03-Measures of Dispersion

* The measurement the scattering of observations / items in a distribution above the avg → a measure of variations or Dispersion.

* They are 2 types :-
1) Absolute Measure
2) Relative Measure

1) Absolute (M) gives us an idea about the amount of Dispersion in a set of obsrvation. They are range, quartile deviation, Mean deviation, standard Deviation & varience.

2) Relative (M) are calculated for the comparison of dispersion in 2 / more than 2 observations. This (M) are a sort of ratio & are called coefficients.

I a) Range :-

It is the difference b/w Largest & smallest of the given values. (i.e)

$\boxed{Range = L - S}$ → Absolute (M)

b) $\boxed{\text{Coefficient of Range} = \frac{L-S}{L+S}}$ → Relative (M)

Q 1) Cal. range & its coefficient —

156, 165, 148, 151, 147, 162.

$Range = L - S$
$= 165 - 147 = 18$

$\text{Coefficient of Range} = \frac{165-147}{165+147} = \frac{18}{312}$

$= 0.057$

2) Cal. range & coefficient —

Mark =	10-20	20-30	30-40	40-50	50-60
No. of student =	8	10	12	8	4

(Don't consider, No freq)

$Range = L - S$
~~= 12 - 4 = 8~~
$= 60 - 10 = 50$

$\text{Coefficient of Range} = \frac{60-10}{60+10} = \frac{50}{70}$

$= 0.714$

II Quartile Deviation :- (QD)

QD is half of the difference b/w 1st & 3rd quartile.

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

Here $Q_3 - Q_1$ → Inter Quartile Range

∴ QD is the semi Inter quartile Range

coefficient of QD is :-

$\boxed{C \text{ of } QD = \frac{Q_3 - Q_1}{Q_3 + Q_1}}$

Q 1) find QD & its coefficient —

391, 384, 591, 407, 672, 777, 733, 1490, 2488, 522

A) 384, ~~38~~ 391, 407, 522, 591, 672, 733, 777, 1490, 2488.

$Q_1 = \frac{n+1}{4}^{th}$ item

$= \frac{10+1}{4} = \frac{11}{4} = 2.75^{th}$ items

$= 2^{nd}$ term $+ 0.75^{th}$ term $(3^{rd} - 2^{nd})$

$= 391 + 0.75(407 - 391)$

$= \underline{403}$

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

$= \frac{3(10+1)}{4}^{th} = \frac{3(11)}{4}^{th} = \frac{33}{4}^{th} = 8.25^{th}$

$= 8^{th}$ term $+ 0.25^{th}$ $(9^{th} - 8^{th})$

$= 777 + 0.25(1490 - 777)$

~~= 777.25 (713) =~~

$= 777 + 178.25$

$= \underline{955.25}$

$QD = \frac{Q_3 - Q_1}{2} = \frac{955.25 - 403}{2}$

$= 276.125$

Coefficient of QD $= \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{552.25}{1358.25}$

$= \underline{0.406}$

2) weekly wages of 52 labours are given below — cal QD & its coefficient.

wages —	No. of weeks	CF
15	1	1
30	4	5
45	8	13
60	21	34
75	10	44
80	8	52
	52	

$N = 52$

$Q_1 = \frac{N+1}{4}^{th}$

$= \frac{53}{4}^{th} = 13.25^{th}$

$= 13^{th} + 0.25^{th} (14^{th} - 13^{th})$

$= 45 + 0.25(60 - 45)$

$= 45 + 0.25(15)$

$= 45 + 3.75 = \underline{48.75}$

$Q_3 = \frac{3(N+1)}{4}^{th} = \frac{3(53)}{4} = \frac{159}{4} = 39.75^{th}$

$= 39^{th} + 0.75^{th}(40^{th} - 39^{th})$

$= 75 + 0.75\,(75 - 75)$

$= \underline{\underline{75}}$

$QD = \frac{Q_3 - Q_1}{2} = \frac{75 - 48.75}{2} = \frac{26.25}{2}$

$= \underline{\underline{13.125}}$

(C) of QD $= \frac{75 - 48.75}{75 + 48.75}$

$= \frac{26.25}{123.75} = \underline{\underline{0.212}}$

3) cal. QD & Q (C) of Dispersion.

X	F	CF
20-30	6	6
30-40	18	24
40-50	25	49
50-60	50	99
60-70	37	136
70-80	30	166
80-90	24	190
90-100	10	200
	200	

Q_1

Q_3

~~$Q_1 = \frac{N+1}{4}^{th} = \frac{201}{4}^{th} = 50.25^{th}$~~

~~$50^{th} + 0.25^{th}(51^{th} - 50^{th})$~~

$Q_1 = l_1 + \frac{\left(\frac{N}{4} - m\right)}{f} \times C$ $\qquad \frac{N}{4} = 50$

$= 50 + \frac{(50 - 49)}{50} \times 10$

$= 50 + \frac{10}{50} = 50 + 0.2$

$= \underline{\underline{50.2}}$

$Q_3 = l_3 + \frac{\left(\frac{3N}{4} - m\right)}{f} \times C$ $\qquad \frac{3N}{4} = 150$

