① find the missing frequency from the following data: Here the given mean is 15.38.

Central value (x)	frequency (f)	fx
10	3	30
12	7	84
14	m	$14m$
16	20	320
18	8	144
20	5	100
	$N = 43+m$	$\Sigma fx = 678+14m$

Here, $\bar{x} = \frac{\Sigma fx}{N}$

$$\bar{x} = \frac{678+14m}{43+m}$$

$$15.38 = \frac{678+14m}{43+m}$$

$$15.38(43+m) = 678+14m$$

$$\Rightarrow 661.34 + 15.38m = 678+14m$$

$$\Rightarrow 1.38m = 16.66$$

$$\therefore m = \frac{16.66}{1.38} = 12.17 \simeq 12$$

Hence missing frequency $= 12$

② An incomplete distribution is given below:

Variable:	0–10	10–20	20–30	30–40	40–50	50–60	60–70
Frequency:	10	20	?	40	?	25	15

It is given that total frequency is 170, the median value is 35.

Using the median formula find the missing frequencies.

$\Rightarrow$

$$\text{Med.} = L_1 + \frac{\frac{N}{2} - c.f}{f} \times i \quad , \quad \frac{N}{2} = \frac{170}{2} = 85\text{th } c.f$$

Here median $= 35$, which lies in the 30–40 c.i. interval

$$35 = 30 + \frac{85 - (10 + 20 + f_1)}{40} \times 10$$

$110 + f_1 + f_2 = 170$

$f_1 + f_2 = 60$

$\therefore f_1 = 35$

$\therefore f_2 = 25$

---

(3) An incomplete distribution is given below:

Marks:	10–20	20–30	30–40	40–50	50–60	60–70	70–80
No. of Students:	12	30	?	65	?	25	18

Total $= 229$, median $= 46$.

Using median formula find the missing frequency.

$\boxed{\text{Ans: } f_1 = 34,\ f_2 = 45,\ \bar{x} = 45.87}$

$\bar{X} = 28, A = 35\ fd^1 = -70, N = 80 + x, C = 10$

$$28 = 35 + \frac{-70}{80+x} \times 10 + x\ 10$$

$$28(80 + x) = 35(80 + x) - 700$$

$$2240 + 28x = 2800 + 35x - 700$$

$$28x - 35x = 2800 - 700 - 2240$$

$$-7x = -140$$

$$x = 20$$

Hence the missing frequency is 20

Median : Median is the size of N/2 th item = 100/2 = 50th item
Median lies in the class 20-30

$$\text{Median} = L_1 + \frac{N/2 - Cf}{6} \times i$$

$L_1 = 20,\ N/2 = 50,\ Cf = 30, f = 27, i = 10$

$$\text{Median} = 20 + \frac{50-30}{27} \times 10 = 20 + \frac{20}{27} \times 10 = 20 + 7.41$$

$$= 27.41$$

MODE

Mode is the most common item of a series. Mode is the value which occurs the greatest number of frequency in a series. It is derived from the French word 'La mode' meaning the fashion. Mode is the most fashionable or typical value of a distribution, because it is repeated the highest number of times in the series. According to Croxton and Cowden, "The mode of a distribution is the value at the point around which the item tend to be most heavily concentrated." Mode is defined as the value of the variable which occurs most frequently in a distribution. The mode in a distribution is that item around which there is a maximum concentration. The chief feature of mode is that it is the size of that item which has the maximum frequency and is also affected by the frequencies of the neighbouring items.

Method of calculation of mode

Mode can often be found out by mere inspection in case of individual observations. The data have to be arranged in the form of an array so that the value which has the highest frequency can be known. For example : 10 persons have the following income :

Rs. 850, 750, 600, 825, 850, 725, 600, 850, 640, 530

850 repeats three times, therefore the mode salary is Rs. 850.

In certain cases that there may not be a mode or there may be more than one mode. For example,

(*a*) 40, 44, 57, 78, 48 (No mode)

(*b*) 45, 55, 50, 45, 40, 55, 45, 45 (Bimodal)

Mode (*i*) = 45 : Mode (*ii*) = 55

When we calculate the mode from data, if there is only one mode in the series, it is called unimodal; if there are two modes, it is called bimodal; if there are three modes, it is called trimodal and if there are more than three modes, it is called multi-modal.

Discrete series

We cannot depend on the method of inspection to find out the mode. In such situations, it is suggested to prepare a grouping table and an analysis table to find out the mode. First we prepare grouping table and then an analysis table.

Steps for the calculation of mode

1. Prepare a grouping table with 6 columns.
2. Write the size of the item in the margin.
3. In column 1, write the frequencies against the respective items.
4. In column 2, the frequencies are grouped in twos.
   (1 and 2, 3 and 4, 5 and 6 and so on)
5. In column 3, the frequencies are grouped in twos, leaving the first frequency. (2 and 3, 4 and 5, 6 and 7 and so on).
6. In column 4, the frequencies are grouped in threes.
   (1, 2 and 3; 4, 5 and 6; 7, 8, and 9; and so on)
7. In column 5, the frequencies are grouped in threes, leaving the first frequency.
   (2, 3 and 4; 5, 6 and 7; 8, 9 and 10 and so on).
8. In column 6, the frequencies are grouped in threes, leaving the first two frequencies. (3, 4 and 5; 6, 7 and 8 so on).
   In all the processes mark down the maximum frequencies by bold letters or by a circle.
9. After grouping the frequencies table, an analysis table is prepared to show the exact size, which has the highest frequency.

