 1	3	9	– 9	– 27	81
1931	– 2	357	0	0	4	0	– 8	16
1932	– 1	358	1	– 1	1	1	– 1	1
1933	0	360	3	0	0	0	0	0
1934	1	361	4	4	1	4	1	1
1935	2	361	4	8	4	16	8	16
1936	3	360	3	9	9	27	27	81
1937	4	359	2	8	16	32	64	256
Total	$\Sigma u = 0$		$\Sigma v = 11$	$\Sigma uv = 51$	$\Sigma u^2 = 60$	$\Sigma u^2 v = -9$	$\Sigma u^3 = 0$	$\Sigma u^4 = 708$

Normal equations are

$$\Sigma v = nA + B\Sigma u + C\Sigma u^2 \Rightarrow 11 = 9A + 0B + 60C \Rightarrow 11 = 9A + 60C$$

$$\Sigma uv = A\Sigma u + B\Sigma u^2 + C\Sigma u^3 \Rightarrow 51 = 0A + 60B + 0C \Rightarrow 51 = 60B \Rightarrow B = \frac{17}{20}$$

$$\Sigma u^2 v = A\Sigma u^2 + B\Sigma u^3 + C\Sigma u^4 \Rightarrow -9 = 60A + 0B + 708C \Rightarrow -9 = 60A + 708C$$

On solving these equations, we get

$$A = \frac{694}{231}, B = \frac{17}{20}, C = -\frac{247}{924}$$

$$v = \frac{694}{231} + \frac{17}{20}u - \frac{247}{924}u^2$$

Putting $v = y - 357$ and $u = x - 1933$, we get

$$y - 357 = \frac{694}{231} + \frac{17}{20}(x-1933) - \frac{247}{924}(x-1933)^2$$

$$\Rightarrow \quad y - 357 = \frac{694}{231} + \frac{17x}{20} - \frac{32861}{20} - \frac{247x^2}{924} - \frac{247}{924}(-3866x) - \frac{247}{924} \times (1933)^2$$

$$\Rightarrow \quad y - 357 = \frac{694}{231} - \frac{32861}{20} - \frac{247}{924}(1933)^2 + \frac{17}{20}x + \frac{247 \times 3866}{924}x - \frac{247}{924}x^2$$

$$\Rightarrow \quad y = 3 - 1643.05 - 998823.36 + 357 + 0.85x + 1033.44x - 0.267x^2$$

$$y = -1000106.41 + 1034.29x - 0.267x^2$$ **Ans.**

Example 12. *Fit a parabolic curve of regression of y on x to the following data :*

x	1.0	1.5	2.0	2.5	3.0	3.5	4.0
y	1.1	1.3	1.6	2.0	2.7	3.4	4.1

Solution. Here, $n = 7$ (odd)

$$u = \frac{x-2.5}{0.5} = 2x - 5$$

Let the equation be $y = a + bx + cx^2$

The transformed equation is

$$y = a + bu + cu^2 \qquad \text{... (1)}$$

x	y	u	u^2	uy	u^2y	u^3	u^4
1.0	1.1	– 3	9	– 3.3	9.9	– 27	81
1.5	1.3	– 2	4	– 2.6	5.2	– 8	16
2.0	1.6	– 1	1	– 1.6	1.6	– 1	1
2.5	2.0	0	0	0	0	0	0
3.0	2.7	1	1	2.7	2.7	1	1
3.5	3.4	2	4	6.8	13.6	8	16
4.0	4.1	3	9	12.3	36.9	27	81
Total	$\Sigma y = 16.2$	$\Sigma u = 0$	$\Sigma u^2 = 28$	$\Sigma uy = 14.3$	$\Sigma u^2y = 69.9$	$\Sigma u^3 = 0$	$\Sigma u^4 = 196$

Normal equations are

$$\Sigma y = na + b\Sigma u + c\Sigma u^2 \qquad \text{... (2)}$$

$$\Sigma uy = a\Sigma u + b\Sigma u^2 + c\Sigma u^3 \qquad \text{... (3)}$$

$$\Sigma u^2y = a\Sigma u^2 + b\Sigma u^3 + c\Sigma u^4 \qquad \text{... (4)}$$

On putting the values of Σu, Σy, Σuy etc. in (2), (3) and (4), we get

$$16.2 = 7a + 0 \times b + 28c \qquad 16.2 = 7a + 28c$$

$$14.3 = 0 \times a + 28b + 0 \times c \quad \Rightarrow \quad 14.3 = 28b$$

$$69.9 = 28a + 0 \times b + 196c \quad \Rightarrow \quad 69.9 = 28a + 196c$$

On solving the above equations, we get

$$a = 2.07, \; b = 0.511, \; c = 0.061$$

On putting the values of a, b, c in (1), we get

$$y = 2.07 + 0.511\,u + 0.061\,u^2 \qquad \text{... (5)}$$

On putting the value of $u = 2x - 5$ in (5), we get

$$y = 2.07 + 0.511\,(2x - 5) + 0.061\,(2x - 5)^2$$

$$y = 2.07 + 1.022x - 2.555 + 0.061\,(4x^2 - 20x + 25)$$

$$\Rightarrow \quad y = 2.07 + 1.022x - 2.555 + 0.244x^2 - 1.22x + 1.525$$

$$\Rightarrow \quad y = 1.04 - 0.198\,x + 0.244\,x^2$$ **Ans.**

Example 13. *Fit a second degree parabola to the following data :*

x	1	2	3	4	5	6	7	8	9	10
y	124	129	140	159	228	289	315	302	263	210

(*U.P., III Semester, Dec. 2009*)

