Unit → 2

Measures of Central Tendency

In this chapter questions related to continuous series are more imp. Partion values and others are lesser imp.

1) Mean: (Arithmetic Mean)

i) For individual series:

$$\text{Mean}(\bar{X}) = \frac{\Sigma x}{n}$$ (which is direct formula for calculating mean)

We can also calculate mean by using deviation as:-

$$\text{Mean}(\bar{X}) = A + \frac{\Sigma d}{n}$$ where, A is assumed mean. & $d = X - A$.

ii) For discrete series:

$$\text{Mean } \bar{X} = \frac{\Sigma fx}{n}$$ (Direct formula)

By using deviation

$$\text{Mean}(\bar{X}) = A + \frac{\Sigma fd}{N}$$ where, f is frequency

By using step-deviation

$$\text{Mean}(\bar{X}) = A + \frac{\Sigma fd'}{N} \times h$$ where, $d' = \frac{X-A}{h}$.

$A \rightarrow$ Assumed mean

$h \rightarrow$ common factor.

iii) For continious series:

The formulas are same as in discrete series but here X is taken as mid-value of each class interval.

⊛. Properties of Arithmetic Mean:

i) Sum of the deviation from mean is always zero.

i.e, $\Sigma(X - \bar{X}) = 0$ (for individual series)

$\Sigma f(X - \bar{X}) = 0$ (for discrete series).

$\frac{1}{h}\Sigma f(X - \bar{X}) = 0$ (for continious series).

ii) Let $\bar{X}_1$ and $\bar{X}_2$ are two means of series having size n_1 and n_2. Then its combined mean is denoted by $\bar{X}_{12}$ and given as:

$$\bar{X}_{12} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1+n_2}$$ Similarly $$\bar{X}_{123} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + n_3\bar{X}_3}{n_1+n_2+n_3}$$

iii) Sum of squares of deviation is minimum when deviation is taken from mean. i.e, $\Sigma(X-A)^2$ is minimum when $A = \bar{X}$.

Mean related questions (Numerical Questions):

Q.1 The mean marks got by 300 students in the subject of statistics are 45. The mean of the top 100 of them was found to be 70 and the mean of last 100 was found to be 20. What is the mean of remaning 100 students.

Soln

Given,

Top 100 students	Remaning 100 students	Last 100 students.
Size $(n_1) = 100$ Average marks i.e, $(\bar{X}_1) = 70$	Size $(n_2) = 100$ Average marks $(\bar{X}_2) = ?$	Size $(n_3) = 100$ Average marks $(\bar{X}_3) = 20$

and average mark of 300 student $(\bar{X}_{123}) = 45$.

Now,

$$\text{Combined mean } (\bar{X}_{123}) = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + n_3\bar{X}_3}{n_1+n_2+n_3}$$

$$\text{or, } 45 = \frac{100\times70 + 100\bar{X}_2 + 100\times20}{100+100+100}$$

$$\text{or, } 135000 = 7000 + 100\bar{X}_2 + 2000$$

$$\text{or, } \bar{X}_2 = 45 \text{ Ans.}$$

Q.2. The mean monthly salery paid to all employes in certain company is Rs 600. The mean monthly saleries paid to male and female were Rs 620 and Rs 520 respectively. Obtain the percentage of male and female employees in the company.

Soln

Given, Combined mean of saleries of male and female $(\bar{X}_{12}) = 600$.

Mean salery of male $(\bar{X}_1) = 620$

Mean salery of female $(\bar{X}_2) = 520$.

Let n_1 and n_2 be the no. of male and female respectively.

We have, $$\text{Combined mean } (\bar{X}_{12}) = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1+n_2}$$

$$\text{or, } 600 = \frac{620n_1 + 520n_2}{n_1+n_2}$$

$$\text{or, } 600n_1 + 600n_2 = 620n_1 + 520n_2$$

or, $600n_2 - 520n_2 = 620n_1 - 600n_1$

or, $20n_1 = 80n_2$

or, $\frac{n_1}{n_2} = \frac{80}{20}$

or, $\frac{n_1}{n_2} = \frac{4}{1}$

i.e, $n_1 : n_2 = 4:1$

Now, Percentage of male $= \frac{4}{4+1} \times 100$

$= 80\%$

& Percentage of female $= (100-80)\%$

$= 20\%$

Q.N.3. The mean of 100 item was 50 later on it was found that 2 items were mis-read as 92 and 8 instead of 192 and 88. Find the correct mean.

Soln.

Given, no. of observation $(n) = 100$

mean $(\bar{x}) = 50$

We have, $(\bar{x}) = \frac{\Sigma x}{n}$

or, $50 = \frac{\Sigma x}{100}$

or, $\Sigma x = 5000$

Incorrect values are 92 and 8

& Correct values are 192 and 88.

Add correct values and subtract incorrect values always to get correct.

Now, Correct $(\Sigma x) = 5000 + 192 + 88 - 92 - 8$

$= 5180.$

So, Correct mean ($\bar{x}$ correct) $= \frac{\Sigma x \text{ (correct)}}{n}$

$= \frac{5180}{100}$

$= 51.80$

Q.N.4. Find the missing frequencies from the following distribution of the sales of shops given at the mean sales of the shops is 24.625.

Sales	No. of shops
0-10	5
10-20	25
20-30	–
30-40	–
40-50	7

Given $N = 80$

Soln

Let x and y be the missing frequencies of class intervals 20-30 and 30-40 respectively. Now for the calculation of x and y we construct table as follows:-

Sales	No. of shops	mid-value (X)	fX
0-10	5	5	25
10-20	25	15	375
20-30	x	25	$25x$
30-40	y	35	$35y$
40-50	7	45	315
	$N = 80 = x+y+37$		$\Sigma fX = 25x+35y+715$

From table

$$x+y+37 = 80$$

i.e. $x+y = 43$ —— (i).

and mean $(\bar{X}) = \dfrac{\Sigma fx}{N}$

or, $24.625 = \dfrac{25x+35y+715}{80}$

or, $25x+35y+715 = 1970$

or, $25x+\quad = 1255$ —— (ii)

Solving (i) and (ii)

$$25x+25y = 43\times25 = 1075$$

$$25x+35y = 1255$$

$$-10y = -180$$

$$y = 18$$

from (i) $x+18 = 43$

or, $x = 43-18 = 25$

∴ Missing frequencies of class 20-30 and 30-40 are 18 and 25 respectively.

Scanned with CamScanner

2> Median:

Median is the value which divides the whole distribution into two equal parts and 50% values are less than median and 50% values are more than median.

Calculation of median

ⓐ For individual series → First arrange the given series into ascending or descending order. Then, $\boxed{\text{median} = \text{value of} \left(\frac{n+1}{2}\right)^{th} \text{item}}$

where, 'n' is no. of observation.

For example: Find the median from the following series.

Soln 15, 14, 18, 12, 9, 6, 8, 4.

For the calculation of median arranging the dat in ascending order → 4, 6, 8, 9, 12, 14, 15, 18.

Now, median $= \text{value of} \left(\frac{n+1}{2}\right)^{th}$ item.

$= \text{value of} \left(\frac{8+1}{2}\right)^{th}$ item

$= \text{value of } (4.5)^{th}$ item

$= \frac{9+12}{2}$

$= 10.5$ is median value.

OR

$\boxed{\text{Value of } 4^{th} \text{ item} + 0.5\,(5^{th} \text{ item} - 4^{th} \text{ item})}$

$= 9 + 0.5\,(12-9)$

$= 9 + 0.5 \times 3$

$= 10.5$

ⓑ. For discrete series → First construct commulative frequency (c.f) distribution (less than).

→ Find median position, $\frac{N+1}{2}$.

→ See the cumulative frequency just equal or greater than $\frac{N+1}{2}$. Then the corresponding value variable is median.

For example Find the median from the following data.

Wages (Rs)	No. of workers (f)
5	20
7	15
8	12
10	15
11	18

Solⁿ

Now, cumulative frequency distribution table is:

Wages (Rs)	frequency	Less than cumulative frequency (c.f)
5	20	20
7	15	35
8	12	47
10	15	62
11	18	80
	N=80.	

and $\frac{N+1}{2} = \frac{80+1}{2}$

$= 40.5$

Since the c.f just greater than 40.5 is 47 whose corresponding value is 8.

ⓒ For continious series → Construct c.f distribution table

→ Find $N/2$

→ See the c.f just equal or greater than $N/2$. Then corresponding class interval is median class interval

→ Now, $\text{median} = l + \frac{N/2 - c.f}{f} \times h.$

where, l = lower limit of median class

N = total frequency (Σf).

$c.f$ = cumulative frequency just preceding median class.

f = frequency.

h = size of c.f (width of class interval).

Example Find out the median from following data.

Wages	No. of worker
10–14	4
15–19	7
20–24	8
25–29	3
30–34	2

Solⁿ

For the calculation of median we construct following table.

inclusive type data is converted into exclusive.

