UNIT II

Descriptive Statistics

Measures of Central Tendency

Types of Data:

(i) Individual observations
(ii) Discrete series
(iii) Continuous series

Example: (i)45, 56,78,97,…….,90

(ii)X (height of the student)	f(no. of student)
155	7
156	4
157	2
158	5
159	1
160	1

	20

(iii)Marks X	No.of students
0-10	0
10-20	0

20-30	1
30-40	1
40-50	2
50-60	1
60-70	5
70-80	4
80-90	4
90-100	1
	20

Measures of Central Tendency:

Individual Observations

Mean = $\frac{\sum x}{n}$

Mean = $A + \frac{\sum d}{n}$ where d=x-A, A-Assumed mean(Individual Observations)

Median = $\frac{n+1}{2}$ th item(Individual Observations)

Mode = the item which is occurred more number of times.

Discrete Series:

Mean = $\frac{\sum fx}{N}$ where $N = \sum f$

Mean = $A + \frac{\sum fd}{N}$ where d=x-A, A-Assumed mean

Mean = $A + \frac{\sum fd}{N} \times i$ where $d = \frac{x - A}{i}$, A-Assumed mean , i= class interval

Median = $\frac{N+1}{2}$ th item

Mode = the item which is occurred more number of times.

Continuous Series:

Mean = $\frac{\sum fm}{N}$ where $N = \sum f$

Mean = $A + \frac{\sum fd}{N}$ where d=m-A, A-Assumed mean

Mean = $A + \frac{\sum fd}{N} \times i$ where $d = \frac{m - A}{i}$, A-Assumed mean , i= class interval

Median = $\frac{N}{2}$th item

Median = $L + \frac{\frac{N}{2} - cf}{f} \times i$

Where L- Lower limit of the median class

$N = \sum f$

Cf-cumulative frequency preceding the median class

f- frequency of the median class

i-class interval

Mode = $M_0 = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$

Where L – lower limit of the modal class

f_0 -frequency of the class preceding the modal class

f_1 - frequency of the modal class

f_2 - frequency of the class succeeding the modal class

i-class interval

Empirical relation:

Mean - Mode =3 (Mean – Median)

(OR)

Mode = Mean - 3 Mean + 3 Median

(OR)

Mode = 3 Median - 2 Mean

Problems :

1.Find the AM , median and mode of the following set of observations:
25,32,28,34,24,31,36; 27,29,30.

Mean= $\frac{\sum x}{n}$

=(25+32+28+34+24+31+36+ 27+29+30)/10=296/10

=29.6

24,25,27,28,29,30,31,32,34,36

Median =(n+1)/2th item

=(10+1)/2 th item=5.5th item

5.5th item =(5th item+6th item)/2=(29+30)/2=29.5

There is no mode.

2. Find the mode of following data: **2,3,2,1,3,2,3,3,2,1,3,3,3,2,2,1,1,3,3,3**

Mode =3 [since 3 come more number of times]

3. Find the mean, median and mode.

x	f
155	30
156	20
157	5
158	5
159	10
160	15
161	5
162	10

Solution:

x	f	fx	cf(cumulative frequency)
155(mode)	30	4650	30
156	20	3120	50
157(Median)	5	785	55
158	5	790	60
159	10	1590	70
160	15	2400	85
161	5	805	90
162	10	1620	100
	$N = \sum f$	$\sum fx$	

	=100	=15760	

N=100

Mean= $\frac{\sum fx}{N}$

=15760/100=157.60

$\overline{X} = 157.60$

Median = (N+1)/2 th item=(100+1)/2th item= 50.5th item

Median =157

Mode =155

4.Find the mean, median and mode for the following data:

Class(x)	frequency(f)
0-10	20
10-20	5
20-30	3
30-40	8
40-50	10
50-60	35
60-70	10
70-80	4
80-90	3
90-100	2
	$N = \sum f$ =100

Solution:

Class(x)	**m**	**frequency(f)**	**fm**	**cf**
0-10	5	20	100	20
10-20	15	5	75	25
20-30	25	3	75	28
30-40	35	8	280	36
40-50	45	10f0	450	46
50-60(median, Mode)	**55**	**35f1**	**1925**	**81**
60-70	65	10f2	650	91
70-80	75	4	300	95
80-90	85	3	255	98
90-100	95	2	190	100
		$N = \sum f$ **=100**	$\sum fm$ **=4300**	