$= 70 + \frac{150 - 136}{30} \times 10$

$= 70 + \frac{140}{30} = 70 + 4.66$

$= \underline{\underline{74.6}}$

$QD = \frac{Q_3 - Q_1}{2} = \frac{74.6 - 50.2}{2} = \frac{24.4}{2}$

$= \underline{\underline{12.23}}$

(C) of QD $= \frac{74.6 - 50.2}{74.6 + 50.2} = \frac{24.4}{124.8}$

$= \underline{\underline{0.195}}$

4) Cal. QD & its (C).

cls	F	Actual cls	CF
0-9	3	-0.5-9.5	3
10-19	9	9.5-19.5	12
20-29	15	19.5-29.5	27 Q_1
30-39	30	29.5-39.5	57
40-49	18	39.5-49.5	75 Q_3
50-59	5	49.5-59.5	80
	80		

$Q_1 = l_1 + \frac{\left(\frac{N}{4} - m\right)}{f} \times c$ $\qquad \frac{N}{4} = 20$

$= 19.5 + \frac{(20 - 12)}{15} \times 10$

$= 19.5 + \frac{80}{15}$

$= 19.5 + 5.333 = \underline{\underline{24.83}}$

$Q_3 = l_3 + \frac{\left(\frac{3N}{4} - m\right)}{f} \times c$ $\qquad \frac{3N}{4} = 60$

$= 39.5 + \frac{(60-57)}{18} \times 10$

$= 39.5 + \frac{30}{18}$

$= 39.5 + 1.666 = \underline{\underline{41.187}}$

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

$= \frac{41.17 - 24.83}{2} = \underline{\underline{8.17}}$

(C) of $QD = \frac{41.17 - 24.83}{41.17 + 24.83}$

$= \frac{16.34}{6600} = \underline{\underline{0.2475}}$

1.1 Mean Deviation :- (MD)

It is defined as the AM of the absoluted values of the (D) of obsrtion from some origin, say mean / median / mode.

$|x| \to -ve \to +ve$

* For a raw data,

MD about mean $= \frac{\Sigma |x - \bar{x}|}{n}$

MD about median $= \frac{\Sigma |x - M|}{n}$

* For a freq data,

MD about Mean $= \frac{\Sigma f |x - \bar{x}|}{\Sigma f}$

MD about Median $= \frac{\Sigma f |x - M|}{\Sigma f}$

* coefficient of MD (about Mean) :-

(C) of MD $= \frac{\text{MD about mean}}{\text{Mean}}$

coefficient of MD (about median) :-

(c) of MD $= \frac{\text{MD about median}}{\text{Median}}$

Q

1) cal MD about Mean —

8, 24, 12, 16, 10, 20

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

$$= \frac{30}{6} = 5$$

$$\bar{x} = \frac{90}{6} = 15$$

x	$x-\bar{x}$ (15)
8	7
24	9
12	3
16	1
10	5
20	5
90	30

M.D
The mean of absolute values of numerical differences b/w the no. of a set of their mean/M/Z

(C) of MD $= \frac{5}{15} = 0.33$

2) cal. MD about Median –

x	$x-M$ (48.5)
20	28.5
23	25.5
30	18.5
32	16.5
46	2.5
51	2.5
56	7.5
57	8.5
57	8.5
78	29.5

$\Sigma|x-M| = 148$

$N = 10$ (even)

$$M = \frac{\frac{n}{2}^{th} + (\frac{n}{2}+1)^{th}}{2}$$

$$= \frac{\frac{10}{2}^{th} + (\frac{10}{2}+1)^{th}}{2} = \frac{5^{th} + 6^{th}}{2}$$

$$= \frac{46+51}{2} = \frac{97}{2}$$

$$= 48.5$$

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

$$= \frac{148}{10}$$

$$= 14.8$$

(c) of MD $= \frac{14.8}{48.85} = 0.305$

3) cal. MD about Mean & its coefficient

c/s	f	x	fx	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$
0-10	5	5	25	18	90
10-20	15	15	225	8	120
20-30	17	25	425	2	34
30-40	11	35	385	12	132
40-50	2	45	90	22	44
	50		1150		420

$$MD = \frac{\Sigma f|x-\bar{x}|}{\Sigma f}$$

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{1150}{50} = 23$$

$$= \frac{420}{50} = 8.4$$

(c) of MD $= \frac{8.4}{23} = 0.365$

4) Cal . MD about Median & its Coefficient

C.I.S	f	CF	x	$\lvert x-13\rvert$	$f\lvert x-13\rvert$
0-4	2	2	2	11	22
5-9	6	8	7	6	36
(10-14)	10	(18)	12	1	10
15-19	7	25	17	4	28
20-24	5	30	22	9	45
	30				141

$$M = l + \frac{\left(\frac{n}{2} - M\right)}{f} \times c \qquad \frac{n}{2} = 15$$