Illustration : 38.

Calculate the mode from the following :

Size	Frequency
10	10
11	12
12	15
13	19
14	20
15	8
16	4
17	3
18	2

Solution : **Grouping table**

Size	Frequency 1	2	3	4	5	6
10	10	22				
11	12		27	37		
12	15	34			46	
13	19		39			54
14	20	28		47		
15	8		12		32	
16	4	7				15
17	3		5	9		
18	2					

Analysis table

Col. No.	Size of item containing maximum frequency				
	11	12	13	14	15
1					
2		1	1	1	
3			1	1	
4			1	1	1
5	1	1	1		
6		1	1	1	
	1	3	5	4	1

The mode is 13, as this size of item repeats five times. But through inspection, we say the mode is 14, because the size 14 occurs 20 times. But this wrong decision is revealed by analysis table.

Calculation of Mode—Continuous Series :

In a continuous series, to find out the mode, we need one step more than those used for discrete series. As explained in the discrete series, modal class is determined by preparing grouping table and analysis tables. Then we apply the following formula :

$$Z = L_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Z = Mode

L_1 = Lower limit of the modal class

f_1 = Frequency of the modal class

f_0 = Frequency of the class preceding the modal class

f_2 = Frequency of the class succeeding the modal class

i = Class interval

Illustration : 39.

Compute the mode from the following series :

Size of item	Frequency
0–5	20
5–10	24
10–15	32
15–20	28
20–25	20
25–30	16
30–35	34
35–40	10
40–45	8

Solution :

Size of item	Frequency 1	2	3	4	5	6
0–5	20	44				
5–10	24		56	76		
10–15	32	60			84	
15–20	28		48			80
20–25	20	36		64		
25–30	16		50		70	
30–35	34	44				60
35–40	10		18	52		
40–45	8					

Alternative formula :

$$Z = L_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

L_1 = the real lower limit of the modal class; i = size of the class interval of the modal class; $\Delta_1 = f_1 - f_0$; $\Delta_2 = f_1 - f_2$.

The above formula takes the following form when elaborated :

$$Z = L_1 + \frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)} i = L_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} i$$

Analysis table

Column		Grouping containing maximum frequency				
1						30–35
2			10–15	15–20		
3		5–10	10–15			
4	0–5	5–10	10–15			
5		5–10	10–15	15–20		
6			10–15	15–20	20–25	
No. of times each occurs	1	3	5	3	1	

From the above analysis table, we find that the mode is 10–15 as its frequencies occur the maximum times. Mode is estimated by the formula :

$$Z = L_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$L_1 = 10 ; f_1 = 32 ; f_0 = 24 ; f_2 = 28 ; i = 5$$

$$Z = 10 + \frac{32 - 24}{2 \times 32 - 24 - 28} \times 5$$

$$= 10 + \frac{40}{12}$$

$$= 10 + 3.33$$

$$= 13.33$$

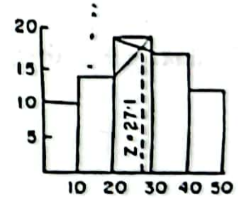
Graphic location of mode

Mode can be located graphically. The steps involved in this process are :

1. Draw a histogram of the given distribution.
2. The highest rectangle will be the modal class.
3. Join the top right and left corners of this rectangle with the top right corner of the rectangle representing the preceding class and the top left corner of the rectangle representing the succeeding class respectively.
4. From the point of intersection of both the lines draw a perpendicular on the X-axis ; the point where the perpendicular meets the X-axis is the value of the mode.

Illustration : 40

Income	Persons
0–10	10
10–20	14
20–30	19
30–40	17
40–50	13

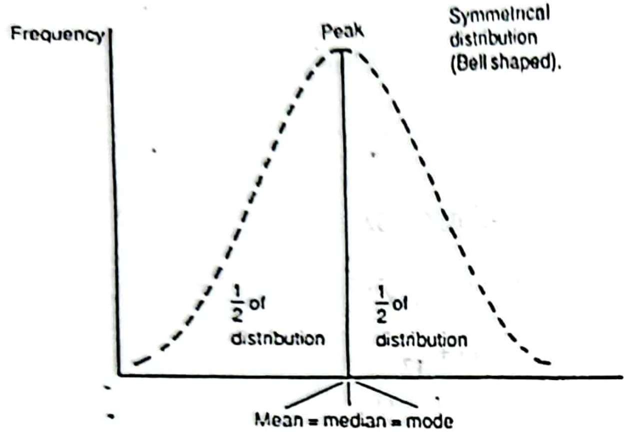


Mode = 27·1

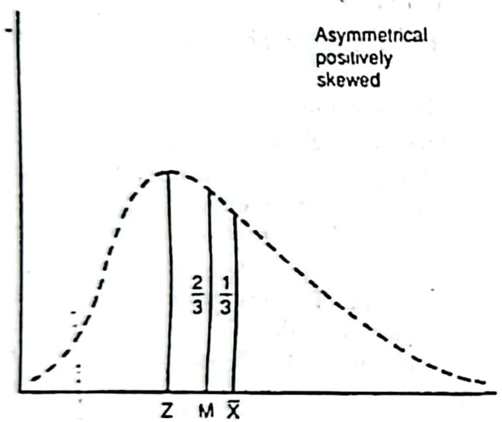
The mode is 27.1.

