"STATISTICS"

"It is the science which deals with the collection, presentation, analysis & interpretation data as well as drawing valid conclusion and making seasonable decision on the basis of such analysis.

Q: The final marks in statistics of 60 studen are given below:
Average them into a frequency distribution table of size 5 showing marks in the col of Tally.

61	69	57	59	70	65	62	57	75	63
69	57	62	60	73	85	75	61	67	69
51	56	54	64	77	95	86	68	69	50
64	56	69	61	73	74	71	67	58	60
51	58	89	92	97	89	68	93	69	59
76	52	75	67	64	99	58	65	63	64

DATA

Discrete (Counted

Continuous (Measurable)

* For frequency distribution we need two columns of class interval & tally.

Class interval

* Total no of classes = $N = \dfrac{\text{Range}}{\text{width / Size of CI}}$

Range = Max − Min Value

$R = X_m - X_n$

Soc:

Max = X_m = 99 , Min = X_n = 50

width size = 5

$N = \dfrac{99-50}{5}$

$N = \dfrac{49}{5} = 9.8 \simeq 10$

CI	Tally	No. of Students (f)
50-54	~~IIII~~	5
55-59	~~IIII~~ ~~IIII~~ I	11
60-64	~~IIII~~ ~~IIII~~ III	13
65-69	~~IIII~~ ~~IIII~~ II	12
70-74	~~IIII~~	5
75-79	~~IIII~~	5
80-84	–	0
85-89	IIII	4
90-94	II	2
95-99	III	3
	Total	60

Q: With Reference to the above data find (i) the highest marks (ii) lowest marks of 3 highest ranking students (iv) the marks of 5 lowest ranking student (v) How many student recieve marks more than 80?

(i) 99 (iii) 95, 97, 99

(ii) 50 (iv) 50, 51, 51, 52, 54

(v) 9 students

Lower class boundary

upper class boundary

no. of student / total stu

(x) Mid Point	Class Boundaries	Relative Freq	%RF
52	49.5 — 54.5	5/60	$\frac{5}{60} \times 100$
57	54.5 — 59.5	14/60	
62	59.5 — 64.5	13/60	
67	64.5 — 69.5	12/60	
72	69.5 — 74.5	5/60	
77	74.5 — 79.5	5/60	
82	79.5 — 84.5	0/60	
87	84.5 — 89.5	4/60	
92	89.5 — 94.5	2/60	
97	94.5 — 99.5	3/60	$\frac{3}{60} \times 100$
		1	100

Total

comulative freq. less than | more than

$C.F_L$	$C.F_m$
5	60
16	55
29	44
41	31
46	19
51	14
51	9
55	9
57	5
60	3

CF_m → mein neeche se upper add krte jayenge. CF_L mein upper se neeche add krenge.

Q: Find Commulative Freq. distribution for less than and more than of the following freq. distribution

CI	2-5	6-9	10-13	14-17	18-21	Total
f	1	3	8	6	2	20

Class Boundaries	f
1.5 — 5.5	1
5.5 - 9.5	3
9.5 - 13.5	8
13.5 - 17.5	6
17.5 - 21.5	2

Commulative Freq. dist for 'less than' type or (F<) is given below

upper class boundary

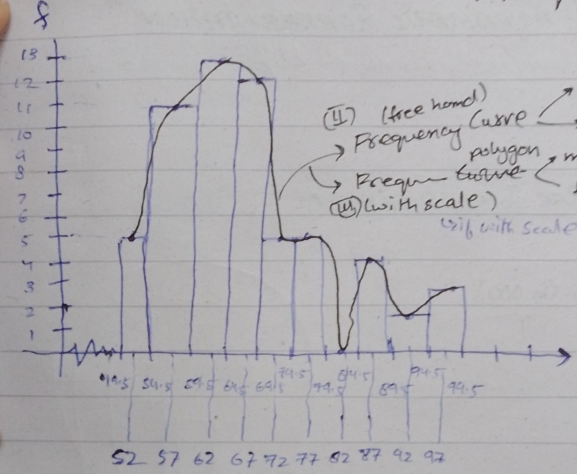
less than U.C.B	C.F
less than 5.5	1
less than 9.5	4
less than 13.5	12
less than 17.5	18
less than 21.5	20

Comm. freq dist for "more than" type or (F>) is given below

more than L.C.B	C.F
1.5 or more	20
5.5 or more	19
9.5 or more	16
13.5 "	8
17.5 or more	2

Graphical (I) Histogram < class boundary / freq.