$$= 9.5 + \frac{(15-8)}{10} \times 5 = 13$$

$$\cancel{= 10 + \frac{28}{10} = 10 + 2.8 = 12.8}$$

$$MD = \frac{\Sigma f\lvert x - M\rvert}{\Sigma f} = \frac{141}{30} = 4.7$$

$$(C) \text{ of } MD = \frac{4.7}{13} = 0.361$$

5) Cal MD about Mean & median

CF	x	f	Actual C.I.S	x	fx	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$	$\lvert x-M\rvert$	$f\lvert x-M\rvert$
3	0-9	3	-0.5-9.5	4.5	13.5	28.25	84.75	29.33	87.99
12	10-19	9	9.5-19.5	14.5	130.5	18.25	164.25	19.33	173.97
27	20-29	15	19.5-29.5	24.5	367.5	8.25	123.75	9.33	139.95
57	30-39	30	29.5-39.5	34.5	1035	1.75	52.5	0.67	20.1
75	40-49	18	39.5-49.5	44.5	801	11.75	211.5	10.67	192.06
80	50-59	5	49.5-59.5	54.5	272.5	21.75	108.75 ~~1367.5~~	20.67	103.35
		80			2620		~~1999.25~~ 1890.5		717.42 ~~717.42~~ ~~10949.07~~

$$MD \text{ about Mean} = \frac{\Sigma f\lvert x-\bar{x}\rvert}{\Sigma f}$$

$$\bar{x} = \frac{2620}{80} = 32.75$$

$$= \frac{1890.5}{80} = 23.63125$$

$$\text{Coefficient of } MD = \frac{23.63125}{32.75} = 0.7215$$

$$\text{Median} = l + \frac{\left(\frac{N}{2} - m\right)}{f} \times c \qquad \frac{N}{2} = 40$$

$$= 29.5 + \frac{(40-27)}{30} \times 10$$

$$= 29.5 + 4.333$$

$$= 33.83$$

$$MD \text{ about Median} = \frac{\Sigma f\lvert x-M\rvert}{\Sigma f}$$

$$= \frac{717.42}{80} \qquad \cancel{= \frac{10949.07}{80}}$$

$$= 8.96775 \qquad \cancel{= 10685.15216}$$

$$\text{Coefficient of } MD, \cancel{= \frac{10635.15216}{33.83}}$$

$$= \frac{8.96775}{33.83}$$

$$= 0.265$$

* **Theorem :-**

For a freq distribution the MD from an avg 'A' is min when A is the median.

Proof
$N = \Sigma f$

We have to show that the MD $= \frac{\Sigma f |x - A|}{N}$ is min when x = Median.

It is enough to show that,

let $'s' = \sum_{i=1}^{n} f_i |x_i - A|$ is min when $A = M$.

let $'x_k'$ be the item / midpoint of the cls just $< A$, so that

$$x_1 < x_2 < x_3 \ldots x_k < A \le x_{k+1} \ldots < x_n$$

$$\therefore \; s = \sum_{i=1}^{n} f_i |x_i - A| \quad \text{——} ①$$

(Expand Σ)

$$= \underbrace{f_1 |x_1 - A| + f_2 |x_2 - A| \ldots + f_k |x_k - A|}_{} + \underbrace{f_{k+1} |x_{k+1} - A| \ldots + f_n |x_n - A|}_{}$$

divide this into 2 set

$$S = \sum_{i=1}^{k} f_i |x_i - A| + \sum_{i=k+1}^{n} f_i |x_i - A|$$

Remove modulus $|\ldots|$

$$S = \sum_{i=1}^{k} f_i (A - x_i) + \sum_{i=k+1}^{n} f_i (x_i - A).$$

$$= \sum_{i=1}^{k} f_i A - \sum_{i=1}^{k} f_i x_i + \sum_{i=k+1}^{n} f_i x_i - \sum_{i=k+1}^{n} f_i A$$

combaining same values

$$= \sum_{i=1}^{k} f_i A - \sum_{i=k+1}^{n} f_i A - \sum_{i=1}^{k} f_i x_i + \sum_{i=k+1}^{n} f_i x_i$$

Take 'A' outside (common)

$$= A \left[\sum_{i=1}^{k} f_i - \sum_{i=k+1}^{n} f_i \right] + \left[-\sum_{i=1}^{k} f_i x_i + \sum_{i=k+1}^{n} f_i x_i \right]$$

(A cannot be changed) (x_i can be changed to its variable)

$$S = A f_1 + f_2 \quad ③$$

where, $f_1 = \sum_{i=1}^{k} f_i - \sum_{i=k+1}^{n} f_i$

$f_2 = -\sum_{i=1}^{k} f_i x_i + \sum_{i=k+1}^{n} f_i x_i$ ④

$f_1 = 0$ [A depend only f_1, min means $f_1 = 0$]

$$f_1 = \sum_{i=1}^{k} f_i - \sum_{i=k+1}^{n} f_i = 0$$

$$\Rightarrow \sum_{i=1}^{k} f_i = \sum_{i=k+1}^{n} f_i$$

(here eq ④ independent of A
$\therefore$ 'S' is min when $f_1 = 0$)

This is possible when A is median.

IV **Standard Deviation :- (SD) (σ)**

Defined as the square root of AM of the squares of the deviations of observations from AM.

* For a row data :-

$$\boxed{\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}}$$

* For a freq data :-

$$\boxed{\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{\Sigma f_i}}}$$

The square of SD $\longrightarrow$ Varience (σ^2).