Wages	No. of workers (f)	less than c.f
9.5–14.5	4	4
14.5–19.5	7	11
19.5–24.5	8	19
24.5–29.5	3	22
29.5–34.5	2	24
	N=24.	

Now, $N/2 = 24/2 = 12$.

The c.f just greater than 12 is 19.

So median class interval is 19.5–24.5.

Now,

$$\text{Median} = l + \frac{N/2 - c.f}{f} \times h$$

$$= 14.5 + \frac{12-11}{8} \times 5$$

$$= 20.15$$

Q. The following table represents the marks of 100 students.

Marks	No. of students (f).
0–20	14
20–40	—
40–60	27
60–80	—
80–100	15

If the median mark is 48 find the missing frequencies.

Solⁿ Let x and y be the missing frequencies of class interval 20–40 and 60–80 respectively.

For the calculation of x and y.

Marks	No. of students (f)	less than c.f.
0–20	14	14
20–40	x	$14+x$
40–60	27	$41+x$
60–80	y	$x+y+41$
80–100	15	$x+y+56$
	$N=100=x+y+56$.	

Scanned with CamScanner

From above table,

$$x+y+56=100$$
$$\text{or, } x+y=44 \quad \text{--- (i)}$$

Since median is 48, So median class interval is 40–60.

Now,

$$\text{median} = l + \frac{N/2 - c.f}{f} \times h$$

$$\text{or, } 48 = 40 + \frac{50-(14+x)}{27} \times 20$$

$$\text{or, } 8 = \frac{(50-14-x)}{27} \times 20$$

$$\text{or, } 20x = 504$$

$$\text{or, } x = 25.2$$
$$= 25$$

From eq^n (i)

$$x+y=44$$
$$\text{or, } 25+y=44$$
$$\text{or, } y=19$$

Hence the missing frequencies of class interval 20-40 and 60–80 are 25 and 19 respectively.

⊛. Partition values:–

Partition values are those values which divides the whole distribution into no. of equal parts. Three types of partition values as follows:-

ⓐ. Quartiles → Quartiles are those values which divides whole distribution into four equal parts, Q_1, Q_2 and Q_3.

25% | 25% | 25% | 25%

Q_1 Q_2 Q_3

ⓑ. Deciles → Deciles are those values which divides whole distribution into 10 equal parts. each of 10%., D_1, D_2 to D_9

10% | 10% | . . . . | 10%

D_1 D_2 . . . D_9.

ⓒ. Percentiles → Percentiles are those values which divides whole distribution into 100 equal parts, P_1 to P_{99}.

1% 1% 1%

P_1 P_2 P_3 . . . . P_{99}

Calculation of partition values:-

S.N	Quartile	Deciles	Percentiles
1.	For Individual Series		
i)	Arrange the given series into ascending or descending order	i) "	i) "
ii)	Quartile (Q_i) = value of $\frac{i(n+1)}{4}^{th}$ item	ii) Deciles (D_j) = value of $\frac{j(n+1)}{10}^{th}$ item	ii) Percentile (P_k) = value of $\frac{k(n+1)}{100}^{th}$ item.
	For discrete Series.		
i)	Construct c.f distribution	"	"
ii)	Find $\frac{i(N+1)}{4}$	Find $\frac{j(N+1)}{10}$	Find $\frac{k(N+1)}{100}$
iii)	~~Find~~ See the c.f just equal or greater than $\frac{i(N+1)}{4}$	"	"
iv)	Corresponding value is the quartile value	"	"
3.	For Continuous Series		
i)	Construct c.f distribution	"	"
ii)	Find $iN/4$	ii) Find $jN/10$	ii) Find $kN/100$.
iii)	See the cf just equal or greater than $iN/4$	"	"
iv)	Then corresponding class interval is quartile class interval.	iv) "	iv) "
v)	Now quartile is given by $Q_i = l + \frac{iN/4 - c.f}{f} \times h$	v) " Decile " $D_j = l + \frac{jN/10 - c.f}{f} \times h$	v) " Percentile " $P_k = l + \frac{kN/100 - c.f}{f} \times h$

where, $i = 1, 2, 3$. $j = 1, 2, \ldots, 9$. & $k = 1, 2, \ldots, 99$.

Q.N.1 The marks distribution of 100 students of a college is as follows:-

Marks	No. of student
10-20	15
20-40	20
40-70	30
70-90	20
90-100	15
	N=100

i) Find the highest marks of the weakest 30% of the studend.

ii) Find the lowest marks of the top 40% of students.

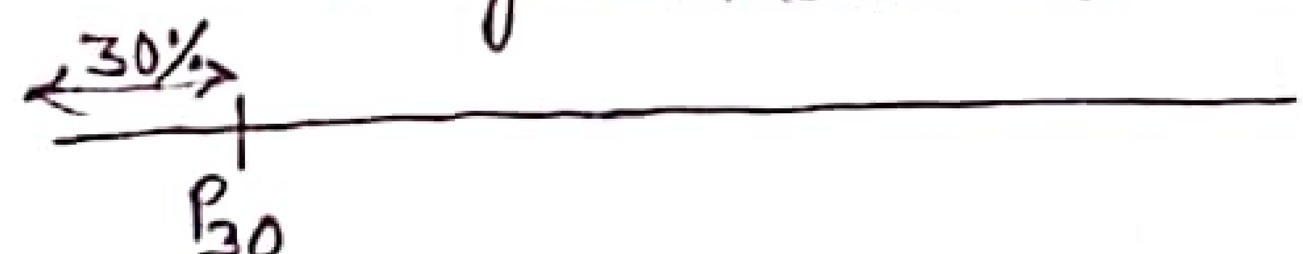
iii) Find the limits and range of marks of middle 50% students.

Soln

For the calculation of partition values we reconstruct the table as follows:-

Marks	No. of students	Less than c.f.
10-20	15	15
20-30	10	25
30-40	10	35
40-50	10	45
50-60	10	55
60-70	10	65
70-80	10	75
80-90	10	85
90-100	15	100
	N=100	

i) Here, P_{30} shows the highest marks of weakest 30% of students.



For P_{30} in continious series $P_{30} = \frac{30N}{100} = 30$

The c.f just greater than 30 is 35, So, P_{30} lies between 30 - 40.

Now, $P_{30} = l + \frac{\frac{30N}{100} - c.f}{f} \times h$

$= 30 + \frac{30-25}{10} \times 10$

$= 35.$

ii) Here, P_{60} shows the lowest marks of top 40% of students.

So, for P_{60}

Class interval $= \frac{60N}{100} = \frac{60 \times 100}{100} = 60$

So, the c.f just greater than 60 is 65. Hence P_{60} lies between interval 60-70.

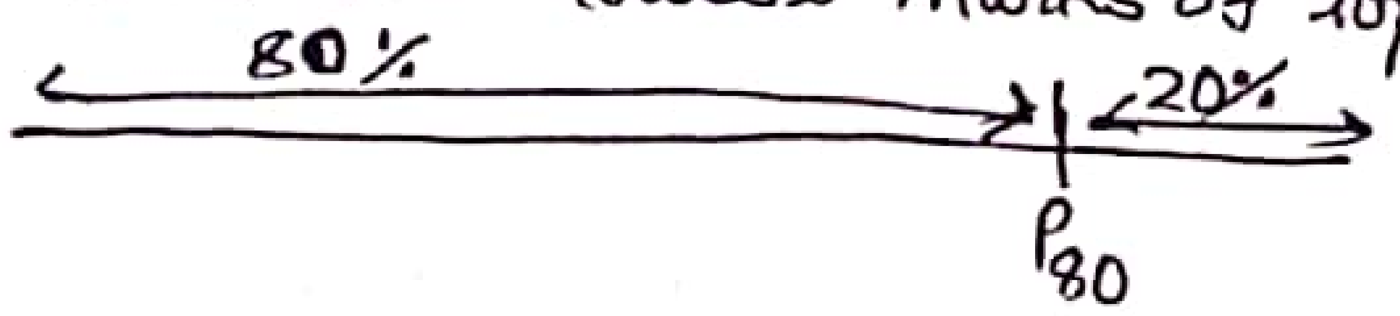
$$\therefore P_{60} = l + \frac{60N/100 - c.f}{f} \times h$$

$$= 60 + \frac{60-55}{10} \times 10$$

$$= 65$$

iii) Here, P_{80} shows the lowest marks of top 20% of students.

