① Class → 01

Date: 30.09.23

Measures of Dispersion: → scatter, variation/spread

Dispersion gives us an extended idea how individual items vary from one another and from the central value.

One of the most important quantities used to characterize a frequency distribution.

There are many types of dispersion measures:

1) Range
2) Inter Quartile Range (IQR)
3) Mean Deviation (MD)
4) Variance / standard deviation
5) Coefficient of variation (CV)

Range: → For ungrouped data

Range is the difference between the largest and smallest observations.

Range = Largest value − Lowest value

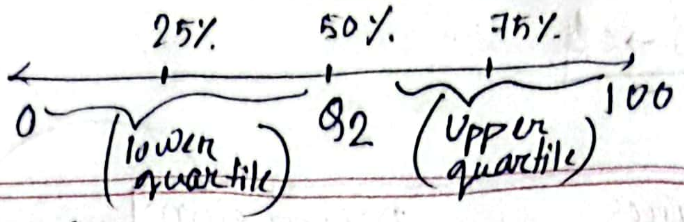
Example: For a set of observations,

90, 110, $\underset{X_n}{\boxed{20}}$ 51, $\underset{X_m}{\boxed{210}}$ and 100.

Range = 210 − 20 = 190 / L = Largest, S = Smallest
(L − S)

Co-efficient of Range / Coefficient of Dispersion:

$$\frac{L-S}{L+S} \times 100 \;/\; \frac{X_m - X_n}{X_m + X_n} \times 100 = \frac{210-200}{210+200} \times 100 = \frac{10 \times 10}{410} = 2.43$$

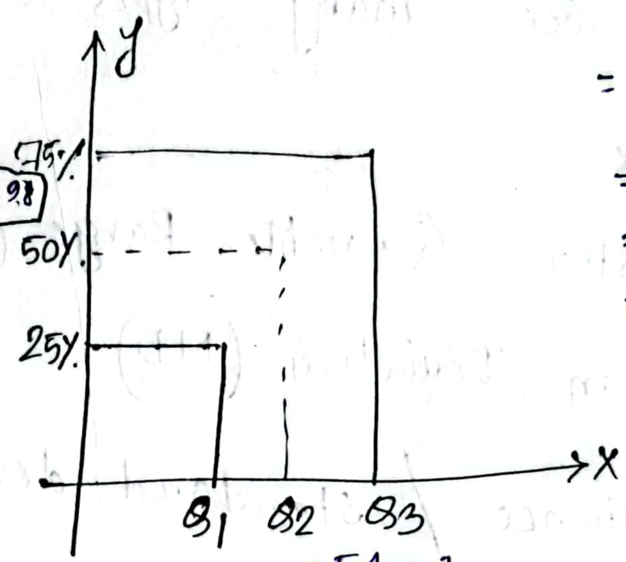


* Interquartile Range:

It is the difference between the third quartile and the first quartile.

$$IQR = Q_3 - Q_1$$

Q_1 = Lower quartile
$= \frac{1}{4}(n+1)$-th term
$= 1.75$ th
$= 1\text{st} + 0.75(2\text{nd} - 1\text{s..}$
$= 71 + 0.75(74 - 7$
$= 73.25$



Example:

$\boxed{71 \quad 74 \quad 85}\ \boxed{90 \quad 96 \quad 98}$

Q_2 = Median

$= \dfrac{\frac{6}{2}\text{th term} + \left(\frac{6}{2}+1\right)\text{th term}}{2}$

$= \dfrac{85+90}{2} = 87.5$

$Q_3 = \frac{3}{4}(n+1)$th
$= 5.25$ th
$= 5\text{th} + 0.25(6\text{th} - 5$
$= 96 + 0.25(98 -$
$= 96.5$

$\therefore IQR = Q_3 - Q_1 = 96.5 - 73.25 = 23.25$ [Ans.]