Solution. Taking $x_0 = 6$ and $y_0 = 228$

x	$u = x - 6$	y	$v = y - 228$	uv	u^2	$u^2 v$	u^3	u^4
1	– 5	124	– 104	520	25	– 2600	– 125	625
2	– 4	129	– 99	396	16	– 1584	– 64	256
3	– 3	140	– 88	264	9	– 792	– 27	81
4	– 2	159	– 69	138	4	– 276	– 8	16
5	– 1	228	0	0	1	0	– 1	1
6	0	289	61	0	0	0	0	0
7	1	315	87	87	1	87	1	1
8	2	302	74	148	4	296	8	16
9	3	263	35	105	9	315	27	81
10	4	210	– 18	– 72	16	– 288	64	256
Total	– 5		– 121	1586	85	– 4842	–125	1333

Let the equation of parabola be $a + bu + cu^2 = v$

Normal equations are

$$\Sigma v = na + b\Sigma u + c\Sigma u^2 \quad \Rightarrow \quad -121 = 10a - 5b + 85c \qquad \text{... (1)}$$

$$\Sigma uv = a\Sigma u + b\Sigma u^2 + c\Sigma u^3 \quad \Rightarrow \quad 1586 = -5a + 85b - 125c \qquad \text{... (2)}$$

$$\Sigma u^2 v = a\Sigma u^2 + b\Sigma u^3 + c\Sigma u^4 \quad \Rightarrow \quad -4842 = 85a - 125b + 1333c \qquad \text{... (3)}$$

$$(1) + 2\,(2) \quad \Rightarrow \quad 3051 = 165b - 165c \qquad \text{... (4)}$$

$$(3) + 17\,(2) \quad \Rightarrow \quad 22120 = 1320b - 792c \qquad \text{... (5)}$$

$$8\,(4) - (5) \quad \Rightarrow \quad 2288 = -528c$$

$$\Rightarrow \quad c = -\frac{13}{3}$$

Putting the value of c in (4), we get

$$3051 = 165b - 165\left(-\frac{13}{3}\right)$$

$$\Rightarrow \quad 3051 = 165b + 715$$

$$\Rightarrow \quad 165b = 2336 \qquad \Rightarrow \quad b = \frac{2336}{165}$$

Putting the values of b and c in (1), we get

$$-121 = 10a - 5\left(\frac{2336}{165}\right) + 85\left(-\frac{13}{3}\right)$$

$$\Rightarrow \quad -121 = 10a - \frac{2336}{33} - \frac{1105}{3}$$

$$\Rightarrow \quad 10a = \frac{10498}{33} \qquad \Rightarrow \quad a = \frac{5249}{165}$$

Putting the values of a, b and c in $a + bu + cu^2 = v$, we get

$$\frac{5249}{165} + \frac{2336}{165}u - \frac{13}{3}u^2 = v$$

Putting the values of $u = x - 6$ and $v = y - 228$, we get

$$\frac{5249}{165} + \frac{2336}{165}(x-6) - \frac{13}{3}(x-6)^2 = y - 228$$

$$\Rightarrow \quad \frac{5249}{165} + \frac{2336}{165}x - \frac{4672}{55} - \frac{13}{3}x^2 + 52x - 156 = y - 228$$

$$\Rightarrow \quad -\frac{13}{3}x^2 + \frac{10916}{165}x + \frac{283}{15} = y$$

$$\Rightarrow \quad y = -4.3333x^2 + 66.1576x + 18.8667 \quad \textbf{Ans.}$$

EXERCISE 58.2

1. Find the values of a, b, c so that $y = a + bx + cx^2$ is the best fit to the data :

x	0	1	2	3	4
y	1	0	3	10	21

Ans. $a = 1$, $b = -3$, $c = 2$

2. Fit a second degree parabola to the following data by Least squares method :

x	0	1	2	3	4
y	1	1.8	1.3	2.5	6.3

Ans. $y = 1.42 - 1.07x + 0.55x^2$

3. Fit a second degree parabola to the following data taking y as dependent variable :

x	1	2	3	4	5	6	7	8	9
y	2	6	7	8	10	11	11	10	9

Ans. $y = -1 + 3.55x - 0.27x^2$

4. Use the least-square method to obtain a parabola that approximates the data :

x	1.0	1.2	1.4	1.6	1.8	2
y	2.345	2.419	2.592	2.863	3.233	3.702

Ans. $y = 3.124 - 2.477x + 1.458x^2$