in dono ve domain draw hoga



(II) (free hand) Frequency Curve < midpoint, freq

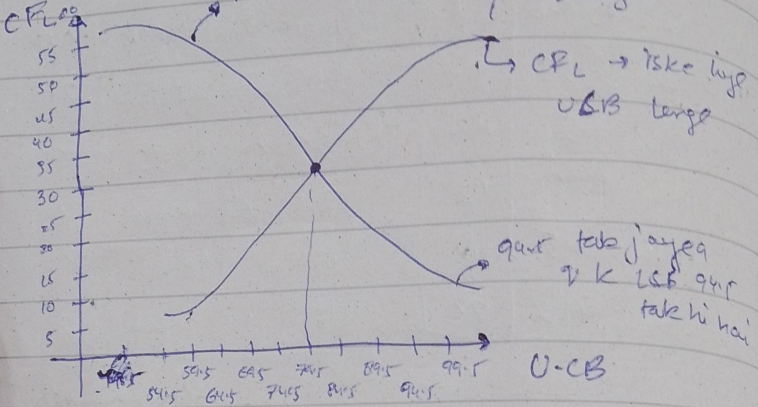
Frequency polygon < midpoint, freq

(III) (with scale)

only with scale.

(iv) Frequency Cummulative Polygon/Curve (OGives)

C·F_L / C·Fm
- C·F_L / C·Fm
- UCB / LCB

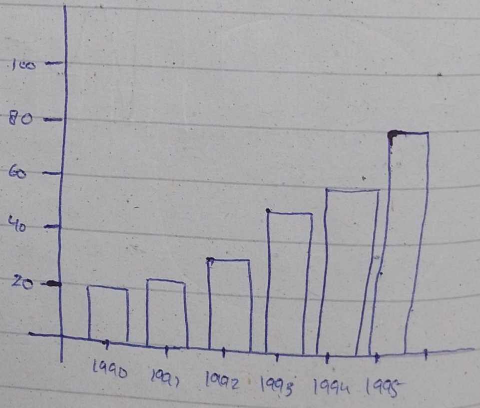


" DIAGRAMATIC REPRESENTATION "

"Representation of data with the help of diagram"

① SIMPLE BAR CHART / DIAGRAM :

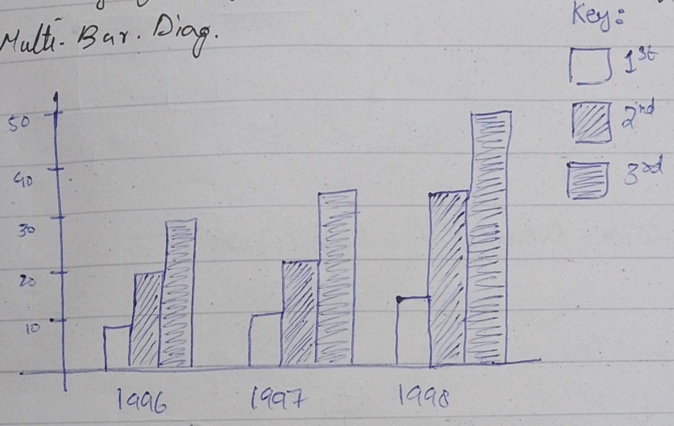
Years	Sales (in 000)
1990	20
1991	22
1992	35
1993	50
1994	60
1995	80



② MULTIPLE BAR DIAGRAM : (used for comparasion b/w two or three phenomenon over no. of years

Years	Division		
	First	second	third
1996	8	18	28
1997	10	20	32
1998	12	32	48

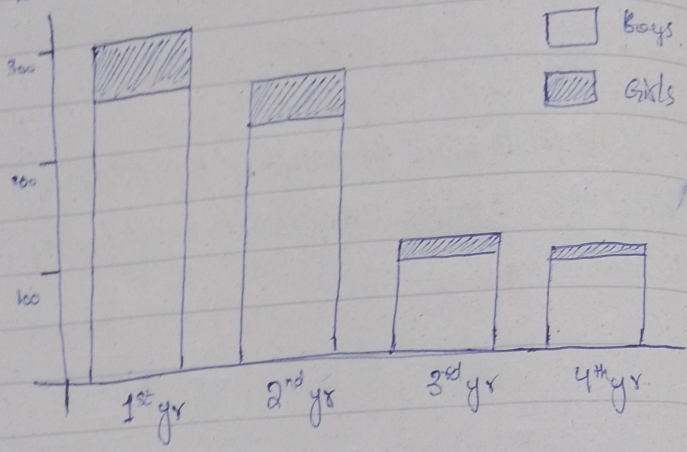
Q: Given above table gives the result of BS (Stats) of a college for 3 years. Represent the data by Multi. Bar. Diag.