$x_1 \quad x_2 \quad x_3 \quad \boxed{x_4} \quad x_5 \quad x_6 \quad x_7$

x_4 → (Median)

x_1, x_2, x_3 (Lower Quarter) → Q_1

x_5, x_6, x_7 (Upper Quarter) → Q_3

Mean Deviation (MD) or ~~A~~ Mean Absolute deviation (MAD): [For grouped and ungrouped data]

$\frac{\Sigma|x-\mu|}{N}$

It is used to compute how far the values in a data set are from the center point.

MD about mean

$$MD / MAD = \frac{\Sigma|x-\mu|}{n} \text{ [Ungrouped] } [\mu = \text{mean}]$$

$$MD / MAD = \frac{\Sigma f_i |x_i - \mu|}{\Sigma f_i} \text{ [Grouped data]}$$

Example: 600 ~~mm~~, 470, 170, 430, 300

Find Mean deviation / Absolute MD.

Soln:

$$\mu = \bar{x} = \frac{\Sigma x}{n} = \frac{1970}{5} = 394$$

x	$\lvert x-\mu \rvert$
600	206
470	76
170	224
430	36
300	94

$$MD = \frac{\Sigma|x-\mu|}{n}$$

$$= \frac{636}{5} \approx 127.2$$

[Ans.]

Mean deviation about Mean:

$$MD(\bar{x}) = \frac{\Sigma |x_i - \bar{x}|}{n}$$

Mean deviation about Median:

$$MD(\tilde{m}) = \frac{\Sigma |x_i - \tilde{m}|}{n}$$

* Variance and standard deviation — uses: how dispersed / far each number from data set

Problem: Ten persons of varying ages were weighed and the following weights in kg were recorded:

(~~sample~~) 110, 125, 125, 147, 117, 125, 136, 157, 124, 110

(Population)

Compute mean deviation about mean, median.

110, 110, 117, 124, 125, 125, 125, 136, 147, 157.

Soln:

Serial No	Weight (x_i)	$\lvert x_i - \bar{x} \rvert$	$\lvert x_i - m \rvert$
1	110	17.6	11
2	125	2.6	4
3	125	2.6	4
4	147	19.4	26
5	117	10.6	4
6	125	2.6	4
7	136	8.4	15
8	157	29.4	36
9	124	3.6	3
10	110	17.6	11

Arithmatic Mean $= \frac{\Sigma x}{n} = \frac{1276}{10}$

$= 127.6$

Now, Median $= \frac{125 + 125}{2}$

$= 125$

Mean about mean Deviation, $M_d(\bar{x}) = \frac{\Sigma|x_i - \bar{x}|}{n} = \frac{114.4}{10}$

$= 11.44$ [check calculation]

and $M_d(\tilde{m}) = \frac{118}{10}$

(M.D. about median) $= 11.8$

[Ans.]

[Problem:] ~~(Population)~~ [Sample]

Class interval	Class marks
	5F
54-57	5
58-61	9
62-65	14
66-69	7
70-73	

Find MD about mean, median, mode, Variance and standard deviation.

Soln:

Originial class Interval	Frequency (f_i)	Cu. Frequency	Midpoint (x_i)	$f_i\lvert x_i-\bar{x}\rvert$	$f_i\lvert x_i-\tilde{m}\rvert$	$f_i\lvert x_i-\text{Mode}\rvert$
53.5 – 57.5	5	5	55.5	46.5	51.4	58.35
57.5 – 61.5	5	10	59.5	26.5	31.4	38.35
61.5 – 65.5	9	19	63.5	11.7	20.61	33.03
median class 65.5 – 69.5	14	33	67.5	37.8	23.94	4.62
69.5 – 73.5	7	40	71.5	46.9	39.97	30.31