N=100

Mean = $\frac{\sum fm}{N}$ where $N = \sum f$

=4300/100=43

Median =(N/2)th item = (100/2)th item =50^{th} item

50^{th} item is 50-60 (the median class)

Median = $L+\dfrac{\frac{N}{2}-cf}{f}\times i$

$= 50+\dfrac{50-46}{35}\times 10$

=50+1.1428=51.1428

Modal class is 50-60

$M_0 = L+\dfrac{f_1-f_0}{2f_1-f_0-f_2}\times i$

$= 50+\dfrac{35-10}{2(35)-10-10}\times 10$

=50+5=55

Another method:

Class(x)	m	frequency(f)	$d=\dfrac{m-55}{10}$	fd
0-10	5	20	-5	-100
10-20	15	5	-4	-20
20-30	25	3	-3	-9
30-40	35	8	-2	-16
40-50	45	10	-1	-10
50-60	55	35	0	0
60-70	65	10	1	10
70-80	75	4	2	8
80-90	85	3	3	9
90-100	95	2	4	8
		$N=\sum f$ =100		$\sum fd$ =-120

Mean = $A+\dfrac{\sum fd}{N}\times i$ where $d=\dfrac{m-A}{i}$, A-Assumed mean , i= class interval

A=55, i=10

Mean= $55+\dfrac{-120}{100}\times 10$

=55-12=43

MEASURES OF DISPERSION

There are five measures of dispersion:

Range,

Inter-quartile range or Quartile Deviation,

Mean deviation,

Standard Deviation,

and Lorenz curve.

Among them, the first four are mathematical methods and the last one is the graphical method.

Range:

Range = L-S

L- Largest values ; S- Smallest value

Coefficient of Range = $\frac{L-S}{L+S}$

Problems:

Example 1: Find the range for the following sets of data:

5, 15, 15, 05, 15,05,15,15,15,15

Range = 15-5=10

Coeff.of range= (L-S)/(L+S)

=10/20=1/2

2. Find the range for the following sets of data:

8 , 7, 15, 11, 12, 5, 13, 11 ,15, 9

Range = 15-5=10

Coeff.of range= (L-S)/(L+S)

=10/20=1/2

Example 2: Calculate the coefficient of range separately for the two sets of data given below:

Set 1	8	10	20	9	15	10	13	28
Set 2	30	35	42	50	32	49	39	33

Solution: It can be seen that the range in both the sets of data is the same:

Set 1 28-8=20

Set 2 50-30=20

Coefficient of range in Set 1 is:

$$\frac{28-8}{28+8} = 0.55$$

Coefficient of range in set 2 is:

$$\frac{50-30}{50+30} = 0.25$$

3.

3. Compute the range

X	**f**
158	15
159	20
160	32
161	35
162	33
163	22
164	20
165	10
166	8
	N=195

Range = 166-158=8

Coeff. Of Range = (166-158)/(166+158)=0.0246

4: Find the range for the following frequency distribution:

Size of Item	**Frequency**
20- 40	7
40- 60	11
60- 80	30
80-100	17
100-120	5
Total	**70**

Here, the upper limit of the highest class is 120

and the lower limit of the lowest class is 20.

Hence, the range is 120 - 20 = 100.

The coefficient of range is calculated by the formula: (L-S)/ (L+S)

Quartile Deviation :

$$QD = \frac{Q_3 - Q_1}{2}$$

Q_1- First quartile

Q_3-3rd quartile

Coefficient of quartile deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

To find Q1:

Q1 = $\frac{N+1}{4}$ **th item (Individual observations)**

Q1 = $\frac{N+1}{4}$ **th item (Discrete series)**

Q1 = $\frac{N}{4}$ **th item (continuous series)**

$$Q1 = L + \frac{\frac{N}{4} - cf}{f} \times i$$

To find Q3:

Q3 = $\frac{3(N+1)}{4}$ **th item (Individual observations)**

Q3= $\frac{3(N+1)}{4}$ **th item (Discrete series)**

Q3 = $\frac{3N}{4}$ **th item (continuous series)**

$$Q3 = L + \frac{\frac{3N}{4} - cf}{f} \times i$$

1. Find the Quartile deviation and the coefficient of QD 3, 8, 6, 10, 12, 9, 11, 10, 12, 7

Solution:

3, 6 , 7, 8,9,10,10,11,12,12

N=10

Q1= 2.75^{th} item

Q1= 2^{nd} item + 0.75(3^{rd} item – 2^{nd} item)

= 6+0.75(7-6)=6.75

Q3= 3(10+1)/4 th item = 8.25^{th} item

Q3=8^{th} item + 0.25(9^{th} item -8^{th} item)

= 11+0.25(12-11)=11.25

QD=(11.25-6.75)/2=2.25

QD= (Q3-Q1)/2

Coefficient of QD=0.25

2. **Compute the Quartile deviation and its relative measure.**

X	**f**
158	15
159	20
160	32
161	35
162	33
163	22
164	20
165	10
166	8
	N=195

Solution:

X	**f**	**cf**
158	15	15
159	20	35
160Q1	32	67
161	35	102
162	33	135
163Q3	22	157
164	20	177
165	10	187
166	8	195
	N=195	

Q1 = $\frac{N+1}{4}$ **th item (Discrete series)**

=(195+1)/4^{th} item=49 th item

Q1=160

Q3= $\frac{3(N+1)}{4}$ **th item (Discrete series)**

=3(195+1)/4=147^{th} item

Q3=163

QD=(Q3-Q1)/2

=(163-160)/2=1.5

CQD=0.0092

3. Find the quartile deviation and its relative measure.

Class(x)	**frequency(f)**
0-10	20
10-20	5
20-30	3
30-40	8
40-50	10
50-60	**35**
60-70	10
70-80	4
80-90	3
90-100	2
	$N = \sum f$ **=100**

Solution:

Class(x)	**frequency(f)**	**cf**
0-10	20	20
10-20	5	25
20-30 Q1	3	28
30-40	8	36
40-50	10	46
50-60Q3	35	**81**
60-70	10	91
70-80	4	95
80-90	3	98
90-100	2	100
	$N = \sum f$ **=100**	

Q1=(100/4)th item =25th item

Q1 lies between 20-30

$$Q1 = L + \frac{\frac{N}{4} - cf}{f} \times i$$

=20 +(25-25)/3 x10

=20

Q3= 3(100)/4th item

= 75 th item

Q3 lies between 50-60

$$Q3 = L + \frac{\frac{3N}{4} - cf}{f} \times i$$

=50+(75-46)/35 X10 = 50+0.829X10
=58.29

QD= (Q3-Q1)/2
=(58.29-20)/2
=19.15
CQD=(58.29-20)/ (58.29+20)
= 0.4890

Mean Deviation:

Mean Deviation about mean:

$$MD = \frac{\sum |D|}{N}$$ **where** $D = x - \bar{x}$ **(Individual Observations)**

(OR)

Mean Deviation about median:

$$MD = \frac{\sum |D|}{N}$$ **where** $D = x - \text{Median}$ **(Individual Observations)**

Mean Deviation about mean:

$$MD = \frac{\sum f\,|D|}{N}$$ **where** $D = x - \bar{x}$ **(Discrete series)**

(OR)

Mean Deviation about median:

$$MD = \frac{\sum f\,|D|}{N}$$ **where** $D = x - \text{Median}$ **(Discrete series)**

Mean Deviation about mean:

$$MD = \frac{\sum f\,|D|}{N}$$ **where** $D = m - \bar{x}$ **(Continuous series)**

(OR)

Mean Deviation about median:

$$MD = \frac{\sum f\,|D|}{N}$$ **where** $D = m - \text{Median}$ **(Continuous series)**

Coefficient Mean Deviation = MD/mean

Coefficient Mean Deviation = MD/median

PROBLEMS:

1. Find the MD of the set of numbers 3, 8, 6, 10, 12, 9, 11, 10, 12, 7

Soln:

X	$\lvert D\rvert=\lvert x-8.8\rvert$
3	5.8
8	0.8
6	2.8
10	1.2
12	3.2
9	0.2
11	2.2
10	1.2
12	3.2
7	1.8
Total= $\sum X=88$	$\sum\lvert D\rvert=22.4$

N=10

Mean = $\overline{X}$ =88/10

$\overline{X}$ =8.8

Mean deviation about mean= $\frac{\sum|D|}{N}$

= 22.4/10=2.24

2. Compute the Mean deviation about median.

X	**f**	**cf**	**$\lvert D\rvert=\lvert x-161\rvert$**	**$f\lvert D\rvert$**
158	15	15	3	45
159	20	35	2	40
160	32	67	1	32
161 Median	35	102	0	0
162	33	135	1	33
163	22	157	2	44
164	20	177	3	60
165	10	187	4	40
166	8	195	5	40
	N=195			$\sum f\lvert D\rvert=334$

Solution:

Median = (N+1)/2 th item=(195+1)/2th item= 98^{th} item

Median =161

Mean Deviation about median:

$$MD=\frac{\sum f\,|D|}{N} \text{ where } D = x - \text{Median} \text{ (Discrete series)}$$

Mean deviation = 334/195=1.712

Coefficient of Mean deviation = MD/Median = 1.712/161=0.010

3. Find the Mean deviation for the following data:

Size of Item	Frequency
2-4	20
4-6	40
6-8	30
8-10	10

Solution:

X	**m**	**f**	**fm**	**|D|=|m-5.6|**	**f|D|**
2-4	3	20	60	2.6	52
4-6	5	40	200	0.6	24
6-8	7	30	210	1.4	42
8-10	9	10	90	3.4	34
		N=100	$\sum fm$ **=560**		$\sum f\,\|D\|$ **=152**

Mean= $\frac{\sum fm}{N}$

=560/100=5.6

Mean Deviation about mean:

$MD = \frac{\sum f\,|D|}{N}$ **where** $D = m - \bar{x}$ **(Continuous series)**

MD=152/100=1.52

CMD=MD/Mean

=1.52/5.6=0.271

Standard Deviation:

$$\sigma = \sqrt{\frac{\sum x^2}{n}}$$ **where** $x = X - \overline{X}$ **(Individual Observations)**

$$\sigma = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$$ **where** $d = X - A$**; A- Assumed mean(Individual Observations)**

$$\sigma = \sqrt{\frac{\sum fx^2}{n}}$$ **where** $x = X - \overline{X}$ **(Discrete series)**

$$\sigma = \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2}$$ **where** $d = X - A$**; A- Assumed mean(Discrete series)**

$$\sigma = \sqrt{\frac{\sum fx^2}{n}}$$ **where** $x = m - \overline{X}$ **(Continuous series)**

$$\sigma = \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2} \times i$$ **where** $d = \frac{m - A}{i}$ **; A- Assumed mean(Continuous series)**

1. Find the Standard deviation of the set of numbers 3, 8, 6, 10, 12, 9, 11, 10, 12, 7

   Soln:

X	$x = X - 8.8$	x^2
3	-5.8	33.64
8	-0.8	0.64
6	-2.8	7.84
10	1.2	1.44
12	3.2	10.24
9	0.2	0.04
11	2.2	4.84
10	1.2	1.44
12	3.2	10.24
7	-1.8	3.24
Total= $\sum X$ =88		$\sum x^2$ =73.6

N=10

Mean = $\overline{X}$ =88/10

$\overline{X}$ =8.8

Standard deviation = $\sigma = \sqrt{\frac{\sum x^2}{n}}$ where $x = X - \overline{X}$

$$\sigma = \sqrt{\frac{73.6}{10}}$$

$$= \sqrt{7.36} = 2.712$$

2. Compute the Standard deviation.

X	**f**	fx	$x = X - 161.5128$	x^2	fx^2
158	15	2370	-3.5128	12.3398	185.0965
159	20	3180	-2.5128	6.3142	126.2833
160	32	5120	-1.5128	2.2886	73.23404
161	35	5635	-0.5128	0.2630	9.2037
162	33	5346	0.4872	0.2374	7.8330
163	22	3586	1.4872	2.2118	48.6588
164	20	3280	2.4872	6.1862	123.7233
165	10	1650	3.4872	12.1606	121.606
166	8	1328	4.4872	20.1350	100.6748
	N=195	31495			856.7055