Relationship between different averages

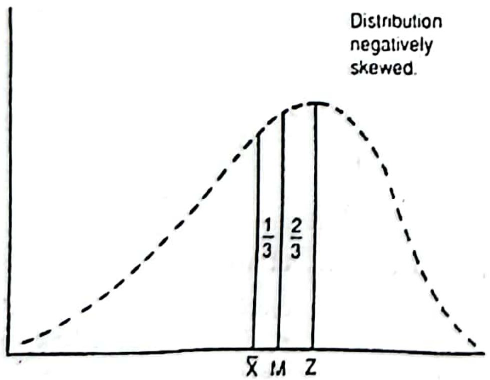
In a symmetrical distribution, Mean ($\bar{X}$), Median (Med) and Mode (Z) will coincide, *i.e.*, Mean = Median = Mode. In an asymmetrical (skewed) distribution, these values will be different.



Asymmetrical positively skewed.



Distribution negatively skewed.



When the distribution is moderately skewed and has greater concentration in the lower values, $\bar{X}$ > Med > Z (Mean > Median > Mode) it means the distribution is positively skewed (skewed to the right).

If the distribution concentrated in higher values, the tail is towards the lower values, then it is negatively skewed. In such cases, Z > Med > $\bar{X}$ (Mode > Median > Mean). As such there is a relationship in a moderately asymmetrical distribution between $\bar{X}$, Med and Z. It is clear from the following : (See page No. 157)

In a moderately asymmetrical distribution the difference between $\bar{X}$ and Z is three times of the differences between $\bar{X}$ and Med. Symbolically (Karl Pearson).

Mean–Median = 1/3 (Mean–Mode)

Mode = 3 Median–2 Mean

Median = Mode + 2/3 (Mean–Mode)

Mode = Mean–3 (Mean–Mode)

Mean–Mode = 3 (Mean–Median).

It follows, that if any two values out of the three are given, the third value can be estimated by applying the above formula.

Illustration : 41

If in a moderately asymmetrical frequency distribution, the values of median and arithmetic mean are 72 and 78 respectively, estimate the value of the mode. (*B.Com.Andhra*)

Solution :

The value of the mode is estimated by applying the following formula :

Mean–Mode = 3 (Mean–Median)

78–Mode = 3 (78–72)

78–Mode = 3 (6)

78–Mode = 18

–Mode = 18–78

–Mode = –60

60 = Mode

Therefore the mode is 60.

Merits :

1. It is easy to understand as well as easy to calculate. In certain cases, it can be found out by inspection.
2. It is usually an actual value as it occurs most frequently in the series.
3. It is not affected by extreme values as in the average.

Smallest size of the collar

$$= 12{\cdot}07 + 0{\cdot}75 = 12{\cdot}82''$$

Hence, the manufacturer should produce collars within the range 12·82″ to 17·14″.

Problem 24. Two cricketers scored the following runs in the several innings. Find who is a better run-getter and who is more consistent players.

A :	42,	17,	83,	59,	72,	76,	64,	45,	40,	32
B :	28,	70,	31,	0,	59,	108,	82,	14,	3,	95

Solution. In order to find out who is better run-getter we will compare the average runs scored and to find out who is more consistent, we will compare the coefficient of variation.

Calculation of Mean and Coefficient of Variation

	Cricketer 'A'			*Cricketer 'B'*	
X	$(X-\bar{X})$ x^2	$(X-\bar{X})^2$ x^2	X	$(X-\bar{X})$ x	$X-\bar{X})^2$ x^2
42	−11	121	28	−21	441
17	−36	1296	70	+21	441
83	+30	900	31	−18	324
59	+ 6	36	0	−49	2401
72	+19	361	59	+10	100
76	+23	529	108	+59	3481
64	+11	121	82	+33	1089
45	− 8	64	14	−35	1225
40	−13	169	3	−46	2116
32	−21	441	95	+46	2116
$\Sigma=530$	$\Sigma x=0$	$\Sigma x^2=4038$	$\Sigma X=490$	$\Sigma x=0$	$\Sigma x^2=13734$

Cricketer 'A' $\quad \bar{X}=\frac{\Sigma X}{N}=\frac{530}{10}=53$ runs

$$\sigma=\sqrt{\frac{\Sigma x^2}{N}}=\sqrt{\frac{4038}{10}}=20{\cdot}09$$

$$C.V.=\frac{\sigma}{\bar{X}}\times 100=\frac{20{\cdot}09}{53}\times 100=37{\cdot}92$$

Cricketer 'B' $\quad \bar{X}=\frac{\Sigma X}{N}=\frac{490}{10}=49$

$$\sigma=\sqrt{\frac{\Sigma x^2}{N}}=\sqrt{\frac{13734}{10}}=37{\cdot}06$$

$$C.V.=\frac{\sigma}{\bar{X}}\times 100=\frac{37{\cdot}06}{49}\times 100=75{\cdot}63$$

$$A=140,\ \Sigma fd=-5,\ N=20,\ i=10$$

$$\therefore \quad \bar{X}=140-\frac{5}{20}\times 10=140-2\cdot5=137\cdot5 \text{ lbs.}$$

$$\therefore \quad \sigma=\sqrt{\frac{\Sigma fd^2}{N}-\left(\frac{\Sigma fd}{N}\right)^2}\times i=\sqrt{\frac{37}{20}-\left(\frac{-5}{20}\right)^2}\times 10$$

$$=\sqrt{1\cdot85-\cdot0625}\times 10=\sqrt{1\cdot7875}\times 10=1\cdot34\times 10=13\cdot4$$

$$\text{C.V.}=\frac{13\cdot4}{137\cdot5}\times 100=9\cdot75$$

Since coefficient of variation is greater in case of weight hence there is greater variation in weight as compared to height.