Now,

$$\text{Mean}, \bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

$$= \frac{2592}{40}$$

$$\bar{x} = 64.8$$

$$\text{Median} = l_m + \frac{h}{f_m}\left(\frac{n}{2} - F_{(m)-1}\right)$$

$$\text{Median} = 65.5 + \frac{4}{14}\left(\frac{40}{2} - 19\right)$$

$$\tilde{m} = 65.78$$

$$\text{Mode} = \overset{\text{H.F.}}{l} + h\left(\frac{\Delta_1}{\Delta_1+\Delta_2}\right)$$

$$\text{Mode} = 65.5 + 4\left(\frac{5}{5+7}\right) \quad \begin{cases}\Delta_1 = 14-9=5\\ \Delta_2 = 14-7=7\end{cases}$$

$$= 67.17$$

Mean dev. about mean $= \frac{\sum f_i |x_i - \bar{x}|}{\sum f_i}$

$= \frac{169.4}{40}$

$= 4.235$

M.D. about Median $= \frac{\sum f_i |x_i - \bar{m}|}{\sum f_i}$

$= \frac{167.42}{40}$

$= 4.18$

and M.D. about Mode $= \frac{\sum f_i |x_i - \text{mode}|}{\sum f_i}$

$= \frac{164.66}{40}$

$= 4.11$

[Ans.]

Standard deviation, S.D. $= \sqrt{\dfrac{\sum f_i (x - \bar{x})^2}{n}}$

Variance, $\sigma^2 / s^2 / var(x) = \dfrac{\sum f_i (x_i - \bar{x})^2}{n}$ [n = Total frequency]

f_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$
5	-9.3	86.49	432.45
5	-5.3	28.09	140.45
9	-1.3	1.69	15.21
14	2.7	7.29	102.06
7	6.7	44.89	314.23
$\sum f_i = 40$			$\sum = 1004.4$

Now, $s^2 = \dfrac{1004.4}{40}$

$= \cancel{25.75}\ 25.11$

and standard deviation, $\sigma = \sqrt{25.11}$

$= 5.010987927$

[Ans.]

The higher the C.V., consistency of uniform, less consistent] data set gets lesser. [less stable, less
> measure of dispersion of data points around mean.

Coefficient of Variation:

The higher the CV, the greater the dispe

$$C.V. = \left(\frac{\text{Standard deviation}}{\text{Mean}}\right) \times 100 \quad [\text{For SD}]$$

$$C.V. = \left(\frac{\text{MD about Mean}}{\text{AM}}\right) \times 100 \quad [\text{For Mean D. about mean}]$$

$$C.V. = \left(\frac{\text{MD about Median}}{\text{Median}}\right) \times 100 \quad [\text{For MD about Median}]$$

$$C.V. = \left(\frac{\text{MD about Mode}}{\text{Mode}}\right) \times 100 \quad [\text{For MD abou Mode}]$$

Previous Problem, [If we want to find Coeff of MD ...]

$$\text{Co-efficient of MD (AM)} = \frac{4.235}{64.8} \times 100$$

$$= 6.53\%.$$

$$\text{Co-efficient of MD (Mode)} = \frac{4.1165}{67.17} \times 100$$

$$= 6.128\%.$$

$$\text{Co-efficient of MD (Median)} = \frac{4.18}{65.79} \times 100$$

$\frac{\sum f_i(x_i}{\sum f_i}$

Problem:

Find Mean, Variance and standard deviation from the following table.

Length	Frequency
20-22	3
23-25	6
26-28	12
29-31	9
32-34	2

Soln:

Length	O.C.I	Frequency (f_i)	x_i	$f_i x_i$	$\lvert x_i - \bar{x} \rvert$
20 – 22	19.5-22.5	3	21	63	6.09
23 – 25	22.5 – 25.5	6	24	144	3.09
26 – 28	25.5-28.5	12	27	324	0.09
29 – 31	28.5 - 31.5	9	30	270	2.91
32 - 34	31.5 – 34.5	2	33	66	5.91
		$\sum f_i = 32$			

Mean $= \frac{\sum f_i x_i}{\sum f_i}$

$= \frac{867}{32}$

$= 27.09$

$f(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
37.08	111.24
9.5481	57.28
0.0081	0.0972
8.4681	76.21
34.92	69.84

Variance, $s^2 = \dfrac{\sum f_i (x_i - \bar{x})^2}{n}$

$= \dfrac{314.6672}{32}$

$= 9.83$

and S.D., $\sigma = \sqrt{9.83}$

$= 3.13$

[Ans.]