~~for~~



For P_{80}

Class interval $= \frac{80N}{100} = \frac{80 \times 100}{100} = 80$

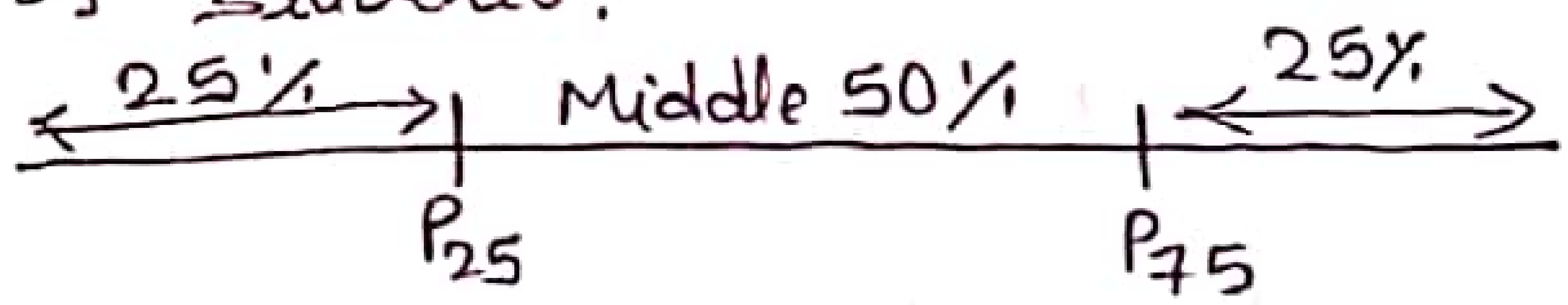
The c.f just greater than 80 is 85. So, P_{80} lies between 80-90.

$$\therefore P_{80} = l + \frac{80N/100 - c.f}{f} \times h$$

$$= 80 + \frac{80-75}{10} \times 10$$

$$= 85$$

iv) Here, P_{25} and P_{75} represents the limits of marks of middle 50% of students.



For P_{25}

Class interval $= \frac{25N}{100} = \frac{25 \times 100}{100} = 25$

The c.f just greater than 25 is 25. So P_{25} lies between 20-30.

$$\text{Now, } P_{25} = l + \frac{25N/100 - c.f}{f} \times h$$

$$= 30$$

~~&~~ For P_{75}

Class interval $= \frac{75N}{100} = \frac{75 \times 100}{100} = 75$

The c.f just equal to 75 is 75. So, P_{75} lies between 70-80.

$$\text{Now, } P_{75} = l + \frac{75N/100 - c.f}{f} \times h$$

$$= 80$$

Hence the limits of marks of middle 50% students are $P_{25} = 30$ & $P_{75} = 8$. & Range of marks of middle 50% students $= 80 - 30$
$= 50$.

Scanned with CamScanner

Q.N.2. 120 students appeared for a certain test and the following marks distribution was obtained as:

Marks	No of students
0-20	10
20-40	30
40-60	36
60-80	30
80-100	14
Total	120

i> Find the limits of marks of middle 30% students

ii> Find the percentage of students getting more than 75.

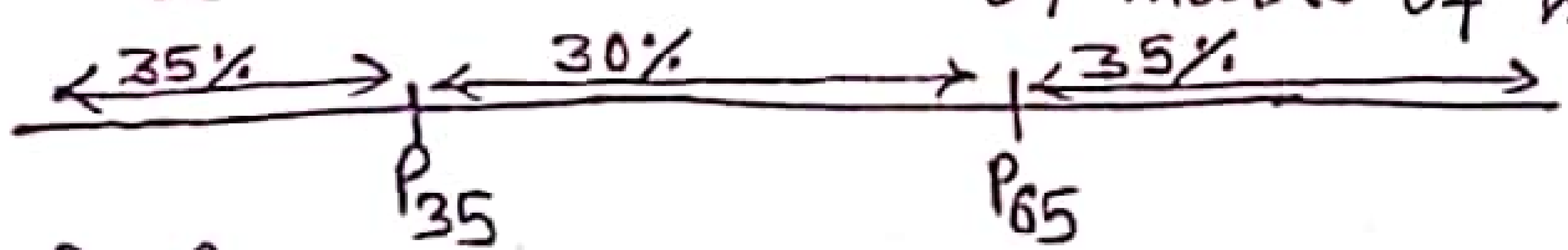
iii> Find the number of student who fail, if 35 marks are required for passing.

Soln

Here,

Marks	No. of students.	Less than c.f
0-20	10	10
20-40	30	40
40-60	36	76
60-80	30	106
80-100	14	120
	N=120	

i> P_{35} and P_{65} shows the limits of marks of middle 30% students.



Now, for P_{35}

$$\text{Class interval} = \frac{35N}{100} = \frac{35 \times 120}{100} = 42$$

The c.f just greater than 42 is 76 So, 42 lies between 40-60 class interval.

$$P_{35} = l + \frac{\frac{35N}{100} - c.f}{f} \times h$$

$$= 40 + \frac{42 - 40}{36} \times 20$$

$$= 41.11$$

ii for P_{65}

$$\text{Class interval} = \frac{65N}{100} = \frac{65 \times 120}{100} = 78.$$

The c.f just greater than 78 is 106 so, P_{65} lies between 60-80.

Scanned with CamScanner

$$P_{65} = l + \frac{\frac{65N}{100} - c.f}{f} \times h$$

$$= 60 + \frac{78-76}{30} \times 20$$

$$= 61.33$$

Hence the limits of marks of middle 30% students are $P_{35} = 41.11$ and $P_{65} = 61.33$.

ii) Here the number of students who got the marks more than

$$75 = \frac{80-75}{20_{\leftarrow C.I}} \times \overset{f}{30} + 14$$

c.f just greater than 75

f भन्दा तलको सबैलाई add गरेको less than भाको भए माथिको सबै add हुन्थ्यो ,

$$= 21.5$$

Now, Percentage of student who got marks more than 75

$$= \frac{21.5}{120} \times 100$$

$$= 17.5$$

$$\approx 18\%$$

iii) Since 35 is pass mark. So, no. of students who fail in exam is who got marks less than 35 is

$$= \frac{35-20}{20} \times 30 + 10$$

just smaller than 35

$$= 32.5$$

$$\approx 33\%$$

Not imp

3. Mode:- Mode is the value of variable which is repeated maximum number of times or which is having maximum frequency.

For Continious series → First we need to find model class interval having maximum frequency then mode is given by,

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

where,

L = lower limit of model C.I

f_1 = frequency of model class interval

f_0 = frequency of previous C.I

f_2 = frequency of next C.I

& h = size of class interval/width.

Q. The model marks for a group of 94 students is 54. Find the number of students getting the marks between 20-40 and 60-80. From the data if the maximum marks of the examination was 100.

Marks	No. of students
0-20	10
20-40	—
40-60	30
60-80	—
80-100	14
Total	N=94

Soln

Let x and y be the missing frequency of class interval 20-40 and 60-80.

Marks	No. of student (f)
0-20	10
20-40	x
40-60	30
60-80	y
80-100	14
	$N=94=x+y+54$

Since mode is 54, model class in C.I is 40-60.

So, $M_0 = L + \frac{f_1-f_0}{2f_1-f_0-f_2} \times h$

or, $54 = 40 + \frac{30-x}{2\times30-x-y} \times 20$

or, $14 = \frac{600-20x}{60-x-y}$

or, $6x - 14y = -240$

or, $3x - 7y = -120$ —(i)

From the above table $x+y+54=94$

or, $x+y=40$ —(ii)

Solving (i) and (ii)×3

$\cancel{3x} + 3y = 120$

$\cancel{3x} - 7y = -120$

$- \quad + \quad \quad +$

$10y = 240$

or, $y = 24$

Putting the value of y in eqn (ii) $x=16$.

Hence the missing frequency of C.I 20-40 and 60-80 are $x=16$ and $y=24$

Note: (a) If the distribution is not symmetrical then Mean $\neq$ Median $\neq$ Mode. Then the relation between mean, mode and median is called empirical relation.

i.e, $\boxed{\text{Mode} = 3\,\text{Median} - 2\,\text{Mean}}$

If the distribution is symmetrical then

$\boxed{\text{Mean} = \text{Median} = \text{Mode}}$

(b). The weighted arithmetic mean is ~~given by~~ denoted by $\bar{x}_w$ and given as
$$\bar{x}_w = \frac{w_1x_1 + w_2x_2 + \dots + w_nx_n}{w_1 + w_2 + \dots + w_n}$$

i.e.
$$\boxed{\bar{x}_w = \frac{\sum wx}{\sum w}}$$

where, $x_1, x_2 \dots x_n$ are variable values & $w_1, w_2, \dots w_n$ are corresponding weights.

Measures of Dispersion (Measures of variability)

i) Range
ii) Semi-inter Quartile Range (or Quartile deviation)
iii) Mean deviation (or Absolute deviation)
iv) Standard deviation
v) Coefficient of variation

These two are mainly important for exam point of view.

i) **Range:** Range is the difference between largest value and smallest value i.e, $X_L - X_S$.