→ Simplified formula for SD :-

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

$$\sigma^2 = \frac{1}{n}\Sigma(x-\bar{x})^2 \qquad (a-b)^2 = (x-\bar{x})^2$$

$$= \frac{1}{n}\Sigma\left(x^2 - 2x^2\bar{x} + \bar{x}^2\right)$$

$$= \frac{1}{n}\left[\Sigma x^2 - \Sigma 2x\bar{x} + \Sigma \bar{x}^2\right]$$

$\Sigma x = x+x \ldots x_n$
$\Sigma 1 = 1+1+1 \ldots 1_n$

$$= \frac{1}{n}\Sigma x^2 - 2\bar{x}\frac{\Sigma x}{n} + \frac{\bar{x}^2}{n}\Sigma 1$$

$$= \frac{1}{n}\Sigma x^2 - 2\bar{x}\times\bar{x} + \frac{\bar{x}^2}{\cancel{n}}\times\cancel{n} \quad \left(\text{since } \Sigma 1 = n \text{ \& } \bar{x} = \frac{\Sigma x}{n}\right)$$

$$= \frac{1}{n}\Sigma x^2 - 2\bar{x}^2 + \bar{x}^2$$

$$= \frac{\Sigma x^2}{n} - (\bar{x})^2$$

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

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

In similar way for a freq data,

$$\boxed{\sigma = \sqrt{\frac{\Sigma f x^2}{N} - \left(\frac{\Sigma f x}{N}\right)^2}} \qquad N = \Sigma f.$$

⇒ Short cut Method :-

a) for a raw data, $\boxed{\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}}$

where $d = x - A$

b) for a freq data, $\boxed{\sigma = c \times \sqrt{\frac{\Sigma f d^2}{N} - \left(\frac{\Sigma f d}{N}\right)^2}}$

where $d = \frac{x-A}{c}$

A = Assumed mean
c = c/s interval

→ Coefficient of SD :-

Coefficient of SD $= \frac{SD}{AM} = \frac{\sigma}{\bar{x}}$

(SD always +ve)

→ Properties of SD :-

1) SD is not affected by change of origin.

Proof: let $x_1, x_2 \ldots x_n$ be a set of n observation

then, $\sigma_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$ —— ①

let $y_i = x_i + c$ $\quad (i = 1,2,3 \ldots n)$ $\quad y_1 = x_1 + c \ldots$

then $\bar{y} = \bar{x} + c$ $\quad \bar{y} = \frac{\Sigma y}{n}$

$\therefore \; y_i - \bar{y} = x_i + \cancel{c} - (\bar{x} + \cancel{c})$

$y_i - \bar{y} = x_i - \bar{x}$

$(y_i - \bar{y})^2 = (x_i - \bar{x})^2$

$\Sigma(y_i - \bar{y})^2 = \Sigma(x_i - \bar{x})^2$ (dividing throughout by n)

$\frac{\Sigma(y_i - \bar{y})^2}{n} = \frac{\Sigma(x_i - \bar{x})^2}{n}$ (taking square root on both side)

$$\sqrt{\frac{\Sigma(y_i-\bar{y})^2}{n}} = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}} = \sigma_x$$

using eq ① we get.

$$\sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}} = \sigma_x \qquad \sqrt{\frac{\Sigma(y_i-\bar{y})^2}{n}} = \sigma_y$$

$$\sigma_y = \sigma_x$$

2) SD is affected by change & scale.

Proof: let $x_1, \dots x_2 \dots x_n$ be a set of observation.

$$\sigma_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} \quad -①$$

let $y_i = cx_i + d$, where $i = 1, 2, 3 \dots n$

$$\bar{y} = c\bar{x} + d$$

$$\therefore y_i - \bar{y} = cx_i + d - (c\bar{x} + d)$$

$$y_i - \bar{y} = cx_i - c\bar{x} = c(x_i - \bar{x})$$

$$(y_i - \bar{y})^2 = c^2 (x_i - \bar{x})^2$$

$$\frac{\Sigma(y_i - \bar{y})^2}{n} = \frac{c^2}{n} \Sigma(x_i - \bar{x})^2$$

taking $\sqrt{\ }$ on both side

$$\sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n}} = \sqrt{\frac{c^2}{n} \Sigma(x_i - \bar{x})^2}$$

$$\sqrt{c^2} = c$$

$$\sigma_x = \frac{\Sigma(x_i - \bar{x})^2}{n}$$

$$\Rightarrow \sigma_y = c\,\sigma_x$$

3) If 1 grp of 'n_1' observation have an AM '$\bar{x}_1$' & SD 'σ_1' & another grp of 'n_2' observation have an AM '$\bar{x}_2$' & SD 'σ_2' then SD, σ of 2 grps combined given by,

$$(n_1+n_2)\sigma^2 = n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 d_1^2 + n_2 d_2^2$$

where $d_1 = \bar{x}_1 - \bar{x}$

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

$\bar{x} \rightarrow$ combined AM.

Proof

let $x_1, x_2 \dots x_{n_1}$ be the observations of 1st grp & $x_1', x_2' \dots x_{n_2}'$ be the obs. of 2nd grp.

$$\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}(x_i), \quad \bar{x}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} x_i'$$

$$\sigma_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2, \quad \sigma_2^2 = \frac{1}{n_2}\sum_{i=1}^{n_2}(x_i' - \bar{x}_2)^2$$

$n\sigma_1^2 = \Sigma(x_i - \bar{x}_1)^2$ ⓐ

then the combined $\sigma^2 \Rightarrow \sigma^2 = \frac{1}{n_1+n_2}\sum_{i=1}^{n_1+n_2}(x_r - \bar{x})^2$. ①

where, $\bar{x} \rightarrow$ combined AM.