(1)

$\boxed{\text{Problem:}}$ Calculate the Variance, SD, Co-efficient of variation from the following weight of 60 apples in grouped data:

Weight (gm)	Midpoints (x_i)	Frequency (f_i)	$f_i x_i$
65-84	74.5	09	670.5
85-104	94.5	10	945
105-124	114.5	17	1945.05
125-144	134.5	10	1345
145-164	154.5	05	~~1345.0~~ 772.5
165-184	174.5	04	~~972.5~~ 698
185-204	194.5	05	972.5
		$\sum f_i = 60$	$\sum f_i x_i = 7350$

$\sum f_i x_i / \sum f_i$

$\boxed{\bar{x} = 122.5}$ ⇒ Mean

$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
-48	2304	20736
-28	784	7840
-8	64	188
12	144	1440
32	1024	5120
52	2704	10816
72	5184	25920
		$\sum f_i(x_i - \bar{x})^2 = 72060$

Variance, $s^2 = \dfrac{\Sigma fi(x_i - \bar{x})^2}{n}$

$= \dfrac{72060}{60}$

$= 1201$

and S.D. $\sigma = \sqrt{1201} = 34.655$

co-efficient of variation:

$$C.V. = \frac{34.655}{\underbrace{122.5}_{(Mean)}} \times 100 = 28.289 \text{ [Ans.]}$$

[*] Standard deviation and variance for Ungrouped data: (Formula)

Variance, $s^2 = \dfrac{\Sigma(x_i - \bar{x})^2}{n}$

and S.D., $\sigma = \sqrt{\dfrac{\Sigma(x_i - \bar{x})^2}{n}}$

For Grouped data:

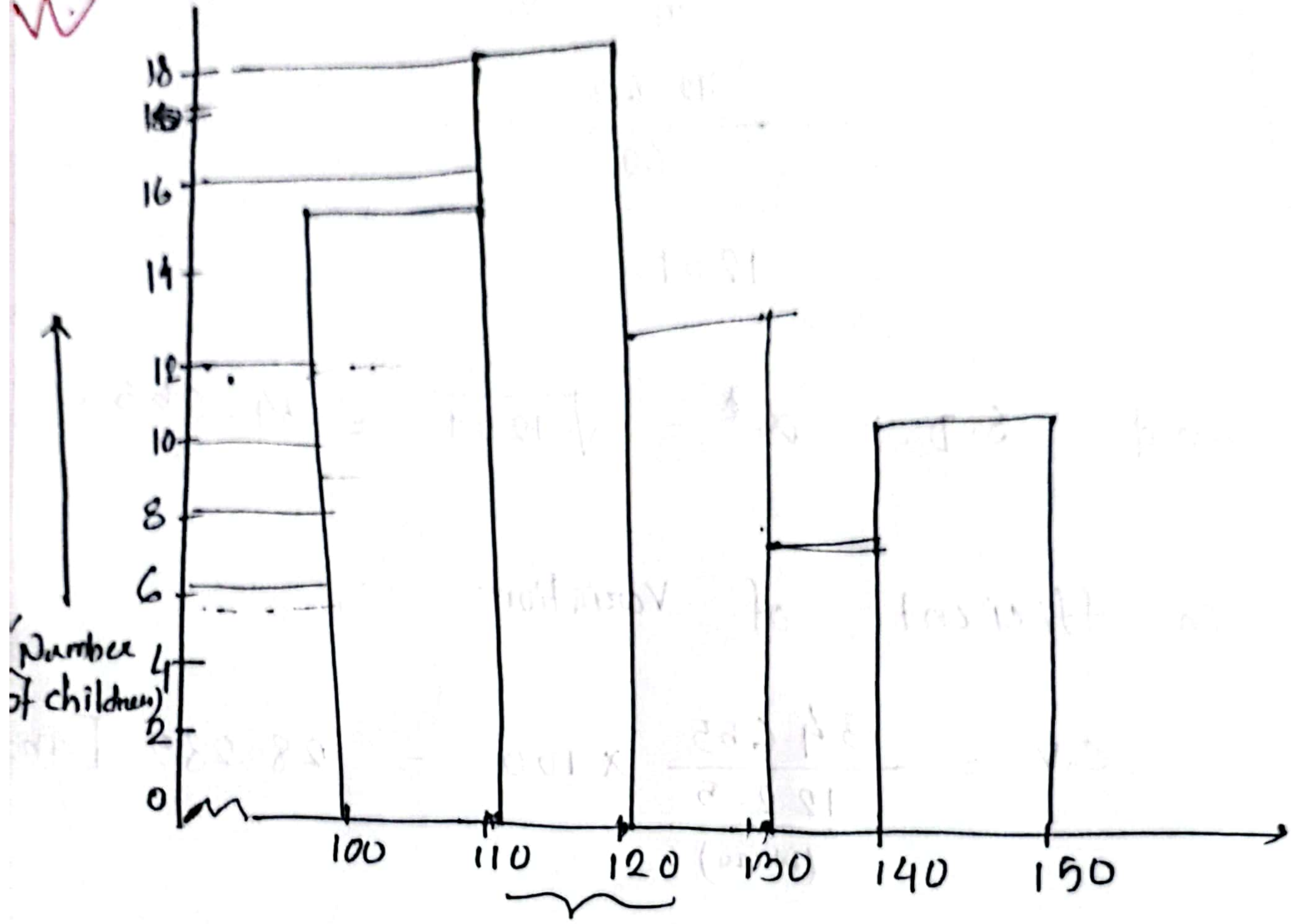
Variance, $s^2 = \dfrac{\Sigma fi\, x_i^2}{\Sigma fi} - \left(\dfrac{\Sigma fi\, x_i}{\Sigma fi}\right)^2$

OR, $\dfrac{\Sigma fi(x_i - \bar{x})^2}{\Sigma fi}$

S.D., $SD = \sqrt{\dfrac{\Sigma fi\, x_i^2}{\Sigma fi} - \dfrac{(\Sigma fi\, x_i)^2}{(\Sigma fi)^2}}$

(3)

Problem:



Frequency Distⁿ

. Standard Deviⁿ

a) find mode

b) Make f.D. Table and find harmonic mean.

c) find S.D. of the F.D.

d) sketch Cumulative Frequency Polygon and hence D_3 and P_{60}.

$\Delta_1 = 18 - 15 = 3$
$\Delta_2 = 18 - 11 = 7$

Soln. a) For the range 110-120 Heighest Frequency occurse. So, mode ~~class~~ class = 110-120

$l + h\left(\frac{\Delta_1}{\Delta_1+\Delta_2}\right)$
$= 110 + 10 \frac{3}{10}$
$= 113$ [Ans.]

b)

Class interval	Frequency (f)	$\frac{f}{x}$	Midpoint(x)
100 - 110	15	0.142	105
110 - 120	18	0.156	115
120 - 130	11	0.088	125
130 - 140	~~7~~ 5	~~0.051~~ 0.037	135
140 - 150	9	0.062	145
	$\Sigma f = 58$	$\Sigma \frac{f}{x} = 0.485$	

$f \to$ Frequency

Formula for Grouped data

$$\text{Harmonic mean} = \frac{\Sigma f}{\Sigma \frac{f}{x}} = \frac{58}{0.485} = 119.587$$ [Ans.]