& Coefficient of Range $= \dfrac{X_L - X_S}{X_L + X_S}$

where, X_L = Largest value & X_S = Smallest value.

ii) **Quartile deviation**

Quartile range $= Q_3 - Q_1$

& Semi-quartile range $= \dfrac{Q_3 - Q_1}{2}$ (or Quartile Deviation)

Coefficient of Quartile deviation $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$.

Q. The following frequency distribution represents the weight of 200 laptops.

Weights in lbs	frequency	Weight in lbs	frequency
4–5	20	8–9	32
5–6	24	9–10	24
6–7	35	10–11	8
7–8	48	11–12	2

Compute the first three quartiles and quartile deviation.

Soln

Weights in lbs	frequency (f)	Cumulative frequency
4–5	20	20
5–6	24	44
6–7	35	79
7–8	48	127
8–9	32	159
9–10	24	183
10–11	8	191
11–12	2	193
	N=193	

For Q_1

$$N/4 = \frac{193}{4} = 48.12$$

the c.f just greater than 48.12 is 79 so, Q_1 lies between 6–7.

$$\text{So, } Q_1 = L + \frac{N/4 - c.f}{f} \times h.$$

$$= 6 + \frac{48.12 - 44}{35} \times 1$$

$$= 6.12$$

For Q_2

$$2N/4 = N/2 = 193/2 = 96.5$$ The c.f just greater than 96.5 is 127 so it lies in 7–8.

$$\text{So, } Q_2 = L + \frac{N/2 - c.f}{f} \times h. = 7 + \frac{96.5 - 79}{1.8} \times 1 = 7.36.$$

For Q_3 $$3N/4 = 3 \times 193/4 = 144.75$$

The c.f just greater than 144.75 is 159 so, it lies in 8–9.

$$Q_3 = L + \frac{3N/4 - c.f}{f} \times h = 8 + \frac{144.75 - 127}{32} \times 1 = 8.64$$

Now,

Coefficient of Quartile deviation $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \dfrac{8.64 - 6.12}{8.64 + 6.12}$

$= 0.17$

& Quartile deviation (Q.D) $= \dfrac{Q_3 - Q_1}{2}$

$= \dfrac{8.64 - 6.12}{2}$

$= 1.26$

Mean deviation (or Absolute deviation):-

i) For Individual series:-

Mean deviation from median $= \frac{1}{n} \sum |x - Med|$.

Mean deviation from mean $= \frac{1}{n} \sum |x - \bar{x}|$.

Mean deviation from mode $= \frac{1}{n} \sum |X - M_0|$.

ii) For Discrete series:-

Mean deviation from mean $= \frac{1}{N} \sum f |x - \bar{x}|$

Mean deviation from median $= \frac{1}{N} \sum f |X - Med|$ $\boxed{\text{where, } N = \sum f}$

Mean deviation from mode $= \frac{1}{N} \sum f |X - M_0|$.

iii) For continious series the formulas are same as in discrete series but here X is taken as mid-values of each class interval.

Coefficient of mean deviation from mean is given by,

$= \dfrac{\text{Mean deviation from mean}}{\text{mean}}$

Similarly

Coefficient of ~~med~~ mean deviation from median $= \dfrac{\text{Mean deviation from median}}{\text{median}}$

& Coefficient of mean deviation from mode $= \dfrac{\text{Mean deviation from mode}}{\text{mode}}$.

Standard Deviation (σ):

i) For individual series

$$\sigma = \sqrt{\frac{1}{n} \sum (x - \bar{x})^2} \quad \text{OR} \quad \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} \quad \text{OR} \quad \sqrt{\frac{\sum x^2}{n} - (\bar{x})^2}$$

- By using deviation: Let $d = x - A$,

$$\sigma = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$$

ii) For discrete series

$$\sigma = \sqrt{\frac{1}{N}\Sigma f(X-\bar{X})^2} \quad \text{OR} \quad \sqrt{\frac{\Sigma fX^2}{N} - \left(\frac{\Sigma fX}{N}\right)^2}$$

by using deviation let $d = X - A$,

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

also by using step-deviation

let $d' = X - A$, where A = assumed value & h = common factor

$$\sigma = h\sqrt{\frac{\Sigma fd'^2}{2}}$$

iii) For continious series the formulas are same as in discrete but here X is taken as mid-value.

Q. Find the standard deviation by using direct and deviation method of following data.

X: 1, 2, 3, 4, 5.

Soln. For the calculation of σ.

X	$(X-\bar{X})$	$(X-\bar{X})^2 = (X-3)$	$d = X-A = X-2$	d^2
1	−2	4	−1	1
2	−1	1	0	0
3	0	0	1	1
4	1	1	2	4
5	2	4	3	9
$\Sigma X = 15$	$\Sigma(X-\bar{X}) = 0$	$\Sigma(X-\bar{X})^2 = 10$	$\Sigma d = 5$	$\Sigma d^2 = 15$

Here, $\bar{X} = \frac{\Sigma X}{N} = 3$.

By using direct formula. $\sigma = \sqrt{\frac{1}{n}\Sigma(X-\bar{X})^2} = \sqrt{\frac{1}{5} \times 10} = \sqrt{2} = 1.414.$

By using deviation

$$\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} = \sqrt{\frac{15}{5} - \left(\frac{5}{5}\right)^2} = \sqrt{3-1} = \sqrt{2} = 1.414$$

Variance (σ^2) → Square of standard deviation is called variance.

⊛. Combined standard Deviation (σ_{12}):

Let x_1 and x_2 be the two series having mean $\bar{x}_1$ and $\bar{x}_2$ with size n_1 and n_2 and standard deviations σ_1 and σ_2 respectively. Then its combined deviation is denoted by σ_{12} and given as:

$$\sigma_{12} = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$

where, $d_1 = \bar{x}_1 - \bar{x}_{12}$

and $d_2 = \bar{x}_2 - \bar{x}_{12}$

$$\bar{x}_{12} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}.$$

Similarly

$$\sigma_{123} = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2) + n_3(\sigma_3^2 + d_3^2)}{n_1 + n_2 + n_3}}$$

where, $d_1 = \bar{x}_1 - \bar{x}_{123}$

$d_2 = \bar{x}_2 - \bar{x}_{123}$

$d_3 = \bar{x}_3 - \bar{x}_{123}$

and $\bar{x}_{123} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3}$

⊛. Coefficient of variation (CV):-

Coefficient of variation is used for comparing the variability of two or more than two series. For comparing variability we have coefficient of variation,

$$C.V = \frac{\sigma}{\bar{X}} \times 100.$$

Let A and B are two series such than $C.V(A) < C.V(B)$ then A is less variable than B OR B is more heterogenous than A OR B is less consistent than A.

Combined coefficient of variation $= \dfrac{\sigma_{12}}{\bar{X}_{12}} \times 100.$

Q1. The score obtained by 10 students in statistics on an IT college are given below. Compute the range and standard deviation.

55, 35, 60, 55, 65, 40, 45, 35, 42.

Soln

For the calculation of Range,

We have Range $= X_L - X_S$

$= 65 - 30$

$= 30,$

For the calculation of S.D.

We have, $\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$

X	$d = x - A$ $= x - 55$	d^2
55	0	0
35	−20	400
60	5	25
55	0	0
65	10	100
40	−15	225
45	−10	100
35	−20	400
42	−13	169
	$\Sigma d = -63$	$\Sigma d^2 = 1419$

Now, $\sigma = \sqrt{\frac{1419}{9} - \left(\frac{-63}{9}\right)^2}$

$= \sqrt{157.66 - 49}$

$= 10.42$

Q2. Compute the mean deviation from the mean of following data.

10, 12, 11, 15, 20, 24, 23, 26, 16.

Soln Here, n=10. We have deviation from mean $= \frac{1}{n} \Sigma |x - \bar{x}|$

X	$x - \bar{x}$ $(x - 17.5)$	$\lvert x - \bar{x} \rvert$
10	−7.5	7.5
12	−5.5	5.5
11	−6.5	6.5
15	−2.5	2.5
20	2.5	6.5
24	6.5	5.5
23	5.5	8.5
26	8.5	1.5
16	−1.5	

Now, $\Sigma x = 175$

$\Sigma (x - \bar{x}) = 0$

$\Sigma |x - \bar{x}| = 47$

Mean $= \frac{\Sigma x}{n} = \frac{175}{10} = 17.5$

$\therefore$ Mean deviation from mean $= \frac{1}{10} \times 47$

$= 4.7$

Q.3 The mean and standard deviation of 200 items are found to be 60 and 20 respectively. If at the time of calculations, two items were wrongly taken as 3 and 67 instead of 13 and 17, find the correct mean and standard deviation. What is the correct coefficient of variation?