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

from ①

split

$$(n_1+n_2)\sigma^2 = \sum_{i=1}^{n_1+n_2}(x_i - \bar{x})^2 = \sum_{i=1}^{n_1}(x_i - \bar{x})^2 + \sum_{i=1}^{n_2}(x_i' - \bar{x})^2$$

adding & subs $\bar{x}_1, \bar{x}_2 \rightarrow$

$$\sum_{i=1}^{n_1}(\underbrace{x_i - \bar{x}_1}_{a} + \underbrace{\bar{x}_1 - \bar{x}}_{b})^2 + \sum_{i=1}^{n_2}(\underbrace{x_i' - \bar{x}_2}_{a} + \underbrace{\bar{x}_2 - \bar{x}}_{b})^2.$$

$(a+b)^2$

$$= \sum_{i=1}^{n_1}\left[(x_i - \bar{x}_1)^2 + 2(x_i - \bar{x}_1)(\bar{x}_1 - \bar{x}) + (\bar{x}_1 - \bar{x})^2\right] + \sum_{i=1}^{n_2}\left[(x_i' - \bar{x}_2)^2 + 2(x_i' - \bar{x}_2)(\bar{x}_2 - \bar{x}) + (\bar{x}_2 - \bar{x})^2\right]$$

constant

give Σ to all x

$$= \sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2 + 2(\bar{x}_1 - \bar{x})\sum_{i=1}^{n_1}(x_i - \bar{x}_1) + (\bar{x}_1 - \bar{x})^2 \times \sum_{i=1}^{n_1} 1 + \sum_{i=1}^{n_2}(x_i' - \bar{x}_2)^2 + 2(\bar{x}_2 - \bar{x})\sum_{i=1}^{n_2}(x_i' - \bar{x}_2) + (\bar{x}_2 - \bar{x})^2 \times \sum_{i=1}^{n_2} 1$$

ⓐ

$$= n_1\sigma_1^2 + 2(\bar{x}_1 - \bar{x}) \times 0 + (\bar{x}_1 - \bar{x}) \times n_1 + n_2\sigma_2^2 + 2(\bar{x}_2 - \bar{x}) \times 0 + (\bar{x}_2 - \bar{x})^2 \times n_2$$

$$= n_1\sigma_1^2 + n_1(\bar{x}_1 - \bar{x})^2 + n_2\sigma_2^2 + n_2(\bar{x}_2 - \bar{x})^2$$

$$= n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 d_1^2 + n_2 d_2^2 \quad -①$$

where $d_1 = (\bar{x}_1 - \bar{x})$

$d_2 = (\bar{x}_2 - \bar{x})$

sub this value $\rightarrow$ $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$

The result can be simplified as follows –

$$\text{let } n_1 d_1^2 + n_2 d_2^2 = n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2$$

$$= n_1\left[\bar{x}_1 - \left(\frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}\right)\right]^2 + n_2\left[\bar{x}_2 - \left(\frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}\right)\right]^2$$

$$= n_1\left[\frac{(n_1+n_2)\bar{x}_1 - (n_1\bar{x}_1+n_2\bar{x}_2)}{n_1+n_2}\right]^2 +$$

$$n_2\left[\frac{(n_1+n_2)\bar{x}_2 - (n_1\bar{x}_1+n_2\bar{x}_2)}{n_1+n_2}\right]^2$$

$$= n_1\left[\frac{\cancel{n_1\bar{x}_1}+n_2\bar{x}_1-\cancel{n_1\bar{x}_1}-n_2\bar{x}_2}{n_1+n_2}\right]^2 + n_2\left[\frac{n_1\bar{x}_2+\cancel{\bar{x}_2 n_2}-n_1\bar{x}_1-\cancel{n_2\bar{x}_2}}{n_1+n_2}\right]^2$$

$$= n_1\left[\frac{n_2(\bar{x}_1-\bar{x}_2)}{n_1+n_2}\right]^2 + n_2\left[\frac{n_1(\bar{x}_2-\bar{x}_1)}{n_1+n_2}\right]^2$$

give sqre →

$$= n_1 \times \frac{n_2^2(\bar{x}_1-\bar{x}_2)^2}{(n_1+n_2)^2} + n_2 \times \frac{n_1^2(\bar{x}_1-\bar{x}_2)^2}{(n_1+n_2)^2}$$

$(\bar{x}_1-\bar{x}_2)^2 = (\bar{x}_2-\bar{x}_1)^2$

$(a-b)^2 = (b-a)^2$

$$= \frac{n_1 n_2}{(n_1+n_2)^2} \times (\bar{x}_1-\bar{x}_2)^2 \underbrace{[n_2+n_1]}_{(n_1+n_2)}$$ ← take common

$$= \frac{n_1 n_2 (\bar{x}_1-\bar{x}_2)^2}{(n_1+n_2)^{\cancel{2}}} \cancel{(n_1+n_2)}$$

$$n_1 d_1^2 + n_2 d_2^2 = \frac{n_1 n_2}{(n_1+n_2)} \times (\bar{x}_1-\bar{x}_2)^2 \quad \text{—②}$$

∴ Sub —② in —①

$$\boxed{(n_1+n_2)\sigma^2 = n_1\sigma_1^2 + n_2\sigma_2^2 + \frac{n_1 n_2}{(n_1+n_2)}(\bar{x}_1-\bar{x}_2)^2}$$

(Combained SD)

$\sigma^2 = \frac{n_1\sigma_1^2 \ldots}{(n_1+n_2)} = \sigma^2 \sqrt{\ldots}$

$\sigma = \ldots$

4) SD cannot be smaller than MD about mean.