~~Do the calculation~~

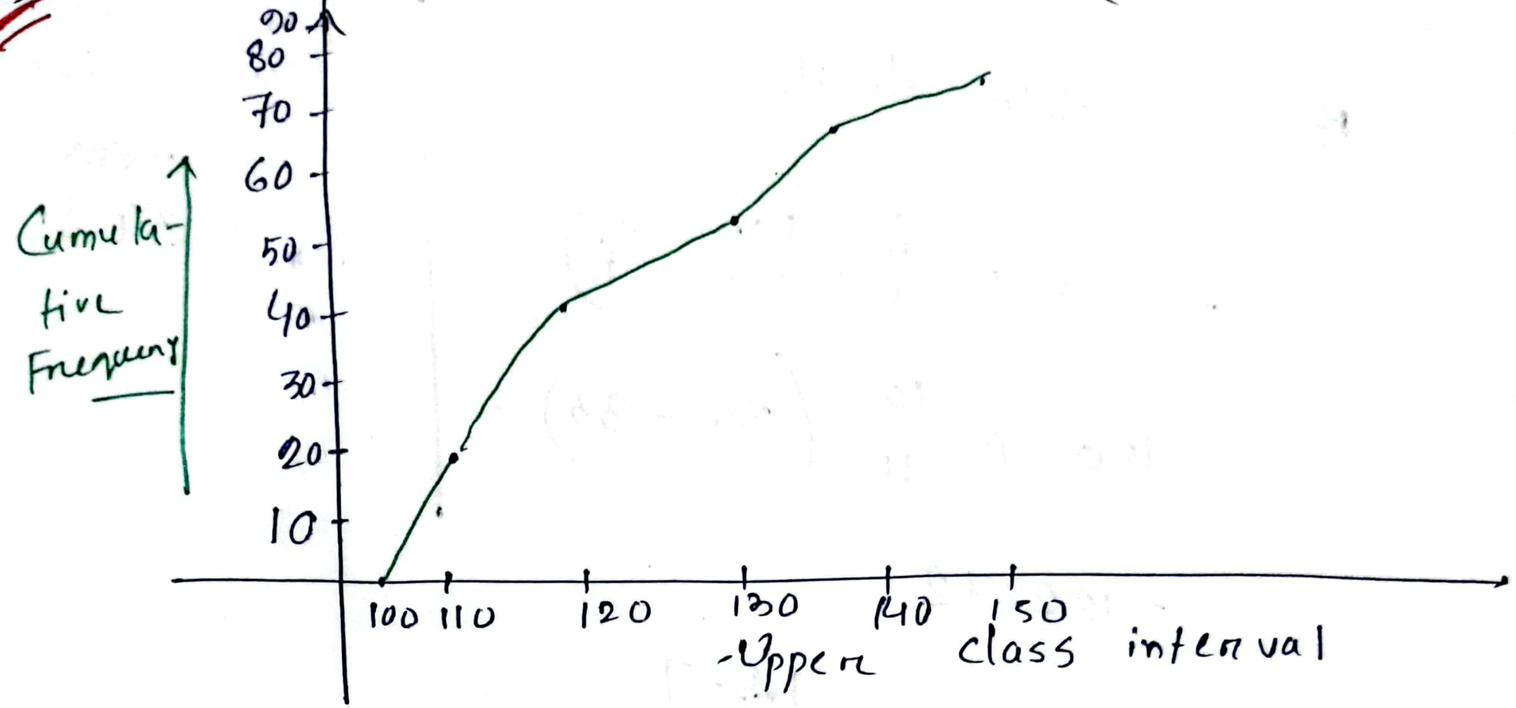
c) $S.D., \sigma = \sqrt{\frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i}}$

f_i = Frequency
x_i = midpoint
$\bar{x}$ = mean

Class Interval	Midpoint (x_i)	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i-\bar{x})^2$
100 - 110	105			
110 - 120	115			
120 - 130	125			
130 - 140	135			
140 - 150	145			