Problem 26. Lives of two models of refrigerators in a recent survey are :

Life (*No. of years*)	*Number of Refrigerators* *Model A*	*Model B*
0— 2	5	2
2— 4	16	7
4— 6	13	12
6— 8	7	19
8—10	5	9
10—12	4	1

What is the average life of each model of these refrigerators ? which model has greater uniformity ?

(*B. Com., Bangalore Univ., 1981 ; MBA, Delhi Univ., 1984 ; Diploma in Mgt., AIMA, 1987*)

Solution. For finding out the average life we will compute arithmetic mean and for determining the model which has greater uniformity, we will compare the coefficient of variation.

Model A : Calculation of Mean and Coefficient of Variation

Life (*No. of Years*)	f	*m.p.* m	$(m-5)/2$ d	fd	fd^2
0— 2	5	1	−2	−10	20
2— 4	16	3	−1	−16	16
4— 6	13	5	0	0	0
6— 8	7	7	+1	+ 7	7
8—10	5	9	+2	+10	20
10—12	4	11	+3	+12	36
	$N=50$			$\Sigma fd=+3$	$\Sigma fd^2=99$

$$\bar{X}=A+\frac{\Sigma fd}{N}\times i$$

$$A=5,\ \Sigma fd=3,\ N=50,\ i=2$$

$$\therefore \quad \bar{X}=5+\frac{3}{50}\times 2=5+0\cdot12=5\cdot12$$

Thus average life of model $A=5\cdot12$ years

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

$$\Sigma fd^2=99,\ \Sigma fd=3,\ N=50,\ i=2$$

$$\therefore \quad \sigma=\sqrt{\frac{99}{50}-\left(\frac{3}{50}\right)^2}\times 2=\sqrt{1\cdot98-0\cdot0036}\times 2$$

$$=\sqrt{1\cdot9764}\times 2=1\cdot406\times 2=2\cdot812$$

$$\text{C.V.}=\frac{\sigma}{\bar{X}}\times 100=\frac{2\cdot812}{5\cdot12}\times 100=54\cdot92\%$$

Model B : Calculation of Mean and Coefficient of Variation

Life (*No. of hours*)	f	*m. p.* m	$(m-5)/2$ d	fd	fd^2
0— 2	2	1	−2	− 4	8
2— 4	7	3	−1	− 7	7
4— 6	12	5	0	0	0
6— 8	19	7	+1	+19	19
8—10	9	9	+2	+18	36
10—12	1	11	+3	+ 3	9
	$N=50$			$\Sigma fd=29$	$\Sigma fd^2=79$

$$\bar{X}=A+\frac{\Sigma fd}{N}\times i$$

$$A=5,\ \Sigma fd=29,\ N=50,\ i=2$$

$$\bar{X}=5+\frac{29}{50}\times 2=5+1\cdot16=6\cdot16$$

The average life of model $B=6\cdot16$

$$\sigma=\sqrt{\frac{\Sigma fd^2}{N}-\left(\frac{\Sigma fd}{N}\right)^2}\times i=\sqrt{\frac{79}{50}-\left(\frac{29}{50}\right)^2}\times 2$$

$$=\sqrt{1\cdot58-\cdot3364}\times 2=\sqrt{1\cdot2436}\times 2=1\cdot115\times 2=2\cdot23$$

$$\text{C.V.}=\frac{\sigma}{\bar{X}}\times 100=\frac{2\cdot23}{6\cdot16}\times 100=36\cdot2.$$

Since coefficient of variation is less for model B, hence refrigerators of model B show greater uniformity.

Problem 27. A factory produces two types of lamps. In an experiment in the working life of these lamps the following results were obtained :

$$\sigma=\sqrt{\frac{\Sigma fd^2}{N}-\left(\frac{\Sigma fd}{N}\right)^2}\times i=\sqrt{\frac{292}{174}-\left(\frac{-40}{174}\right)^2}\times 5$$

$$=\sqrt{1{\cdot}678-0{\cdot}053}=1{\cdot}275\times 5=6{\cdot}375$$

$$C.V.=\frac{\sigma}{\bar{X}}\times 100=\frac{6{\cdot}375}{21{\cdot}35}\times 100=29{\cdot}86$$

Problem 34. Mean and standard deviations of 100 items are found to be 40 and 10. If at the time of calculations two items are wrongly taken as 30 and 70 instead of 3 and 27, find the correct mean and standard deviation.

Solution. **Calculation of correct mean and standard deviation.**

Correct Mean. $\bar{X}=\frac{\Sigma X}{N}$

$\Sigma X=N\bar{X}$

$N=100,\quad \bar{X}=40$

$\therefore\quad \Sigma X=100\times 40=4000$

But this is not correct ΣX

Correct ΣX = Incorrect ΣX − wrong items + correct items

$=4000-30-70+3+27=3930$

Correct Mean $=\frac{\text{Correct }\Sigma X}{N}=\frac{3930}{100}=39{\cdot}3$

Correct Standard deviation : $\sigma^2=\frac{\Sigma X^2}{N}-(\bar{X})^2$

$\sigma=10,\quad N=100,\quad \bar{X}=40$

$$100=\frac{\Sigma X^2}{100}-(40)^2$$

$10000=\Sigma X^2-1{,}600\times 100$

$\Sigma X^2=1{,}60{,}000+10{,}000=1{,}70{,}000$

But this is not correct ΣX^2

Correct ΣX^2 = Incorrect ΣX^2 − Square of wrong items + Square of correct items

$=1{,}70{,}000-(30)^2-(70)^2+(3)^2+(27)^2$

$=1{,}70{,}000-900-4{,}900+9+729=1{,}64{,}938$

Correct $\sigma=\sqrt{\frac{\text{correct }\Sigma X^2}{N}-(\text{correct }\bar{X})^2}$

$$=\sqrt{\frac{164938}{100}-(39{\cdot}3)^2}=\sqrt{1649{\cdot}38-1544{\cdot}49}$$

$$=\sqrt{104{\cdot}89}=10{\cdot}24$$

Thus the correct mean is 39·3 and the correct standard deviation is 10·24.