$\boxed{SD \geq MD}$

for a raw data,

$SD = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}}$, $MD(mean) = \frac{\Sigma|x_i-\bar{x}|}{n}$ —②

$\therefore SD^2 = \frac{1}{n}\Sigma(x_i-\bar{x})^2$ —①

$= \frac{1}{n}\Sigma|x_i-\bar{x}|^2$ from ① & ②

$= \frac{1}{n}\Sigma z_i^2$ where $z_i = |x_i-\bar{x}|$

$\therefore MD = \frac{1}{n}\Sigma|x_i-\bar{x}| = \frac{1}{n}\Sigma z_i$

$SD^2 - MD^2 = \frac{1}{n}\Sigma z_i^2 - \left(\frac{1}{n}\Sigma z_i\right)^2$

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

$= \sigma_z \geq 0$ $\sigma_z = \frac{\Sigma z_i^2}{n} - \left(\frac{\Sigma z_i}{n}\right)^2$

$\Rightarrow SD^2 - MD^2 \geq 0$

$SD^2 \geq MD^2$

Taking square root on both side.

$SD \geq MD$

Q 1) Cal. SD of

x	$x-\bar{x}$	$(x-\bar{x})^2$
23	−10	100
25	−8	64
28	−5	25
31	−2	4
38	5	25
40	7	49
46	13	169
231		436

$\sigma = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}}$

$n = 7$

$\bar{x} = \frac{\Sigma x}{n} = 33$

$\sigma = \sqrt{\frac{436}{7}} = \sqrt{62.28} = 7.892$

2) Cal. SD

x	x^2
1	1
2	4
3	9
4	16
5	25
6	36
7	49
8	64
9	81
10	100
55	385

using simplified form of eq −

$\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$ $n = 10$

$= \sqrt{\frac{385}{10} - (5.5)^2}$ $\bar{x} = \frac{\Sigma x}{n} = 5.5$

$= \sqrt{38.5 - 30.25}$

$= \sqrt{8.25} = 2.872$

3) Cal. SD using short cut method

x	$d = x - A$	d^2
42	-8	64
48	-2	4
50	0	0
62	12	144
65	15	225
	17	437

n=5

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

$d = x - A$

$A = 50$ (centre value)

$$= \sqrt{\frac{437}{5} - \left(\frac{17}{5}\right)^2}$$

$$= \sqrt{\frac{437}{5} - 11.56}$$

$$= \sqrt{87.4 - 11.56}$$

$$= \sqrt{75.84} = 8.708$$

5) SD.

X	F	$(x - 13.7)^2$	$f(x-\bar{x})^2$	fx
10	2	13.69	27.38	20
12	4	2.89	11.56	48
14	10	0.09	0.9	140
16	3	5.29	15.87	48
18	1	18.49	18.49	18
	20		64.2	274

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

$$= \sqrt{\frac{69.2}{20}} = 1.86$$

$\bar{x} = \frac{\Sigma fx}{\Sigma f}$

$= \frac{274}{20}$

$= 13.7$

5) Cal. SD.

X	F	fx	$(x-39.2)^2$	$f(x-\bar{x})^2$
10	3	30	852.64	2557.92
20	5	100	368.64	1843.2
30	7	210	84.64	592.48
40	20	800	0.64	12.8
50	8	400	116.64	933.12
60	7	420	432.64	3028.48
	50	1960		8968

6) Cal. range & SD

Mrk	No.	fx	$(x-51.6)^2$	$f(x-\bar{x})^2$
43	1	43	73.96	73.96
48	2	96	12.96	25.92
50	3	150	2.56	7.68
49	4	196	6.76	27.04
51	5	255	0.36	1.8
60	6	360	70.56	423.36
55	7	385	12.24	85.68
52	8	416	0.16	1.28
49	9	441	6.76	60.84
50	10	500	2.56	25.6
	55	2842		733.16

$$Range = L - S$$
$$= 60 - 43 = 17$$

$$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}}$$
$$= \sqrt{\frac{733.16}{55}}$$
$$= \sqrt{13.33} = 3.651$$

$$\bar{x} = \frac{\Sigma fx}{\Sigma f}$$
$$= \frac{2842}{55}$$
$$= 51.6$$

5.A)

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{1960}{50} = 39.2$$

$$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}}$$
$$= \sqrt{\frac{8968}{50}} = \sqrt{179.36} = 13.39$$

24/1/2022

=> Coefficient of variation :- (CV)

* It is the % variation in the mean.

(i.e) $\boxed{CV = \frac{SD}{\bar{x}} \times 100}$

* The grp which has less CV → More consistent / more uniform / more stable.
* More CV indicates greater variability / less consistency / less uniformity / less stability.