Mean= $\frac{\sum fx}{N}$

= 31495/195=161.5128

$\sigma = \sqrt{\frac{\sum fx^2}{n}}$ **where** $x = X - \overline{X}$ **(Discrete series)**

$$= \sigma = \sqrt{\frac{856.7055}{195}}$$

=2.09603

3.Find the standard deviation for the following data:

Size of Item	**Frequency**
2-4	20
4-6	40
6-8	30
8-10	10

Coefficient of Variation:

$$CV = \frac{\sigma}{\overline{X}} \times 100$$

4. Find the Range, Quartile deviation , Mean deviation and standard deviation for the following data:

Size of Item	Frequency
2-4	20
4-6	40
6-8	30
8-10	10

Solution:

range =

10-2=8

Size of Item	Mid-points (m)	Frequency (f)	cf	fm	|D|=|m-5.6|	f |D|	d=(m-A)/2	fd	f(d^2)
2-4	3	20	20	60	2.6	52	-1	-20	20
4-6(Q1)	**5**	**40**	**60**	200	0.6	24	0	0	0
6-8Q3	**7**	**30**	**90**	210	1.4	42	1	30	30
8-10	9	10	100	90	3.4	34	2	20	40
	Total	**100**		**560**		**152**		**30**	**90**

Here N=100

$$MD = \frac{\sum f\,|D|}{N} \text{ **where** } D = m - \bar{x}$$

Mean=5.6

$$MD = \frac{152}{100}$$

MD=1.52

Q_1=(N/4)th item

=25th item

Q_1 lies in 4-6

Q_1= $L + \frac{\frac{N}{4} - cf}{f} \times i$

$$= 4 + \frac{25 - 20}{40} \times 2$$

Q_1=4.25

Q_3=(3N/4)th item

=75th item

Q_3 lies in 6-8

Q_3= $L + \frac{\frac{3N}{4} - cf}{f} \times i$

$$= 6 + \frac{75 - 60}{30} \times 2$$

Q_3=7

QD $= \frac{Q_3 - Q_1}{2}$

=(7-4.25)/2=1.375

Coefficient of QD=0.244

$\sigma = \sqrt{\frac{\sum fx^2}{n}}$ **where** $d = m - \overline{X}$

$\sigma = \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2} \times i$ **where** $d = \frac{m - A}{i}$ **; A- Assumed mean**

$$\sigma = \sqrt{\frac{90}{100} - \left(\frac{30}{100}\right)^2} \times 2$$

=1.8

CV=32.1

6. Compute Range, QD, MD and SD.

X	**f**	**cf**	**fx**	**|D|=|X-mean|**	**|D|=|X-median|**	**f|D(mean)|**	**f|D|(median)**	**x=X-mean**	**x^2**	**$f(x^2)$**
158	15	15	2370	3.5128	3	52.692	45	-3.5128	12.3398	185.0965
159	20	35	3180	2.5128	2	50.256	40	-2.5128	6.3142	126.2833
$Q_1$160	32	67	5120	1.5128	1	48.4096	32	-1.5128	2.2886	73.23404
M161	35	102	5635	0.5128	0	17.948	0	-0.5128	0.2630	9.2037
162	33	135	5346	0.4872	1	16.0776	33	0.4872	0.2374	7.8330
$Q_3$163	22	157	3586	1.4872	2	32.7184	44	1.4872	2.2118	48.6588
164	20	177	3280	2.4872	3	49.744	60	2.4872	6.1862	123.7233
165	10	187	1650	3.4872	4	34.872	40	3.4872	12.1606	121.606
166	8	195	1328	4.4872	5	35.8976	40	4.4872	20.1350	100.6748
	N=195		31495			338.6152	334			856.7055

Mean=161.5128

Median=(195+1)/2th item=98^{th} item=161

Q1=(195+1)/4th item=49^{th} item=160

Q3=3(195+1)/4th item=147^{th} item=163

QD=1.5

CQD=0.0092

MD about mean=1.736

CMD about mean=1.736/161.5128=0.01075

MD about median=1.7128

CMD about median=1.7128/161=0.01063

SD= $\sqrt{\frac{856.7055}{195}} = 2.096$

CV=(2.096/161.5128)X100=1.2977

Range=8

7. The following are scores of two batsmen A and B in a series of innings:

A: 12 115 6 73 7 19 119 36 84 29

B: 47 12 16 42 4 51 37 48 13 0

Who is better scorer and who is more consistent?