Problem 35. Following are the marks obtained by two student A and B in 10 tests of 100 marks each :

Tests	1	2	3	4	5	6	7	8	9	10
Marks obtained by A	44	80	76	48	52	72	68	56	60	54
Marks obtained by B	48	75	54	60	63	69	72	51	57	66

If the consistency of performance is the criterion for awarding a prize, who should get the prize ?

(M.A., Kurukshetra Univ., 1986)

Solution. **Computation of coefficient of variation**

	Marks obtained by A			*Marks obtained by B*		
Tests	X	$(X-\bar{X})$ $\bar{X}=61$	$(X-\bar{X})^2$ x^2	X	$(X-61)$ d	d^2
1	44	−17	289	48	−13	169
2	80	+19	361	75	+14	196
3	76	+15	225	54	− 7	49
4	48	−13	169	60	− 1	1
5	52	− 9	81	63	+ 2	4
6	72	+11	121	69	+ 8	64
7	68	+ 7	49	72	+11	121
8	56	− 5	25	51	−10	100
9	60	− 1	1	57	− 4	16
10	54	+ 7	49	66	+ 5	25
	$\Sigma X=610$	$\Sigma(X-\bar{X})=0$	$\Sigma x^2=1370$	$\Sigma X=615$	$\Sigma d=5$	$\Sigma d^2=745$

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

$$\bar{X}=\frac{\Sigma X}{N}=\frac{610}{10}=61$$

$$\sigma=\sqrt{\frac{\Sigma x^2}{N}}=\sqrt{\frac{1370}{10}}=11{\cdot}7$$

$$C.V.=\frac{11{\cdot}7}{61}\times 100=19{\cdot}18$$

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

$$\bar{X}=\frac{\Sigma X}{N}=\frac{615}{10}=61{\cdot}5$$

$$\sigma=\sqrt{\frac{\Sigma d^2}{N}-\left(\frac{\Sigma d}{N}\right)^2}=\sqrt{\frac{745}{10}-\left(\frac{5}{10}\right)^2}$$

$$=\sqrt{74{\cdot}5-{\cdot}25}=8{\cdot}62$$

$$C.V.=\frac{8{\cdot}62}{61{\cdot}5}\times 100=14{\cdot}02$$

Since coefficient of variation is less for student B, hence student B should get the prize.

Problem 36. Find the actual class groups from the data given below :

d :	−3	−2	−1	0	1	2	3
f :	10	15	25	25	10	10	5

You are given that the mean of the distribution is 31 and standard deviation 15·9. (*M. Com., Jammu Univ., 1986*)

Solution. In order to ascertain the class groups, we need two values—the class interval and the assumed mean. From the formula for finding out standard deviation we can determine the class interval and from the formula for calculating mean we can ascertain the value of assumed mean.

Computations for Determining Class Groups

d	f	fd	fd^2
−3	10	−30	90
−2	15	−30	60
−1	25	−25	25
0	25	0	0
+1	10	+10	10
+2	10	+20	40
+3	5	+15	45
	$N=100$	$\Sigma fd=-40$	$\Sigma fd^2=270$

$$\sigma=\sqrt{\frac{\Sigma fd^2}{N}-\left(\frac{\Sigma fd}{N}\right)^2}\times i=\sqrt{\frac{270}{100}-\left(\frac{-40}{100}\right)^2}\times i$$

$$\sigma=\sqrt{2{\cdot}7-{\cdot}16}\times i=\sqrt{2{\cdot}54}\times i$$

But $\sigma=15{\cdot}9$ (given)

$\therefore$ $15{\cdot}9=1{\cdot}59\ i$

$$i=\frac{15{\cdot}9}{1{\cdot}59}=10$$

$$\bar{X}=A+\frac{\Sigma fd}{N}\times i$$

$\bar{X}=31,\ \Sigma fd=-40,\ N=100,\ i=10.$

We can find out A

$$31=A-\frac{40}{100}\times 10$$

$\therefore$ $A=31+4=35$

Hence the mid value of the group from which deviations have been calculated is 35. Since the common factor is 10 the lower and upper limits of this class would be 30—40. This class would correspond to 0 in the question given. Other classes can be easily ascertained and will be as follows :

0−10, 10−20, 20—30, 30—40, 40−50, 50—60, 60—70

Problem 37. A group has $\bar{X}=10$, $N=60$, $\sigma^2=4$. A sub-group of this has $\bar{X}_1=11$, $N_1=40$, $\sigma_1^2=2{\cdot}25$. Find the mean and standard deviation of the other sub-group.

Solution. We are given combined mean=10 and combined standard deviation=2.