Soln

Here, $n = 200, \bar{x} = 60, \sigma = 20 \quad \therefore \sigma^2 = 400$

We have,

$$\bar{x} = \frac{\Sigma x}{n} \text{ or, } \Sigma x = n\bar{x}$$

$$= 200 \times 60$$

We have,

$$= 12000$$

$$\sigma^2 = \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2$$

$$\therefore \Sigma x^2 = n(\sigma^2 + \bar{x}^2)$$

$$= 200(400 + 3600)$$

$$= 8{,}00{,}000$$

If the wrong items 3 and 67 are replaced by the correct values 13 and 17 respectively, we get,

$$\text{Correct } \Sigma x = 12000 - 3 - 67 + 13 + 17$$

$$= 11960$$

$$\text{Correct } \Sigma x^2 = 800000 - 3^2 - 67^2 + 13^2 + 17^2$$

$$= 800000 - 9 - 4489 + 169 + 289$$

$$= 795960$$

$$\therefore \text{Corrected Mean } (\bar{x})_{correct} = \frac{\text{Correct } \Sigma x}{n}$$

$$= \frac{11960}{200}$$

$$= 59.8$$

$$\text{Correct s.d. } \sigma = \sqrt{\frac{\text{Correct } \Sigma x^2}{n} - (\text{Correct mean})^2}$$

$$= \sqrt{\frac{795960}{200} - (59.8)^2}$$

$$= 20.09$$

$$\therefore \text{Correct C.V} = \frac{\text{Correct } \sigma}{\text{Correct } \bar{x}} \times 100\% = \frac{20.09}{59.8} \times 100\% = 33.60\%$$

Q.4 Following are sample of ages of 6 male and 6 female enrolled in hardware training program of computer.

Age of male	23	30	18	25	32	40
Age of female	20	35	44	27	41	38

Find which gender has uniform age.

Solution

Age of male (x)	x^2	Age of female (y)	y^2
23	529	20	400
30	900	35	1225
18	324	44	1936
25	625	27	729
32	1024	41	1681
40	1600	38	1444
$\Sigma x = 168$	$\Sigma x^2 = 5002$	$\Sigma y = 205$	$\Sigma y^2 = 7415$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{168}{6} = 28$$

$$\sigma_x = \sqrt{\frac{1}{n-1}\left(\Sigma x^2 - n\bar{x}^2\right)} = \sqrt{\frac{1}{6-1}\left(5002 - 6\times 28^2\right)} = \sqrt{59.6} = 7.72$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{205}{6} = 34.16$$

$$\sigma_y = \sqrt{\frac{1}{n-1}\left(\Sigma y^2 - n\bar{y}^2\right)} = \sqrt{\frac{1}{6-1}\left(7415 - 6\times(34.16)^2\right)} = \sqrt{82.71} = 9.09$$

$$C.V_x = \frac{\sigma_x}{\bar{x}} \times 100\% = \frac{7.72}{28} \times 100\% = 27.57\%$$

$$C.V_y = \frac{\sigma_y}{\bar{y}} \times 100\% = \frac{9.09}{34.16} \times 100\% = 26.61\%$$

Here, $C.V_y < C.V_x$

Hence female have uniform age.

Q.5. Sample of polyethene bags from two manufacturers, A and B are tested by a prospective buyer for bursting pressure and the results are as follows:-

Bursting Pressure (lb.)	Number of Bags	
	A	B
5.0–9.9	2	9
10.0–14.9	9	11
15.0–19.9	29	18
20.0–24.9	54	32
25.0–29.9	11	27
30.0–34.9	5	13

Which set of bags has more uniform pressure? If prices are the same which manufacturer's would be preferred by the buyer? Why?

Solution.

Let X be the mid value so that $d' = \frac{X - 17.45}{5}$. Let f and f' be the number of bags of manufacturers A and B respectively.

Class	m.v. (X)	$d' = (X-17.45)/5$	f	f'	fd'	fd'^2	f'd'	$f'd'^2$
5-9.9	7.45	-2	2	9	-4	8	-18	36
10-14.9	12.45	-1	9	11	-9	9	-11	11
15-19.9	17.45	0	29	18	0	0	0	0
20-24.9	22.45	1	54	32	54	54	32	32
25-29.9	27.45	2	11	27	22	44	54	108
30-34.9	32.45	3	5	13	15	45	39	117
		$\Sigma d'=3$	N=110	N'=110	$\Sigma fd'=78$	$\Sigma fd'^2=160$	$\Sigma f'd'=96$	$\Sigma f'd'^2=304$

For manufacturer A

Mean, $\bar{X}_A = A + \frac{\Sigma fd'}{N} \times h$

$= 17.45 + \frac{78}{110} \times 5$

$= 21$

S.D. $(\sigma_A) = h\sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2}$

$= 5\sqrt{\frac{160}{110} - \left(\frac{78}{110}\right)^2}$

$= 5\sqrt{1.45 - 0.50}$

$= 5 \times 0.975$

$= 4.875$.

C.V. $(A) = \frac{\sigma_A}{\bar{X}_A} \times 100\%$

$= \frac{4.875}{21} \times 100\%$

$= 23.2\%$

For manufacturer B

Mean, $\bar{X}_B = A + \frac{\Sigma f'd'}{N} \times h$

$= 17.45 + \frac{96}{110} \times 5$

$= 21.8$

S.D. $(\sigma_B) = h\sqrt{\frac{\Sigma f'd'^2}{N} - \left(\frac{\Sigma f'd'}{N}\right)^2}$

$= 5\sqrt{\frac{304}{110} - \left(\frac{96}{110}\right)^2}$

$= 5\sqrt{2.76 - 0.76}$

$= 5 \times 1.41$

$= 7.05$

C.V. $(B) = \frac{\sigma_B}{\bar{X}_B} \times 100\%$

$= \frac{7.05}{21.8} \times 100\%$

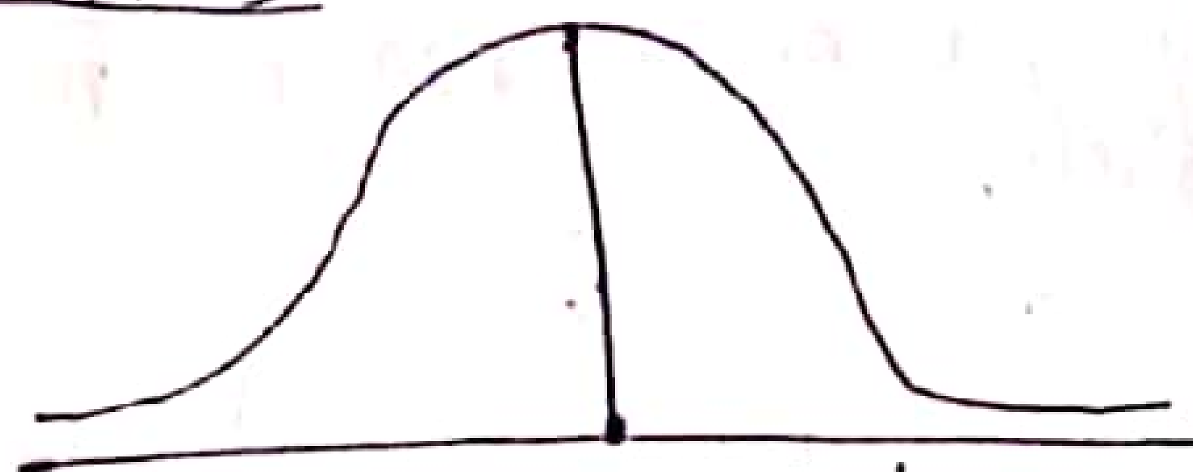
$= 32.34\%$

Since the coefficient of variation for A is less than that for B, the set of bags manufactured by A has more uniform pressure. Hence the buyer would prefer to buy bags manufactured by A.

⑧. Skewness:-

Lack of symmetry is the skewness. OR Skewness gives us an idea about the shape of the distribution so, there are three types of skewness.

i) Zero skewness (Symmetrical)

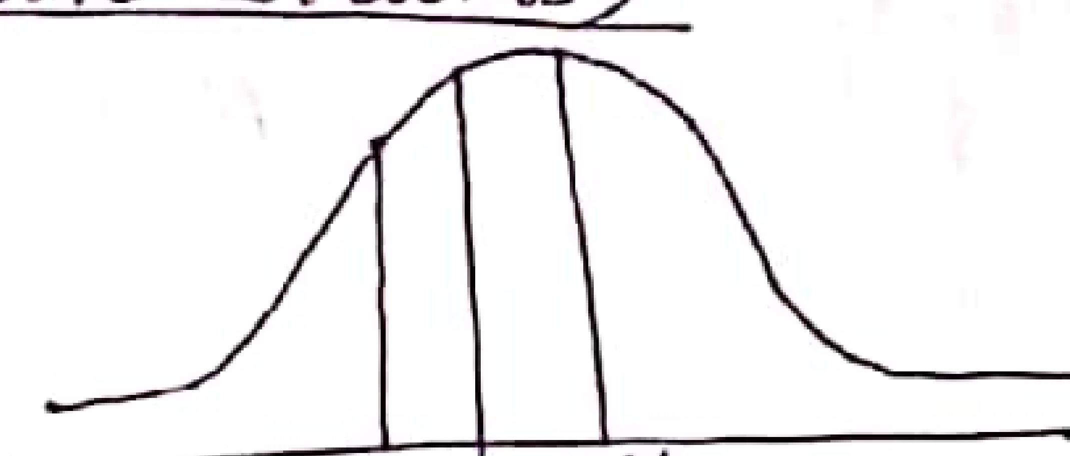


$\bar{X} = M_0 = Med$

ii) Right skewness (Positive Skewness)

$M_0 < Med < \bar{X}$

iii) Left skewness (Negative skewness)



$\bar{X} < Med < M_0$

Coefficients of Skewness:

1) Karl Pearson's coefficient of skewness → It is based on Mean, Mode & Standard deviation. Absolute measure of skewness $= \bar{X} - M_0$.