Solution:

X	x=X-mean	x^2	Y	y=Y-mean	y^2
12	-38	1444	47	20	400
115	65	4225	12	-15	225
6	-44	1936	16	-11	121
73	23	529	42	15	225
7	-43	1849	4	-23	529
19	-31	961	51	24	576
119	69	4761	37	10	100
36	-14	196	48	21	441
84	34	1156	13	-14	196
29	-21	441	0	-27	729
total=500		17498	270		3542

N=10

$$\overline{X} = \frac{\sum x}{n}$$

=500/10=50

$$\overline{Y} = \frac{\sum y}{n}$$

270/10=27

$$\sigma_x = \sqrt{\frac{\sum x^2}{n}}$$

$$\sigma_x = \sqrt{\frac{17498}{10}} = \sqrt{1749.8}$$

$$\sigma_x = 41.8306$$

$$\sigma_y = \sqrt{\frac{\sum y^2}{n}}$$

$$\sigma_y = \sqrt{\frac{3542}{10}} = \sqrt{354.2} = 18.846$$

CV for A:

$$cv = \frac{\sigma_x}{\overline{\overline{X}}} \times 100$$

83.6612%

CV for B:

CV= $\frac{\sigma_y}{\overline{\overline{Y}}} \times 100$

69.6%

Mean of A > Mean of B

Therefore A is the better player.

CV for A > CV for B

So here B is the consistent player.

MEASURES OF SKEWNESS

Skewness

Literal meaning of skewness is lack of symmetry. It measures the degree of departure of a distribution from symmetry and reveals the direction of scatterdness of the items.

A frequency distribution is said to be symmetrical when values of the variables equidistant from their mean have equal frequencies. If a frequency distribution is not symmetrical, it is said to be asymmetrical or skewed. Any deviation from symmetry is called skewness.

According to *Morris Humberg* Skewness refers to the asymmetry or lack of symmetry in the shape of a frequency distribution.

According to *Croxton & Cowden* When a series is not symmetrical it is said to be asymmetrical or skewed.

According to *Simpson & Kafka* Measures of skewness tell us the direction and the extent of skewness. In a symmetrical distribution the mean, median and mode are identical. The more we move away from the mode, the larger the asymmetry or skewness.

Symmetrical curve

The figure , given below, presents the shape of a symmetrical curve which is bell shaped having no skewness. The value of mean (M), median (M_d) and mode (M_o) for such a curve would be identical.

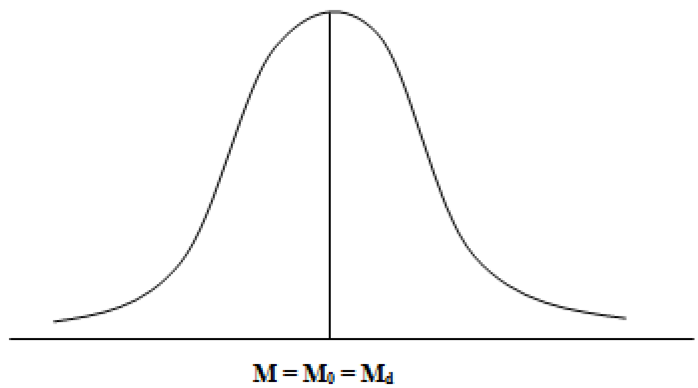

Fig. 5.1 Symmetrical distribution

In a symmetrical distribution the values of mean, median and mode coincide. The spread of the frequencies is the same on both sides of the centre point of the curve. For a symmetrical distribution Mean = Median = Mode.