Just Do the calculation

d) Cumulative Frequency Polygon (Ogive Curve)



C.I.	Frequency	Cumu Frequency	Upper C.I.
100 - 110	15	15	110
110 - 120 (Q. class)	18	33	120
120 - 130 (Per. class)	11	44	130
130 - 140	5	49	140
140 - 150	9	58	150
	$\Sigma f = 58$		

D_3 = 3rd decile

$= l + \frac{h}{f}\left(\frac{kN}{10} - c_f\right)$

$= 110 + \frac{10}{18}(18 - 15)$

$= 111.67$

$\frac{kN}{10} = \frac{3 \times 58}{10}$

$= 17.4$

(see the previous page table)

P_{60} = 60th decile

$= l + \frac{h}{f}\left(\frac{kN}{100} - c_f\right)$

$= 120 + \frac{10}{11}(36 - 33)$

$= 122.72$

$\frac{kN}{100} = \frac{60 \times 58}{100}$

$= 34.8$

[Ans.]

(2) Divide the range by the average value.

from 110 to 130 cm, Range = 18 − 11

= 7

Averag value = $\frac{18+11}{2} = 14.5$

∴ Percentage = $\frac{7}{14.5} \times 100\%$

$= 48.27\%$ [Ans.]

Dispersion

Absolute Measure (calculate normal dispersion)

1) Range
2) Quartile deviation
3) Mean Deviation
4) Standard deviation
5) Variance

Relative measure (used to compare the distribution of two or more data sets)

1) Coefficient of Range
2) Coefficient of quartile deviation
3) Coefficient of S.D.
4) Coefficient of variance variation.

$\boxed{*}$ Range for grouped data:

Mass(kg)	Frequency(f)	O.C.I.
41 – 45	7	$\boxed{40.5}$ - 45.5
46 - 50	10	45.5 - 50.5
51 - 55	15	50.5 – 55.5
56 - 60	12	55.5 - 60.5
61 - 65	6	60.5 – $\boxed{65.5}$

Find the range and co-efficient of range.

Soln: Range = L - S

= Upper Limit of highest c.I - Lower Limit of Lowest class interval.

= 65.5 – 40.5 = 25

co-efficient of Range = $\frac{L-S}{L+S} \times 100$

$= \frac{65.5-40.5}{65.5+40.5} \times 100$

= 23.58%. [Ans.]

Date: 03.10.

(2)

Problem The following data was received by testing 60 bulbs of two diff companies:

Lenght of Life (in hours)	Company A	Company B
700 – 900	18	6
900 – 1100	16	42
1100 – 1300	26	12

Calculata S.D. and coeff of variation also state which company bulb are more uniform?

Soln: For company A,

C.I.	x_i	f_i	$f_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
700 – 900	800	18	14400	-226.67	51379.28	924827.04
900 – 1100	1000	16	16000	-26.67	711.288	113804
1100 – 1300	1200	26	31200	173.33	30043.28	781125.

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

$$= 1026.67$$

$$\text{Variance}, s^2 = \frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}$$

$$= \frac{1717332.8}{60}$$

$$= 28622.21$$

$$\therefore S.D, \sigma = \sqrt{s} = 169.181$$

$$\frac{S.D}{Mean} \times 100$$

$$C.V. = \frac{169.81}{1026.67} \times 100$$

$$= 16.53\%$$

similarly, for company B

calculate like Company A

we get, for company B, $S.D = 107.7$

Coefficient of variation $= 10.56\%$

As, the C.V. of company B gets lesser than cv of company A. that is the observations of company B are near around the central value. so, company B's bulbs are more uniform.

[Ans.]

Note:

$\boxed{*}$ A smaller standard deviation signifies that data points are closely packed together while a larger one indicates a more spread out dataset.