Coefficient of skewness $\{S_k(P)\} = \dfrac{\bar{X} - M_0}{\sigma}$

And Karl Pearsons coefficient lies between -1 to $+1$.

i.e, $-1 \leq S_k(P) \leq +1$.

If mode is not defined then we have $\boxed{Mode = 3Med - 2\bar{X}}$

Now,

$\boxed{\text{Range} = -3 \leq S_k(P) \leq +3}$

2) Bowley's coefficient of skewness → It is based on quartiles Q_1, Q_2 (Median), Q_3.

Coefficient of skewness $S_k(B) = \dfrac{(Q_3 - Med) - (Med - Q_1)}{(Q_3 - Med) + (Med - Q_1)}$

or, $S_k(B) = \dfrac{Q_3 + Q_1 - 2Med}{Q_3 - Q_1}$

$\boxed{\text{Range} -1 \leq S_k(B) \leq +1}$

Q1. The sum of 20 observations is 300 and its sum of squares is 5000 and median is 15. Find the coefficient of skewness and coefficient of variation.

Soln

Given,

$n = 20$

$\Sigma x = 300$

$\Sigma x^2 = 5000$

Median $= 15$

We have,

$$S_K(P) = \frac{3(\bar{X} - Med)}{\sigma}$$

$$\text{st. dev. } (\sigma) = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

$$= \sqrt{\frac{5000}{20} - (25)^2}$$

$$= \sqrt{250 - 225}$$

$$= \sqrt{25}$$

$$= 5$$

$$\text{Now, } S_K(P) = \frac{3(15-5)}{5}$$

$$\text{or, } S_K(P) = 0$$

$$\text{\& } C.V = \frac{\sigma}{\bar{X}} \times 100\%$$

$$= \frac{5}{15} \times 100\%$$

$$= 33.33\%$$

Q2. For a moderately skewed data, the arithmetic mean is 200, the coefficient of variation is 8% and Karl Pearson's coefficient of skewness is 0.3. Find mode and median.

Soln

Given,

Mean $(\bar{X}) = 200$

Coefficient of variation $(C.V) = 8\%$

$S_K(P) = 0.3$

Mode $= ?$

Median $= ?$

We have,

$$C.V = \frac{\sigma}{\bar{X}} \times 100$$

$$\text{or, } 8 = \frac{\sigma}{200} \times 100$$

$$\text{or, } \sigma = 16$$

$$\text{and } S_K(P) = \frac{\bar{X} - M_o}{\sigma} \quad \text{or, } 0.3 = \frac{200 - M_o}{16} \quad \text{or, } M_o = 195.2$$

Now,

$$M_o = 3\,\text{Median} - 2\,\text{Mean}$$

or, $195.2 = 3\,\text{Med} - 2 \times 200$

or, $3\,\text{Med} = 195.2 + 400$

or, $\text{Med} = \frac{195.2 + 400}{3}$

or, $\text{Med} = 198.4$

Q.3. Compute Karl Pearson's coeffcient from following data.

Mid value of income	150	250	350	450	550	650	750	850
No of staff in IT company	80	105	120	165	100	90	60	40

Soln. For the calculation of Karl Pearson's coefficient of skewness we have, $S_k(P) = \frac{\bar{X} - M_o}{\sigma}$.

Mid-value (X)	Interval	frequency (f)	$d' = \frac{X-A}{h} = (X-50)/100$	fd'	fd'^2
150	100-200	80	-3	-240	720
250	200-300	105	-2	-210	420
350	300-400	120	-1	-120	120
450	400-500	165	0	0	0
550	500-600	100	1	100	100
650	600-700	90	2	180	360
750	700-800	60	3	180	540
850	800-900	40	4	160	540
		N=760		$\Sigma fd' = 50$	$\Sigma fd'^2 = 2900$

Here, $\text{Mean}(\bar{X}) = A + \frac{\Sigma fd'}{N} \times h = 450 + \frac{50}{760} \times 100 = 456.57$

$$\text{Standard deviation}(\sigma) = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times h$$

$$= \sqrt{\frac{2900}{760} - \left(\frac{50}{760}\right)^2} \times 100$$

$$= 195.22$$

and $M_o = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h = 400 + \frac{165 - 120}{2 \times 165 - 120 - 100} \times 100$

$= 440.90$.

Now, $S_K(P) = \frac{\bar{X} - M_o}{\sigma} = \frac{456.57 - 440.40}{195.22}$

$= 0.08$

Q.4. In a distribution, the difference of the two quartiles is 15 and their sum is 35 with median 20. Find the coefficient of skewness.

Solution: Here, ~~$S_k(B)=0,\ Q_1=45$~~

$$Q_3-Q_1=15$$
$$Q_3+Q_1=35$$
$$Med = 20.$$

$$S_k(B)=\frac{Q_3+Q_1-2Md}{Q_3-Q_1}=\frac{35-2\times20}{15}=-0.333$$

The coefficient of skewness $S_k(B)=-0.333<0$, hence the distribution is negatively skewed.

Q5. For a symmetrical distribution, first quartile is 45 and semi-inter quartile range is 5, find the median.

Solution.

Here, $S_k(B)=0,\ Q_1=45,\ Q_3-Q_1=5$

$$Md=?$$

$$Q_3-Q_1=5$$
or, $Q_3=5+Q_1$
or, $Q_3=5+45$
$=50$

We know, $$S_k(B)=\frac{Q_3+Q_1-2Md}{Q_3-Q_1}$$

or, $0=\dfrac{50+45-2Md}{5}$

or, $0=95-2Md$

or, $2Md=95$

or, $Md=\dfrac{95}{2}$

or, $Md=47.5$

⊛. **Kurtosis:** It measures the peakedness or flatness of curve. According to their peakedness and flatness there are three types of curves.

i) Leptokurtic → Highest peaked curve.
ii) Mesokurtic → It is normal curve neither more flat nor more peaked.
iii) Platykurtic → More flat curve among three curves.

⊛. **Percentile coefficient of kurtosis:**

$$\boxed{K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}}$$

Conditions
i) If value of $K = 0.263$ then the curve is mesokurtic.
ii) If value of $K > 0.263$ then the curve is leptokurtic.
iii) If value of $K < 0.263$ then the curve is platykurtic.

Q1. Find percentile coefficient of kurtosis from following data.
3, 14, 22, 7, 16, 25, 11, 19, 27.

Solution:
Arranging given data in ascending order; 3, 7, 11, 14, 16, 19, 22, 25, 27.

$n = 9$.

$$P_{10} = \left(\frac{10(n+1)}{100}\right)^{th} \text{item} = \frac{10(9+1)}{100} = 1^{st} \text{ item} = 3.$$

$$P_{90} = \left(\frac{90(n+1)}{100}\right)^{th} \text{item} = \frac{90(9+1)}{100} = 9^{th} \text{ item} = 27$$

$$Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{item} = \frac{(9+1)}{4} = 2.5^{th} \text{ item}$$
$$= \frac{2^{nd} \text{ item} + 3^{rd} \text{ item}}{2}$$
$$= \frac{7+11}{2}$$
$$= 9.$$

$$Q_3 = \left[\frac{3(n+1)}{4}\right]^{th} \text{item} = \frac{3\times(9+1)}{4} = 7.5^{th} \text{ item}$$
$$= \frac{7^{th} \text{ item} + 8^{th} \text{ item}}{2}$$
$$= \frac{22+25}{2}$$
$$= 23.5$$

Now,
$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \frac{23.5 - 9}{2(27-3)}.$$
$$= 0.302$$

Here, $K = 0.302 > 0.263$, hence the distribution is leptokurtic.

⊛. Moments:

1) Central moment: (μ_r)

$\mu_r \rightarrow$ rth central moment.

ⓐ For individual series

$$\boxed{\mu_r = \frac{\Sigma (x-\bar{x})^r}{n}}$$

$$\mu_1 = \frac{\Sigma (x-\bar{x})^1}{n}$$

$$= 0 \qquad [\because \Sigma (x-\bar{x}) = 0]$$

μ_1 is first moment about mean or, 1st central moment.