Also $N=60,\ N_1=40,\ \bar{X}_1=11$

$N_2=(N-N_1)=(60-40)=20$

$$\bar{X}_{12}=\frac{N_1\bar{X}_1+N_2\bar{X}_2}{N_1+N_2}$$

Substituting the values

$$10=\frac{(40\times 11)+20\bar{X}_2}{60}$$

$600=440+20\ \bar{X}_2$

$\therefore$ $20\bar{X}_2=160$

$\bar{X}_2=8$

Thus the mean of the second group is 8.

$$\sigma_{12}=\sqrt{\frac{N_1\sigma_1^2+N_2\sigma_2^2+N_1d_1^2+N_2d_2^2}{N_1+N_2}}$$

$$=\sqrt{\frac{40\times(2{\cdot}25)+20\sigma_2^2+N_1(11-10)^2+N_2(8-10)^2}{60}}$$

$$=\sqrt{\frac{(40\times 2{\cdot}25)+20\sigma_2^2+(40\times 1)+(20\times 4)}{60}}$$

$$=\sqrt{\frac{90+20\sigma_2^2+40+80}{60}}$$

$\sigma_{12}^2=4$ given

$\therefore$ $4=\dfrac{210+20\sigma^2}{60}$

$\therefore$ $240=210+20\sigma^2$

$$20\sigma^2=240-210$$
$$20\sigma^2=30$$
$$\sigma^2=\frac{30}{20}=1{\cdot}5$$
$$\therefore \quad \sigma=\sqrt{1{\cdot}5}=1{\cdot}225$$

Problem 38. The mean of 5 observations is 4·4 and the variate is 8·24. If three of the five observations are 1, 2 and 6, find the other two.

Solution. $\bar{X}=\frac{\Sigma X}{N}$

$\therefore \quad \Sigma X=N\bar{X}$

Hence $\quad N=5,\ \bar{X}=4{\cdot}4$

$$\Sigma X=5\times 4{\cdot}4=22$$

Let the two missing items be x_1 and x_2

$$\therefore \quad 1+2+6+x_1+x_2=22$$

or $\quad x_1+x_2=22-9$

or $\quad x_1+x_2=13$

$$\sigma^2=\frac{\Sigma X^2}{N}-(\bar{X})^2$$
$$8{\cdot}24=\frac{\Sigma X^2}{N}-(4{\cdot}4)^2$$
$$41{\cdot}2=\Sigma X^2-91{\cdot}36\times 5$$
$$\Sigma X^2=96{\cdot}80+41{\cdot}2$$
$$\therefore \quad \Sigma X^2=138$$
$$\Sigma X^2=x_1^2+x_2^2+1^2+2^2+6^2$$
$$\Sigma X^2=x_1^2+x_2^2+41$$
$$x_1^2+x_2^2=138-41$$
$$x_1^2+x_2^2=97$$
$$(x_1+x_2)^2=x_1^2+x_2^2+2x_1x_2$$
$$(13)^2=97+2x_1x_2$$
$$169=97+2x_1x_2$$
$$2x_1x_2=169-97$$
$$x_1x_2=36$$
$$x_1+x_2=13 \qquad \ldots(i)$$
$$(x_1-x_2)^2=x_1^2+x_2^2-2x_1x_2$$
$$(x_1-x_2)^2=97-2(36)$$
$$(x_1-x_2)^2=97-72$$
$$(x_1-x_2)^2=25$$
$$x_1-x_2=5$$

$$x_1+x_2=13 \qquad \ldots(ii)$$
$$x_1-x_2=5 \qquad \text{from}\ldots(i)$$
$$2x_1=18 \qquad \ldots(ii)$$
$$\therefore \quad x_1=9$$

$\therefore$ Putting the value of x_1 in equation (*i*)

$$9+x_2=13$$
$$\therefore \quad x_2=4$$

Thus the two missing values are $x_1=9,\ x_2=4$.

Problem 39. (*a*) Mean of 200 items is 80 and their standard deviation is 10. Find the sum of the items and also the sum of squares of all the items.

Solution. $\bar{X}=\frac{\Sigma X}{N}$

$$N\bar{X}=\Sigma X$$
$$N=200,\ \bar{X}=80$$

Sum of the items

or $\quad \Sigma X=200\times 80=16{,}000$

$$\sigma=\sqrt{\frac{\Sigma X^2}{N}-(\bar{X})^2}=10=\sqrt{\frac{\Sigma X^2}{200}-(80)^2}$$

Squaring both sides :

$$100=\frac{\Sigma X^2}{200}-6{,}400$$
$$20{,}000=\Sigma X^2-12{,}80{,}000$$
$$\Sigma X^2=12{,}80{,}000+20{,}000=13{,}00{,}000$$

(*b*) Coefficient of variation of two series are 60% and 80%. Their standard deviations are 20 and 16. What are their arithmetic means ? (*B. Com., Bangalore Univ., 1986*)

Solution. Series A : $C.V.=\frac{\sigma}{\bar{X}}\times 100$

Given $\quad C.V.=60,\ \sigma=20$

$$\because \quad 60=\frac{20}{\bar{X}}\times 100$$
$$60\bar{X}=2000$$
$$\bar{X}=\frac{2000}{60}=33{\cdot}33$$

Series B $\quad C.V.=80,\ \sigma=16$

$$80=\frac{16}{\bar{X}}\times 100$$
$$80\bar{X}=1600 \text{ or } \bar{X}=20$$

Thus the arithmetic mean of the two series are 33·33 and 20 respectively.