③ SUBDIVIDE / COMPONENT BAR DIAG. : (to show the division of the total into variours parts (component)

Q The table gives the no. of students in a college during the yr 1990-1991. Represent the data by sub-divided bar diag.

CLASS	No. of Student Boys	Girls	Total
1st yr	250	50	300
2nd yr	220	30	250
3rd yr	80	20	100
4th yr	70	10	80

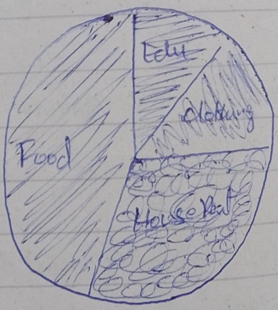


④ SECTOR DIAGRAM / PIE DIAG / CHART:

$$\frac{\text{Component Value}}{\text{Whole Quantity}} \times 360^\circ$$

$\frac{1200}{3000} \times 360^\circ = 144^\circ$

Item	Expenditure
Food	1200 (144°)
House Rent	1000 (120°)
Clothing	500 (60°)
Education	300 (36°)
Total	3000



"MEASURE OF CENTRAL TENDANCY"

"A certain value that singly defines the characteristics of a given set of data"

↱ Arithmetic mean of Arith-seq gives middle value.

(1) Arithemetic Mean ⟶ Sum / No. of terms

(2) Geometric Mean ⟶ $\log^{-1}\left[\frac{\Sigma \log n}{n}\right]$

(3) Harmonic Mean ⟶ $n / \Sigma \frac{1}{x}$
(Apply on inverse relation)

↱ if even no. of terms than mean = avg of two middle no.s

(4) Median ⟶ Middle most value
Positional Mean

(5) Mode ⟶ Most occuring value.

Q: Calculate (i) Mean (AM) (ii) GM (iii) HM (iv) Median (v) Mode of the following data.

2, 3, 3, 9, 6, 6, 12, 11, 8, 2, 3, 5, 7, 5, 4, 4, 5, 12, 9

(I) ~~AM = Mean~~ $= \bar{X} = \frac{\Sigma x}{N}$ or $\frac{\Sigma x}{n}$

"For Ungrouped Data"

1. $AM = Mean = \frac{\Sigma x}{N} = \frac{\Sigma x}{n}$

2. $GM = Antilog\left(\frac{\Sigma \log x}{n}\right)$

3. $HM = \frac{n}{\Sigma \frac{1}{x}}$

4. $Median = \left(\frac{n+1}{2}\right)^{th}$ value

5. Mode → most repeated value
   → most occuring value
   → most fashionable value

"For Grouped Data"

1. $AM = Mean = \bar{X} = \frac{\Sigma fx}{\Sigma f}$ or $\frac{\Sigma x}{n}$

2. $GM = Antilog\left(\frac{\Sigma f \log x}{\Sigma f}\right)$

3. $HM = \frac{\Sigma f}{\Sigma \frac{f}{x}}$

4. $Median = l + \frac{h}{f}\left(\frac{\Sigma f}{2} - CF\right)$

where,

l = lower class boundary

h = width / Size of class boundary

f = Median group freq.

CF = Commulative Freq.

5. $$\text{Mode} = l + \left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right) * h$$

where,

l = lower class boundary

f_m = Max frequency

h = Width / Size of Class Bound.

f_1 = Preceding Freq. (max k uper waali)

f_2 = Proceeding Freq. (max k neeche waali)

x	$\log x$	$1/x$
2	0.3010	0.5
2	0.3010	0.5
3	0.4771	0.33
3	0.4771	0.33
3	0.4771	0.33
4	0.6020	0.25
4	0.6020	0.25
5	0.6989	0.20
5	0.6989	0.20
5	0.6989	0.20
6	0.7781	0.169
6	0.7781	0.169
7	0.8450	
8	0.9030	
9	0.9542	
9	0.9542	
11	1.0460	
12	1.0790	0.083
12	1.0790	0.083
$\sum x = 116$	$\sum \log x = 13.749$	$\sum 1/x = 4.18$

$\bar{X} = \frac{\sum x}{n} = \frac{116}{19}$ $\boxed{A.M = 6.05}$

$G.M = \text{Antilog}\left[\frac{13.747}{19}\right]$

$\boxed{G.M = 5.37}$

$H.M = n / \sum 1/x$

$= 19/4.18$

$\boxed{H.M = 4.536}$

$\boxed{A.M > G.M > H.M}$

Median = 5

Mode = 3 & 5

→ Mean, Median, Mode nikalna hay or ~~dat~~ class intervals inclusive diye ae hai unko exclusive krenge, jese neeche waale sawal mein hai.