Positively skewed curve

A positively skewed curve has a longer tail towards the higher values of X i.e. the frequency curve gradually slopes down towards the higher values of X. In a positively skewed distribution the mean is greater than the median and then mode and the median lies in between mean and mode. The frequencies are spread over a greater range of values on the high value end of the curve (the right hand side) as is clear from the Figure

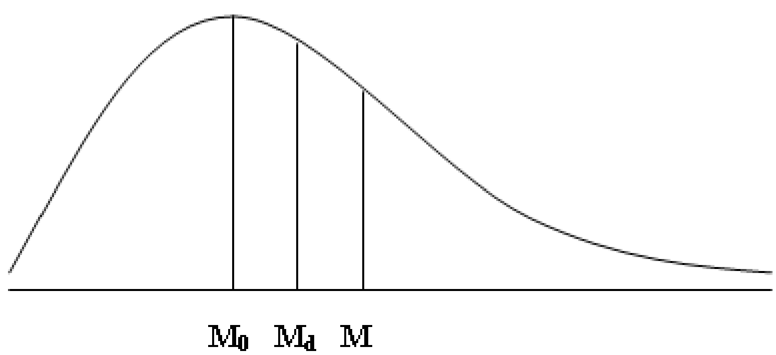

Fig. Positively skewed distribution

For a positively skewed distribution Mean > Median > Mode.

Negatively skewed curve

A negatively skewed curve has a longer tail towards the lower values of X i.e. the frequency curve gradually slopes down towards the lower values of X as shown in Figure.

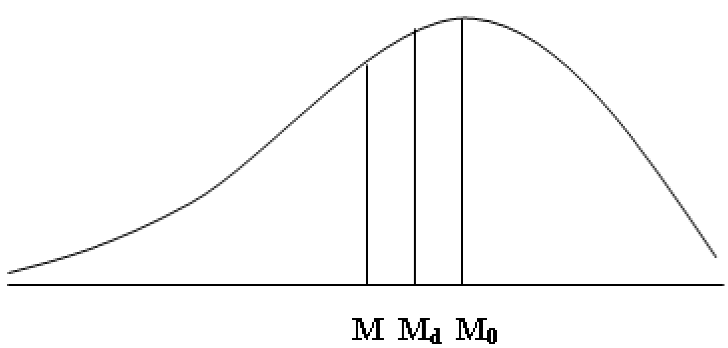

Fig. Negatively skewed distribution

In the negatively skewed distribution the mode is the maximum and mean is the least. The median lies in between mean and mode. The elongated tail in negatively skewed distribution is on the left hand side as would be clear from Figure. For a negatively skewed distribution,

Mean < Median < Mode.

Karl pearson's coefficient of skewness

The first coefficient of skewness as defined by Karl Pearson is

$$\text{Coefficient of skewness} = \frac{\text{Mean} - \text{Mode}}{\text{Std.deviation}} = \frac{M - M_0}{\sigma}$$

This measure is based on the fact that the mean and the mode are drawn widely apart. Skewness will be positive if mean > mode and negative if mean < mode. There is no limit to this measure in theory and this is a slight drawback. But in practice the value given by this formula is rarely very high and its value usually lies between -1 and +1.

It may also be written as $\frac{3(\text{Mean} - \text{Median})}{\sigma}$ as Mode = 3 Median - 2 Mean

This coefficient is a pure number without units since both numerator and denominator have the same dimensions. The value of this coefficient lies between -3 and +3.

Problems:

1. Calculate the Karl Pearson's coefficient of Skewness :
   3, 8, 6, 10, 12, 9, 11, 10, 12, 7

   Solution:
   Ascending order : 3, 6, 7,8, 9,10,10,11,12,12

   Mean= $\frac{\sum x}{n}$

   =88/10=8.8

We can't define Mode.