Q
1) Cal. mean, SD & CV.

c/s	F	x	fx	$(x-14.2)^2$	$f(x-\bar{x})^2$
0-6	5	3	15	125.44	627.2
6-12	12	9	108	27.04	324.48
12-18	15	15	225	0.64	9.6
18-24	10	21	210	46.24	462.4
24-30	3	27	81	163.84	491.52
	45		639	363.2	1915.2

$\bar{x} = \frac{\Sigma fx}{\Sigma f}$

$= \frac{639}{45}$

$= \cancel{10.65}$

$= 14.2$

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

$\cancel{= \frac{363.2}{45} = 8.07}$

$= \sqrt{\frac{1915.2}{45}} = \sqrt{42.56}$

$= 6.52$

$CV = \frac{SD}{\bar{x}} \times 100$

$= \frac{6.52}{14.2} \times 100 = 0.4591 \times 100$

$= 45.91$

2) The runs scored by 2 bats man in 5 ~~inings~~ inings are given below, find who is the more consistent bat man?

(x) A	(x) B	CV of A. $(x-44)^2$	~~CV of B~~ ~~$(x-49)^2$~~	$(x-49)^2 \to$ CV of B
25	10	361	~~1556~~	1521
50	70	36	~~676~~	441
45	50	1	~~36~~	1
30	20	196	~~576~~	841
70	95	676	~~2601~~	2116
		1270	~~5045~~	4920

$\bar{x} = \frac{\Sigma x}{n} = \frac{220}{5} = 44$

$\sigma_A = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$

$= \sqrt{\frac{1270}{5}}$

$= \sqrt{254} = 15.93$

$CV(A) = \frac{SD}{\bar{x}} \times 100$

$= \frac{15.93}{44} \times 100$

$= 0.3620 \times 100 = 36.20$

$\sigma_B = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{4920}{5}} = 31.36$

~~$= \sqrt{\frac{5045}{5}} = \sqrt{1009} = 31.76$~~

~~CV (B)~~

$\bar{x} = \frac{245}{5} = 49$

$CV(B) = \frac{SD}{\bar{x}} \times 100$

$= \frac{31.36}{49} \times 100 = 0.64 \times 100 = 64$

~~$= 6.273 \times 100$~~ ~~$= 6.273$~~

A has less CV $\therefore$ A is more consistent.

3)

C I S	f	x	fx	$(x-26.14)^2$	$f(x-\bar{x})^2$
0-10	3	5	15	446.8996	1340.6988
10-20	10	15	150	124.0996	1240.996
20-30	36	25	900	1.2996	46.7856
30-40	18	35	630	78.4996	1412.9928
40-50	3	45	135	355.6996	1067.0988
	70		1830		5108.572

$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{1830}{70} = 26.14$

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

$= \sqrt{\frac{5108.572}{70}} = \sqrt{72.9796}$

$= 8.542$

$CV = \frac{SD}{\bar{x}} \times 100$

$= \frac{8.542}{26.14} \times 100 = 32.68$

4) Cal. SD

x	$(x-53.4)^2$
42	129.96
48	29.16
50	11.56
62	73.96
65	134.56
267	379.2

$\bar{x} = \frac{\Sigma x}{n} = \frac{267}{5} = 53.4$

$\sigma = \sqrt{\frac{\Sigma (x-\bar{x})^2}{n}}$

$= \sqrt{\frac{379.2}{5}}$

$= \sqrt{75.84}$

$\sigma = 8.70$

5) In 2 factories A & B engaged in the same industry in an area, the avg weekly wage in ₹ & SD R as follows.

factory	Avg	SD	No. of workers
A	36	5	250
B	38	4.5	150

find the avg & SD of workers in 2 factories taken together.

☆

A) $(n_1+n_2)\sigma^2 = n_1\sigma_1^2 + n_2\sigma_2^2 + n_1d_1^2 + n_2d_2^2$

$d_1 = \bar{x}_1 - \bar{x}$

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

$n_1 = 250 \quad n_2 = 150$

$\sigma_1 = 5 \quad \sigma_2 = 4.5$

$\bar{x}_1 = 36 \quad \bar{x}_2 = 38$

$\bar{x}$ = combined AM

$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2} = \frac{250\times36+150\times38}{250+150}$

$= \frac{9000+5700}{400} = \frac{14700}{400}$

$= 36.75$

$d_1 = \bar{x}_1 - \bar{x}$

$= 36 - 36.75 = -0.75$

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

$= 38 - 36.75 = 1.25$

$\sigma^2 = \frac{250\times25+150\times20.25+250\times0.5625+150\times1.5625}{(250+150)}$

$= \frac{6250+3037.5+140.625+234.375}{400}$

$= \frac{\cancel{2353478.125}}{\cancel{400}} = \frac{9662.5}{400}$

$= 24.15$

$\sigma^2 = 24.15$

$\sigma = \sqrt{24.15}$

$= 4.914$

6) From the data given below find which series is more consistent.