not in exam only GM

Q2: Calculate (1) AM (2) GM (3) HM (4) Median & mode of the following freq. dist.

inclusive → exclusive

C.I. marks	'f' no. of stdu.	mid point x	fx	C.B	C.f
50-54	5	52	260	49.5 – 54.5	5
55-59	4 f_1	57	627	54.5 – 59.5	16
60-64	13 f_m	62	806	59.4 – 64.5	29 → C.F
65-69	12 f_2	67	804	64.5 – 69.5	46
70-74	5	72	360	69.5 – 74.5	56
75-79	5	77	385	74.5 – 79.5	51
80-84	0	82	0	79.5 – 84.5	51
85-89	4	87	348	84.5 – 89.5	55
90-94	2	92	184	89.5 – 94.5	57
95-99	3	97	291	94.5 – 99.5	60
	$\Sigma f = 60$		$\Sigma fx = 4065$		

→ l

$\frac{\Sigma f}{2} = 30$ hum C.F mein ye value dhoondenge or uske C.B ki ... C.B = L hoga

$$AM = \frac{\Sigma fx}{\Sigma f} = \frac{4065}{60} = 67.75$$

$$Med = l + \frac{h}{f}\left(\frac{\Sigma f}{2} - C.F\right) = 69.5 + \frac{5}{12}\left(\frac{60}{2} - 29\right)$$

64.5

l se ek preceding val.

$$= 69.5 + 5/12(30-29)$$

$$Med = 64.9$$

$$Mod = 62.83$$

22k-4187

Q#1.

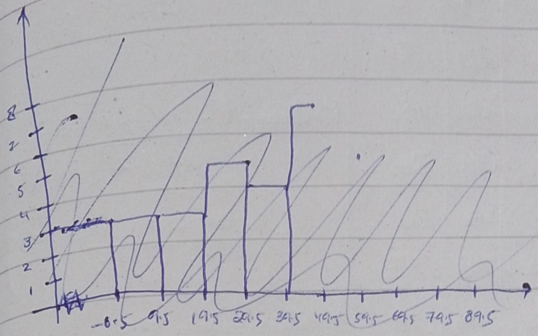
$N = \dfrac{80-0}{10}$

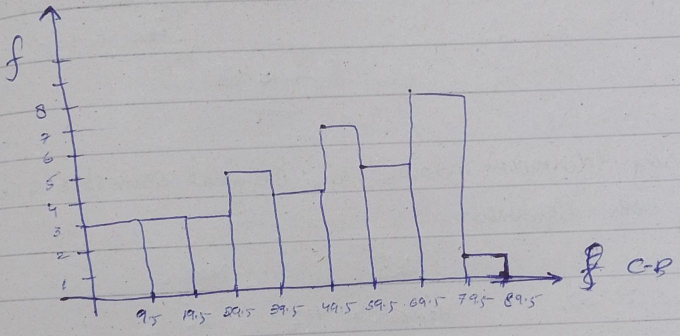
$N = 8$

i) Highest: 80
Lowest: 0

Stem & Leaf Plot

CI	Tally	frequency	Stem	Leaf	C.B
0 – 9	III	3	0	0 4 6	-0.5 – 9.5
10 – 19	III	3	1	0 0 4	9.5 – 19.5
20 – 29	III	3	2	0 2 6	19.5 – 29.5
30 – 39	卌	5	3	0 2 4 6 8	29.5 – 39.5
40 – 49	IIII	4	4	0 4 4 8	39.5 – 49.5
50 – 59	卌 II	7	5	0 0 2 2 4 6 8	49.5 – 59.5
60 – 69	卌	5	6	2 4 4 6 8	59.5 – 69.5
70 – 79	卌 III	8	7	0 0 2 8 4 4 6 6	69.5 – 79.5
80 – 89	I	1	8	0	79.5 – 89.5
		$\Sigma f = 39$			


8
7
6
5
4
3
2
1
-0.5
9.5
19.5
29.5
39.5
49.5
59.5
69.5
79.5
89.5


f
8
7
6
5
4
3
2
1
9.5
19.5
29.5
39.5
49.5
59.5
69.5
79.5
89.5
C-B

C.I	f	x	fx	CF
0-4	2	2	4	2
4-8	5	6	30	7
8-12	8	10	80	15
12-16	11	14	154	26
16-20	12	18	216	38
20-24	9	22	198	47
24-28	4	26	104	51
28-32	1	30	30	52
	$\Sigma f = 52$		$\Sigma fx = 816$	