Median = (n+1)/2th item = (10+1)/2th item

=5.5 th item

Median = (5^{th} item+6^{th} item)/2=(9+10)/2=9.5

X	$x = X - 8.8$	x^2
3	-5.8	33.64
8	-0.8	0.64
6	-2.8	7.84
10	1.2	1.44
12	3.2	10.24
9	0.2	0.04

11	2.2	4.84
10	1.2	1.44
12	3.2	10.24
7	-1.8	3.24
Total= $\sum X=88$		$\sum x^2=73.6$

Standard deviation $=\sigma=\sqrt{\dfrac{\sum x^2}{n}}$ where $x=X-\overline{X}$

$$\sigma=\sqrt{\frac{73.6}{10}}$$

$$=\sqrt{7.36}=2.712$$

Karl Pearson's coefficient of skewness= $\dfrac{3(\overline{X}-\text{Median})}{\sigma}$

=3(8.8-9.5)/2.712=-0.77433

Negative skewed.

2. Find the Karl Pearson's coefficient of Skewness :

x	f
155	30
156	20
157	5
158	5
159	10
160	15
161	5
162	10

Solution:

X	f	fX	x=X-mean	x^2	f x^2
155(mode)	30	4650	-2.60	6.76	20.28
156	20	3120	-1.60	2.56	51.20
157	5	785	-0.60	0.36	1.8
158	5	790	0.40	0.16	0.8
159	10	1590	1.40	1.96	19.6
160	15	2400	2.40	5.76	86.4

161	5	805	3.40	11.56	57.8
162	10	1620	4.40	19.36	193.6
	$N=\sum f$ =100	$\sum fx$ =15760			$\sum f\ x^2$ =431.48

N=100

Mean= $\frac{\sum fx}{N}$

=15760/100=157.60

$\overline{X}=157.60$

Mode =155

Standard deviation = $\sqrt{\frac{\sum fx^2}{N}}$ x=X-Mean

$=\sqrt{\frac{431.48}{100}}$ =2.077

Skewness= $\frac{\overline{X}-\text{Mode}}{\sigma}$

=(157.60-155)/2.077=1.252

Positively skewed.

3.Find coefficient of skewness for the following data:

Class(x)	**frequency(f)**
0-10	20
10-20	5
20-30	3
30-40	8
40-50	10
50-60	35
60-70	10
70-80	4
80-90	3
90-100	2
	$N=\sum f$ **=100**

Solution:

Class(x)	**m**	**frequency(f)**	**fm**	x=m-mean	x^2	f x^2
0-10	5	20	100	-38	1444	28880
10-20	15	5	75	-28	784	3920

20-30	25	3	75	-18	324	972
30-40	35	8	280	-8	64	512
40-50	45	10f0	450	2	4	40
50-60(Mode)	**55**	**35f1**	**1925**	**12**	**144**	**5040**
60-70	65	10f2	650	22	484	4840
70-80	75	4	300	32	1024	4096
80-90	85	3	255	42	1764	5292
90-100	95	2	190	52	2704	5408
		$N=\sum f$ **=100**	$\sum fm$ **=4300**			**59000**

N=100

Mean = $\frac{\sum fm}{N}$ where $N=\sum f$

=4300/100=43

Modal class is 50-60

$$M_0 = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 50 + \frac{35-10}{2(35)-10-10} \times 10$$

=50+5=55

Standard deviation= $\sqrt{\frac{\sum fx^2}{N}}$ x=m-Mean

$$= \sqrt{\frac{59000}{100}} = \sqrt{590}$$

=24.28

Skewness=(43-55)/24.28=-0.4942

Negatively skewed.

4.. Find the coefficient of skewness from the following data.

A=22.5

X	f	m	cf	d=(m-A)/5	fd	d^2	fd^2	
0-5	2	2.5	2	-4	-8	16	32	
5-10	5	7.5	7	-3	-15	9	45	
10-15	7	12.5	14	-2	-14	4	28	

Q1 15-20	13	17.5	27	-1	-13	1	13	
M 20-25	21	22.5	48	0	0	0	0	
Q3 25-30	16	27.5	64	1	16	1	16	
30-35	8	32.5	72	2	16	4	32	
35-40	3	37.5	75	3	9	9	27	
Total	75				-9		193	

Q1=15+ $\frac{18.75-14}{13}\times 5$ **=16.827**

Median= 20+ $\frac{37.5-27}{21}\times 5$ **=22.5**

Q3=27.57

Sk=(27.57+16.827-45)/(27.57-16.827)=-0.055

Negatively skewed.