(c) For a distribution, the co-efficient of variation is 22·5% and the value of arithmetic average is 7·5. Find out the value of Standard Deviation. (B. Com., Delhi Univ. 1983)

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

$\bar{X} = 7{\cdot}5,\ C.V. = 22{\cdot}5$

$22{\cdot}5 = \frac{\sigma}{7{\cdot}5} \times 100$

$100\sigma = 22{\cdot}5 \times 7{\cdot}5 = 168{\cdot}75$

$\sigma = \frac{168{\cdot}75}{100} = 1{\cdot}6875$

Problem 40. The first of two sub-groups has 100 items with mean 15 and standard deviation 3. If the whole group has 250 items with mean 15·6 and standard deviation $\sqrt{13{\cdot}44}$, find the standard deviation of the second sub-group.

Solution. $N_1 = 100,\ \bar{X}_1 = 15,\ \sigma_1 = 3$

$N_1 + N_2 = 250,\ \bar{X}_{12} = 15{\cdot}6,\ \sigma_{12} = \sqrt{13{\cdot}44}$

Since $N_1 + N_2 = 250,\ N_2 = (250 - 100) = 150$

$$\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

$$15{\cdot}6 = \frac{100 \times 15 + 150\,\bar{X}_2}{250}$$

$3900 = 1500 + 150\,\bar{X}_2$

$150\,\bar{X}_2 = 2400$ or $\bar{X}_2 = 16$

or

$$\sigma_{12}^2 = \frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}$$

$d_1 = (\bar{X}_{12} - \bar{X}_1) - (15{\cdot}6 - 15) = {\cdot}6$

$d_2 = (\bar{X}_{12} - \bar{X}_2) = (15{\cdot}6 - 16) = -{\cdot}4$

$$13{\cdot}44 = \frac{100(3)^2 + 150\sigma_2^2 + 100({\cdot}6) + 150(-{\cdot}4)^2}{250}$$

$3360 = 900 + 150\sigma_2^2 + 36 + 24$

$150\sigma^2 = 3360 - 960$

$\sigma_2^2 = 16$ or $\sigma_2 = 4$

The standard deviation of the second sub-group is 4.

Problem 41. *Find the standard deviation of the first n natural numbers 1, 2, 3...n.*

Solution. If X denotes items, standard deviation is given by the formula :

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

$$\Sigma X = 1 + 2 + 3 + n \ldots = \frac{n(n+1)}{2}$$

(Sum of first n natural numbers)

$$\Sigma X^2 = 1^2 + 2^2 + 3^2 + \ldots n^2 = \frac{n(n+1)(2n+1)}{6}$$

(Sum of the squares of n natural numbers)

$$\therefore \quad \frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2 = \frac{n(n+1)(2n-1)}{6n} - \left[\frac{n(n+1)}{2n}\right]^2$$

$$= \frac{(n+1)(2n+1}{6} - \frac{(n+1)^2}{4}$$

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

$$= \frac{n+1}{2}\left[\frac{4n+2-3n-3}{12}\right] = \frac{(n+1)(n-1)}{12}$$

$$\sigma = \sqrt{\frac{n^2-1}{12}}$$

Problem 42. (*i*) Calculate range, standard deviation and coefficient of variation in respect of the marks obtained by 10 students given below :

50, 55, 57, 49, 54, 61, 64, 59, 58, 56,

(*ii*) How would your results be affected if it is decided to increase the marks of each of the above students by 5.

Solution. Range $= L - S$

$L = 64,\ S = 49$

$\therefore$ Range $= 64 - 49 = 15$

Computation of Standard Deviation

X	$(X-56)$ d	d^2
50	−6	36
55	−1	1
57	+1	1
49	−7	49
54	−2	4
61	+5	25
64	+8	64
59	+3	9
58	+2	4
56	0	0
$\Sigma X = 563$	$\Sigma d = 3$	$\Sigma d^2 = 193$

$$=\sqrt{2.793-.947}\times 2=1.359\times 2=2.72$$

$$\bar{X}=A+\frac{\Sigma fd}{N}\times i=19.5-\frac{2929}{3010}\times 2=19.5-1.95=17.55$$

$$\bar{X}\pm 3\sigma=17.55\pm 3(2.72)=9.39-25.71.$$

Problem 46. You are provided with the following raw sums in a statistical investigation of two variables, X and Y:

$$\Sigma X=235,\ \Sigma Y=250,\ \Sigma X^2=6750,\ \Sigma Y^2=6840$$

Ten pairs of values are included in the survey. Compute the standard deviation of the X and Y variables.

Solution. $\sigma_x=\sqrt{\frac{\Sigma X^2}{N}-(\bar{X})^2}$

$\Sigma X^2=6750$ (given)

$$\bar{X}=\frac{\Sigma X}{N}=\frac{235}{10}=23.5$$

$$\therefore \quad \sigma_x=\sqrt{\frac{6750}{10}-(23.5)^2}$$

$$=\sqrt{675-552.25}=\sqrt{122.75}=11.1$$

$$\sigma_y=\sqrt{\frac{\Sigma Y^2}{N}-(\bar{Y})^2}$$

$\Sigma Y^2=6840$ given

$$\bar{Y}=\frac{\Sigma Y}{N}=\frac{250}{10}=25$$

$$\sigma_y=\sqrt{\frac{6840}{10}-(25)^2}\ \sqrt{684-625}=\sqrt{59}=7.68$$

Problem 47. The following table gives the length of life of 400 radio tubes :

Length of life (hours)	*No. of radio tubes*	*Length of life (hours*	*No. of radio tubes*
300—399	12	800—899	60
400—499	32	900—999	32
500—599	64	1000—1099	26
600—699	76	1100—1199	10
700—799	88		

Calculate (*i*) The average length of life of a radio tube
(*ii*) The standard deviation of the length

Solution. $S.D.=\frac{3}{2}Q.D.$ and $S.D.=\frac{5}{4}M.D.$

Hence $\frac{3}{2}\ Q.D.=\frac{5}{4}M.D.$

$$Q.D.=\frac{5}{4}\times\frac{2}{3}M.D=\frac{10}{12}M.D.$$

$$Q.D.=\frac{10}{12}\times 24=20$$

(*b*) For a group of observations the value of Quartile deviation is 20. What would be the most likely value of variance for that group?