Similarly $\mu_2 = \sigma^2 =$ variance.

$$\mu_3 = \frac{\Sigma (x-\bar{x})^3}{n}$$

& $\mu_4 = \frac{\Sigma (x-\bar{x})^4}{n}$.

ⓑ. For Discrete series

$$\boxed{\mu_r = \frac{\Sigma f (x-\bar{x})^r}{N}}$$ where $N = \Sigma f$

$\mu_1 = 0$

$\mu_2 = \sigma^2 =$ variance

$\mu_3 = \frac{\Sigma f (x-\bar{x})^3}{N}$

& $\mu_4 = \frac{\Sigma f (x-\bar{x})^4}{N}$.

ⓒ. For continious series formulas are same as in discrete series but here X is taken as mid-value of each class interval.

ii) Raw moment: (μ_r')

For individual series $\boxed{\mu_r' = \frac{\Sigma (X-A)^r}{n} = \frac{\Sigma d^r}{n}}$ $\because d = X-A$.

$\mu_1' = \frac{\Sigma x}{n} - A = \bar{X} - A$. or, $\boxed{\bar{X} = \mu_1' + A}$

$\mu_2' = \frac{\Sigma d^2}{n}$

$\mu_3' = \frac{\Sigma d^3}{n}$

& $\mu_4' = \frac{\Sigma d^4}{n}$

For discrete series

$\boxed{\mu_r' = \frac{\Sigma f d^r}{N}}$ $\because d = X-A$.

So, $\mu_1' = \frac{\Sigma f d}{N}$

$\mu_2' = \frac{\Sigma f d^2}{N}$

$\mu_3' = \frac{\Sigma f d^3}{N}$

& $\mu_4' = \frac{\Sigma f d^4}{N}$

Relation between Raw and Central moments:

i) $\mu_1 = \mu_1' - \mu_1' = 0$

ii) $\mu_2 = \mu_2' - (\mu_1')^2$

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

iv) $\mu_4 = \mu_4' - 4\mu_3'\cdot\mu_1' + 6\mu_2'\cdot(\mu_1')^2 - 3(\mu_1')^4$

Q1. Compute first four central moments from following series.

Class interval	0-10	10-20	20-30	30-40	40-50
frequency	5	7	9	7	5

Soln. For the calculation of first four moments

Class interval	Mid-value (X)	frequency	$(X-\bar{X})$ $=X-25$	$f(X-\bar{X})$	$f(X-\bar{X})^2$	$f(X-\bar{X})^3$	$f(X-\bar{X})^4$
0-10	5	5	-20	-100	2000	-40000	800000
10-20	15	7	-10	-70	700	-70000	70000
20-30	25	9	0	0	0	0	0
30-40	35	7	10	70	700	7000	70000
40-50	45	5	20	100	2000	80000	800000
		N=33		$\Sigma f(X-\bar{X}) = 0$	5400	0	1740000

Here, Mean $(\bar{X}) = \frac{\Sigma fX}{N}$

$$= \frac{5\times5 + 15\times7 + 25\times9 + 35\times7 + 45\times5}{33}$$

$$= 25$$

Now,

$$\mu_1 = \frac{0}{N} = 0$$

$$\mu_2 = \frac{5400}{33} = 263.63$$

$$\mu_3 = \frac{0}{33} = 0$$

$$\mu_4 = \frac{1740000}{33} = 52727.27$$

Q.2. Compute the first four moments about arbitrary point 4 from following distribution.

x	2	3	4	5	6
f	1	3	7	2	1

Soln.

Scanned with CamScanner

For the calculation of first four moment about arbitary point A = 4.

x	f	$(X-A) = X-4$	$f(X-A)$	$f(X-A)^2$	$f(X-A)^3$	$f(X-A)^4$
2	1	-2	-2	4	-8	16
3	3	-1	-3	3	-3	3
4	7	0	0	0	0	0
5	2	1	2	2	2	2
6	1	2	2	4	8	16
	$N=14$		$\Sigma f(X-A) = -1$	$\Sigma f(X-A)^2 = 13$	$\Sigma f(X-A)^3 = -1$	$\Sigma f(X-A)^4 = 37$

Now,

$$\mu_1' = \frac{\Sigma f(X-A)}{N} = \frac{-1}{14} = 0.071$$

$$\mu_2' = \frac{\Sigma f(X-A)^2}{N} = \frac{13}{14} = 0.928$$

$$\mu_3' = \frac{\Sigma f(X-A)^3}{N} = \frac{-1}{14} = 0.071$$

$$\mu_4' = \frac{\Sigma f(X-A)^4}{N} = \frac{37}{14} = 2.642$$

Q3. The first four raw moments about an arbitrary origin of a frequency distribution are -2, 14, -20, 50. Find first four central moments.

Soln Given, First four raw moment about origin are $\mu_1' = -2$, $\mu_2' = 14$, $\mu_3' = 20$ and $\mu_4' = 50$.

Now,

$$\mu_1 = \mu_1' - \mu_1' = 0$$

$$\mu_2 = \mu_2' - (\mu_1')^2 = 14 - (-2)^2$$
$$= 14 - 4$$
$$= 10.$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$$
$$= -20 - 3 \times 14 \times (-2) + 2(-2)^3$$
$$= -20 + 84 - 16$$
$$= 48$$

and $\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2' \cdot \mu_1'^2 - 3(\mu_1')^4$
$$= 50 - 4(-20)(-2) + 6 \times 14 \times (-2)^2 - 3(-2)^4$$
$$= 50 - 160 + 336 - 48$$
$$= 178$$

//

* Coefficient of skewness in terms of moment.

Coefficient of skewness $\boxed{\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}}$

Conditions:

If $\gamma_1 = 0$ i.e, no skewness

If $\gamma_1 < 0$ i.e, negative skewness

& if $\gamma_1 > 0$ i.e, positive skewness.

*. Coefficient of Kurtosis in terms of moment:

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

Also, $\gamma_2 = \beta_2 - 3$

if $\beta_2 = 3$ then the curve is mesokurtic.

if $\beta_2 > 3$ then the curve is leptokurtic.

if $\beta_2 < 3$ then the curve is platykurtic.

Q1. If the first four moments about mean are 0, 2.8, −2. and 24.5 respectively. Compute coefficient of skewness and kurtosis and comment upon result.

Soln. Given,

first-four central moments are $\mu_1 = 0$, $\mu_2 = 2.8$, $\mu_3 = -2$ and $\mu_4 = 24.5$.

Now,

$$\text{Coefficient of skewness } (\gamma_1) = \frac{\mu_3}{\mu_2^{3/2}}$$

$$= \frac{-2}{(2.8)^{3/2}}$$

$$= -0.427$$

Since γ_1 is less than 0, so there is negative skewness.

& Coefficient of kurtosis $(\beta_2) = \frac{\mu_4}{\mu_2^2}$

$$= \frac{24.5}{(2.8)^2}$$

$$= 3.125$$

Since $\beta_2 > 3$ so, the curve is leptokurtic.

Q2. The first three moments of a distribution about value 5 of a distribution are 2, 20 and 40 respectively. Compute measure of central tendency, dispersion and skewness.

Soln

The first three moments about the value 5 are

$\mu_1' = 2$, $\mu_2' = 20$ and $\mu_3' = 40$.

(i) We have, $\mu_1' = \bar{x} - A$

or, $\bar{x} = \mu_1' + A$

$= 2 + 5$

$= 7$

So measure of central tendency $(\bar{x}) = 7$

(ii) We have, $\mu_2 = \mu_2' - (\mu_1')^2$

$= 20 - (2)^2$

$= 16$

$= \sigma^2$

Here, $\mu_2 = 16$ & $\sigma \Rightarrow 4$

So, measure of dispersion is 4.

(iii). and $\mu_3 = \mu_3' - 3\mu_2' \cdot \mu_1' + 2(\mu_1')^3$

$= 40 - 3\times20\times2 + 2\times2^3$

$= 40 - 120 + 16$

$= -64$.

Now, coefficient of skewness $(\gamma_1) = \dfrac{\mu_3}{\mu_2^{3/2}}$

$= \dfrac{-64}{(16)^{3/2}}$

$= \dfrac{-64}{64}$

$= -1$

Q3. If the S.D of a symmetrical distribution is 11, what ~~is~~ must be the value of fourth moment about the mean in order that the distribution be i) Mesokurtic ii) Platykurtic iii) Leptokurtic.

Soln.