C.I.s	10-20	20-30	30-40	40-50	50-60	60-70
Series A	10	16	30	40	26	18
B	22	18	32	34	18	16

CV of A

x	f	fx	$f(x-42.85)^2$
15	10	150	7756.225
25	16	400	5097.96
35	30	1050	1848.675
45	40	1800	184.9
55	26	1430	3838.185
65	18	1170	8831.205
	140	6000	27557.15

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{6000}{140} = 42.85$$

$$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}} = \sqrt{\frac{27557.15}{140}} = \sqrt{196.83} = 14.02$$

$$CV = \frac{SD}{\bar{x}} \times 100 = 32.74$$

CV of B

x	f	fx	$f(x-39)^2$
15	22	330	12672
25	18	450	3528
35	32	1120	512
45	34	1530	1224
55	18	990	4608
65	16	1040	10816
	140	5460	33360

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{5460}{140} = 39$$

$$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}} = \sqrt{\frac{33360}{140}} = 15.43$$

$$CV = \frac{SD}{\bar{x}} \times 100 = 39.58$$

A is more consistent.

★ ⇒ Relationship among different measures of Dispersion :-

1) QD $\approx$ (approx) $\frac{2}{3}$ SD
2) MD $\approx \frac{4}{5}$ SD
3) QD $\approx \frac{5}{6}$ MD
4) QD : MD : SD = 10 : 12 : 15
5) Range $\simeq 6 \times$ SD

⇒ Root Mean Square Deviation :- (S)

It is defined as the squares of the deviations of observations from any value say A. (S).

For raw data —

$$\boxed{S = \sqrt{\frac{\Sigma (x-A)^2}{n}}}$$

for freq data —

$$\boxed{S = \sqrt{\frac{\Sigma f(x-A)^2}{\Sigma f}}}$$

Result :-

SD is the least root mean square deviation.

Proof

we have to show that $\sigma < S$.

let $x_1, x_2, \dots x_n$ be set of n observation

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

$$S = \sqrt{\frac{\sum(x-A)^2}{n}} \Rightarrow S^2 = \frac{\sum(x-A)^2}{n}$$

Adding $\bar{x}$ & subtracting $\bar{x}$, $\quad nS^2 = \sum(x-A)^2$ --- (2)

$$nS^2 = \sum\left((x-\bar{x}) + (\bar{x}-A)\right)^2 \longrightarrow (a+b)^2$$

$$nS^2 = \sum\left[(x-\bar{x})^2 + 2(x-\bar{x})(\bar{x}-A) + (\bar{x}-A)^2\right]$$

$$= \sum(x-\bar{x})^2 + 2(\bar{x}-A)\sum(x-\bar{x}) + \underbrace{\sum(\bar{x}-A)^2}_{n \text{ times}}$$

in eq-(1)
$n\sigma^2 = \sum(x-\bar{x})^2$
$\sum(x-\bar{x}) = 0$.

$$= n\sigma^2 + 2(\bar{x}-A)\times 0 + n(\bar{x}-A)^2$$

$$\cancel{n}S^2 = \cancel{n}\sigma^2 + \cancel{n}(\bar{x}-A)^2$$

$$S^2 = \sigma^2 + (\bar{x}-A)^2$$

here $(\bar{x}-A)^2 > 0$

$\therefore S^2 = \sigma^2 +$ a positive no.

$\Rightarrow S^2 > \sigma^2$ (taking root on both sides)

$S > \sigma$.

24/1/2022

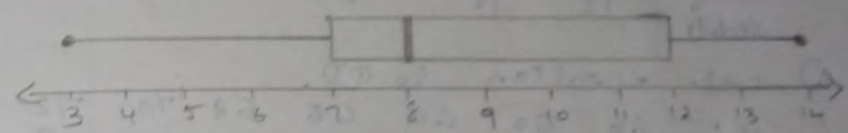
⇒ Box plot :-

They are a type of graph that can help visually organised data. To graph a box plot the following data points must be calculated — the min value, the 1st quartile, median, the 3rd quartile & max value.

once the box plot is graphed, you can display & compare distribution of data.

Q
1) construct a Box plot for the follow data
7, 3, 14, 9, 7, 8, 12

A) ascending order →
3, 7, 7, 8, 9, 12, 14
min value - 3   max - 14.

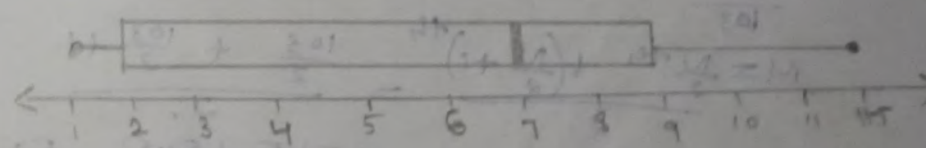


$Q_1 = 7$ → $\frac{n+1}{4} = \frac{8}{4} = 2^{nd} \to 7$

$Q_3 = 12$ → $\frac{3(n+1)}{4} = \frac{24}{4} = 6^{th} \to 12$

$M = 8$ (checked) → $\frac{n+1}{2} = \frac{8}{2} = 4^{th} \to 8$

2) construct a B.P —
1, 1, 2, 2, 4, 6, 6.8, 7.2, 8, 8.3, 9, 10, 10, 11.5



$$M = \frac{6.8 + 7.2}{2} = \underline{\underline{7}}$$

$$Q_1 = 2$$

$$Q_3 = 9$$