Solution : $Q.D.=\frac{2}{3}S.D.$

or $\frac{3}{2}Q.D.=S.D.$

$$S.D.=\frac{3}{2}\times 20=30$$

Variance $=\sigma^2=(30)^2=900$

Problem 49. The mean and S.D. calculated from 20 observations are 15 and 10 respectively. If an additional observation 36, left out through oversight, be included in the calculations, find the correct mean and standard deviation.

Solution. $\bar{X}=\frac{\Sigma X}{N}$

$$N\bar{X}=\Sigma X$$

$$N=20,\ \bar{X}=15$$

$\therefore$ $\Sigma X=20\times 15=300.$ But this is incorrect

Correct $\Sigma X=300+36=336$

Correct $N=21$

$\therefore$ Correct $\bar{X}=\frac{336}{21}=16$

$$\sigma^2=\frac{\Sigma X^2}{N}-(\bar{X})^2$$

$$100=\frac{\Sigma X^2}{20}-(15)^2$$

$$2000=\Sigma X^2-4500$$

$$\Sigma X^2-4500=2000$$

$\therefore$ $\Sigma X^2 = 6500$

Correct $\Sigma X^2 = 6500 + (36)^2 = 7796$

Correct $\sigma^2 = \frac{\text{correct } \Sigma X^2}{N} - (\text{correct } \bar{X})^2$

$$= \frac{7796}{21} - (16)^2 = \frac{7796 - 5376}{21} = \frac{2420}{21}$$

Correct $\sigma = \sqrt{\frac{2420}{21}} = \sqrt{115 \cdot 24} = 10 \cdot 73$

Thus correct $\bar{X} = 16$ and $\sigma = 10 \cdot 73$.

Problem 50. Weights of eight students in kilograms are as follows. If the weighing machine shows weight less by two kilograms, find mean and coefficient of variation.

48, 53, 46, 56, 47, 52, 48, 50

(*B. Com., Kerala, 1987*)

Solution. Calculation of Mean and Coefficient of Variation

X	$(X-\bar{X})$ x	x^2
48	−2	4
53	+3	9
46	−4	16
56	+6	36
47	−3	9
52	+2	4
48	−2	4
50	0	0
$\Sigma X = 400$	$\Sigma x = 0$	$\Sigma x^2 = 82$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{400}{8} = 50$$

Correct $\bar{X} = 50 + 2 = 52$ kgs.

Since the weight of each student is recorded less by 2 kg., th arithmetic mean would tend to be increased by 2 kg.

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{82}{8}} = 3 \cdot 2$$

The standard deviation will not be affected when the weight is recorded less by 2 kg for every student. It is so because the standard deviation is independent of change of scale *i.e.*, the standard deviation of 10, 20, 30 will be the same as the standard deviation of 8, 18 and 28 or 12, 22 and 32. But the same constant must be added or deducted throughout the series.

Problem 51. The arithmetic mean and the standard deviation of a series of 20 items were calculated by a student as 20 cm. and 5 cm. respectively. But while calculating them an item 13 was misread as 30. Find the correct arithmetic mean and the correct standard deviation. (*B. Com., Mysore Univ., 1985*)

Solution.

Correct mean $N = 20$, $\bar{X} = 20$ and $\sigma = 5$

$$\bar{X} = \frac{\Sigma X}{N}$$

$$\Sigma X = N\bar{X}$$

or $20 \times 20 = 400$ (*i.e.*, ΣX)

But this is incorrect ΣX

Correct $\Sigma X = 400 - 30 + 13 = 383$

Correct Mean $= \frac{383}{20} = 19 \cdot 15$

Correct Standard Deviation $\sigma^2 = \frac{\Sigma X^2}{N} - (\bar{X})^2$

$$25 = \frac{\Sigma X^2}{20} - (20)^2$$

$$20 \times 25 = \Sigma X^2 - 8000$$

$$\Sigma X^2 = 8500$$

Correct $\Sigma X^2 = 8500 - (30)^2 + (13)^2 = 8500 - 900 + 169 = 7769$

Correct $\sigma = \sqrt{\frac{\text{correct } \Sigma X^2}{N} - (\text{correct } \bar{X})^2}$

$$= \sqrt{\frac{7769}{20} - (19 \cdot 15)^2} = \sqrt{388 \cdot 45 - 366 \cdot 72}$$

$$= \sqrt{21 \cdot 73} = 4 \cdot 66$$

Thus correct $\bar{X} = 19 \cdot 15$ and correct $\sigma = 4 \cdot 66$.

Problem 52. Calculate the standard deviation and the coefficient of variation of the data giving the kilowatt-hours of electricity used in one month by 75 residential consumers :

Consumption (kilowatt-hours)	*Numbers of Consumers*	*Consumption (kilowatt-hours)*	*Number of Consumers*
5—25	4	85—105	14
25—45	6	105—125	5
45—65	14	125—145	7
65—85	22	145—165	3

(*B. Com., Calcutta Univ., 1985*)