S.d. dev $(\sigma) = 11$

$\Rightarrow \sigma^2 = 11^2 = 121$

If the distribution is mesokurtic then, $\beta_2 = \frac{\mu_4}{\mu_2^2} = 3.$

$$\Rightarrow \frac{\mu_4}{(121)^2} = 3$$

$$\Rightarrow \mu_4 = 43923$$

If the distribution is leptokurtic then,

$$\beta_2 = \frac{\mu_4}{\mu_2^2} > 3$$

$$= \frac{\mu_4}{(121)^2} > 3$$

$$= \mu_4 > 43923.$$

Five number summary:

Five number summary is the set of five values X_S, Q_1, Q_2, Q_3 and X_L. Five number summary is useful for finding skewness whether the distribution is positively skewed, negatively skewed or zero skewed.

where, $X_S \rightarrow$ smallest value
$Q_1 \rightarrow$ First quartile / Lower quartile.
$Q_2 \rightarrow$ second quartile / median
$Q_3 \rightarrow$ third quartile / Upper quartile
$X_L \rightarrow$ Largest value

Q. Construct the five number summary from the following values;

Soln: 10, 18, 9, 6, 5, 11, 15, 20.

First arranging the given data in ascending order
5, 6, 9, 10, 11, 15, 18, 20.

Here, No. of observations (n) = 8.

$X_S = 5$ and $X_L = 20$.

For Q_1

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

= value of $\left(\frac{8+1}{4}\right)^{th}$ item

= value of $(2.25)^{th}$ item

= value of 2nd item + 0.25 (3rd − 2nd) item

$= 6 + 0.25\,(9-6)$

$= 6.75$

For Q_2

Value of $\left(\frac{n+1}{2}\right)^{th}$ item

$= $ value of $\left(\frac{8+1}{2}\right)^{th}$ item

$= $ value of $(4.5)^{th}$ item

$= $ value of 4^{th} item $+ 0.5(5^{th} - 4^{th})$

$= 10 + 0.5(11 - 10)$

$= 10.5$

For Q_3 value of $\frac{3(n+1)}{4}^{th}$ item

$= $ value of $3\left(\frac{9}{4}\right)^{th}$ item

$= $ value of 6.75^{th} item

$= $ value of 6^{th} item $+ 0.75(7^{th} - 6^{th})$

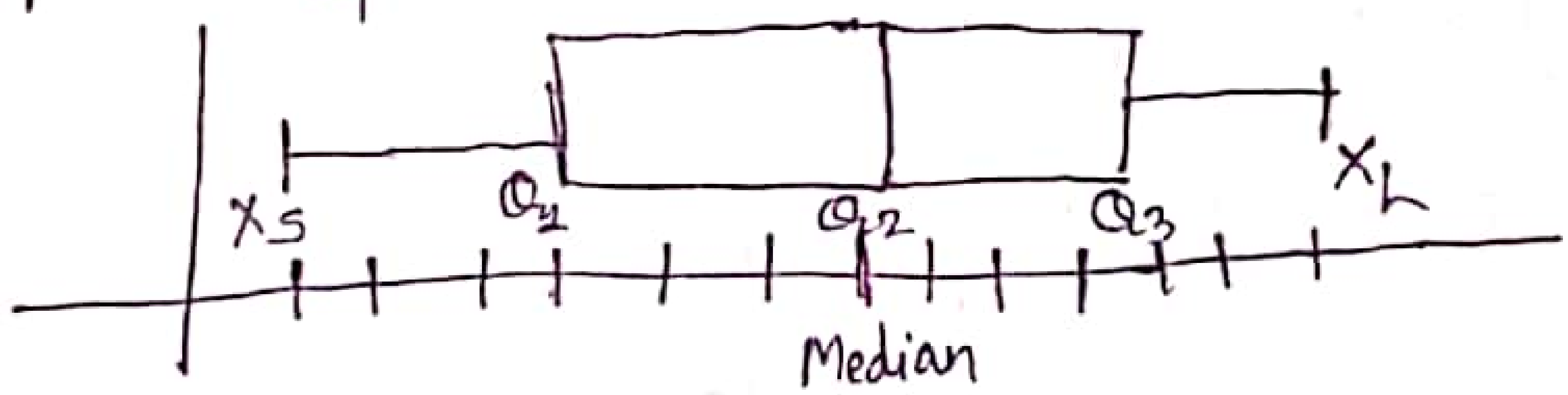
$= 15 + 0.75(18 - 15)$

$= 17.25$

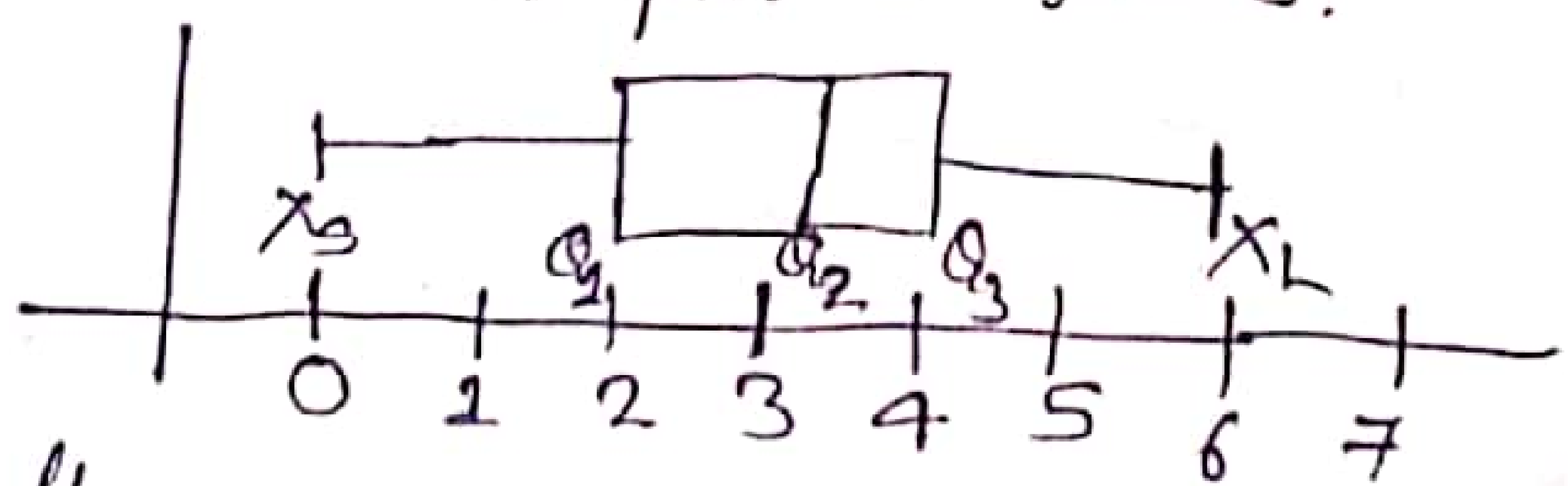
$\therefore$ Five number summary is $\{X_S, Q_1, Q_2, Q_3, X_L\}$

$= \{5, 6.75, 10.5, 17.25, 20\}$.

Box-Whisker plot → It is the graphical presentation of five number summary which contains the values X_S, Q_1, Q_2, Q_3 and X_L. These values can be plotted in graph along X-axis and Y-axis.



Eg. Let five number summary is $\{0, 2, 3, 4, 6\}$, then we construct box-plot as follows.



Note: If the length from Q_1 to Q_2 is equal to distance from Q_2 to Q_3 then the given data is zero-skewed or symmetrical.

Q. The following data represents number of laptos passed from a checkpoint in a week.

52 47 35 40 28 95 62.

Compute (a). Mean, three quartiles, sample standard deviation from above data.

(b) Check whether there are outliers or not.

(c). If outliers are present then delete the outliers and compute mean, three quartiles and sample standard deviation.

Soln. For the calculation of mean and quartiles.

28, 35, 40, 47, 52, 62, 95.

Now, Mean $= \frac{\Sigma_i X}{n} = \frac{28+35+40+47+52+62+95}{7} = 51.28$

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

$= $ value of $\left(\frac{7+1}{4}\right)^{th}$ item

$=$ value of 2^{nd} item

$= 35$

$Q_2 = 47$

For Q_3 value of $3\left(\frac{n+1}{4}\right)^{th}$ item $=$ value of 6^{th} item $= 62$..

We know,

Sample of standard deviation$(S) = \sqrt{\frac{1}{n-1}\Sigma(X-\bar{X})^2}$

$= \sqrt{\frac{1}{n-1}(\Sigma_i X^2 - n\bar{x}^2)}$

X	X^2
52	2704
47	2200
35	1225
40	1600
28	784
45	9028
62	3844
	$\Sigma X^2 = 21391$

$\therefore S = \sqrt{\frac{1}{7-1}(21391 - 7(51.28)^2}$

or, $S = 22.29$.

(b) Now,

Inter quatile range

Lower outlier limit $= Q_1 - 1.5 \times I.R$

$= 35 - 1.5(62-35)$

$= 35 - 1.5 \times 27$

$= -5.5$

Upper outlier limit $= Q_3 + IR \times 1.5$

$= 62 + 27 \times 1.5$

$= 102.5$

(c). No any aultier layers are present in this.