Mean $= \bar{X} = \frac{\Sigma fx}{\Sigma f} = \frac{816}{52} = 15.$

Median $= l + \frac{h}{f}\left(\frac{\Sigma f}{2} - CF\right)$

$= 11.5 + \frac{5}{11}\left(\frac{52}{2} - 15\right)$

$\boxed{\text{Median} = \tilde{X} = 16.4}$

Mode $= l + \left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right) \times h$

$\boxed{\text{Mode} = 17}$

Q: Using information below, calc the final semester grade of both Students:

			Ali	Nasir
Homework	15%	Homework	92	100
Quiz	10%	Quiz	74	82
Lab	20%	Lab	83	95
Test	25%	Test	76	70
Final	30%	Final	88	76

For Ali: these are called weights

$\bar{X} = 0.15(92) + 0.10(74) + 0.20(83) + 0.25(75) + 0.30(88)$

$0.15 + 0.10 + 0.20 + 0.25 + 0.30 = 1$

$\boxed{\bar{X} = 83.2}$

→ Same for Nasir

Q: Rachel mixes 5 gallons of a 20% antifreeze solution with 10 gallons of a 50% antifreez sol to form a new solution with a different antifreez concentration
(a) Will the new sol have a concentration that is closer to 20% or 50%. (b) Calculate the concentration of the new antifreez sol.

S_1	S_2	S_3
$w_1 \to 5$ gal	$w_2 \to 10$ gal	15 gal
$x_1 \to 20\%$	$x_2 \to 50\%$	$\bar{X}_w \to$ Weighted mean.

(a) The new concentration would be closer to 50% as the volume of 50% conc. Solution is 10 gallons.

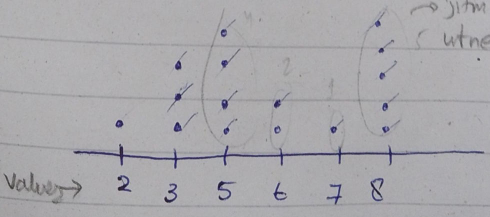
(b) $$X_w = \frac{W_1X_1 + W_2X_2}{W_1 + W_2} = \frac{5(20\%) + 10(50\%)}{15} = 40\%$$

"Dot Plots"

Q: Make Dot Plot of following data:
5, 8, 3, 6, 8, 5, 8, 2, 3, 8, 5, 8, 3, 5, 7, 6

Values	f
2	1
3	3
5	4
6	2
7	1
8	5

→ jitni frequency utne dots



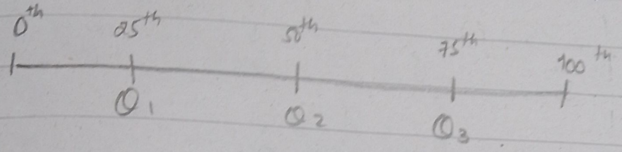
" STEM LEAF PLOT "

Q: 1.2, 2.3, 1.5, 2.4, 3.6, 1.8, 2.7, 3.2, 4.1, 2.9, 4.5, 7.6, 5.8, 9.3,
10.6, 12.4, 10.9

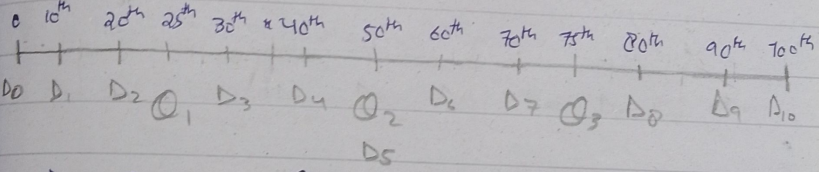
Stem	Leaf
1	2 5 8
2	3 4 7 9
3	2 6
4	1 5
5	8
6	→ leave empty
7	6
8	
9	3
10	6 9
11	
12	4

"QUARTILES, DECILES & PERCENTILES"

QUARTILES: Divides data into 4 equal parts.
DECILES: Divides data into 10 equal parts.
PERCENTILE: Divides data into 100 equal parts.



→ Data is dived into 4 equal parts.
→ "Q_2" is the median of whole data, Q_1 is median for lower half & Q_3 for upper half.



→ $D_5 = P_{50} = Q_2$, $D_4 = (P_{40})$ → percentile.

"The 70th percentile is the data point where 70% of the entire data is less than or equal to the data point. It also means that 30% of the data is greater than or equal to the data point."

$Q_1 = \frac{1}{4}(n+1)^{th}$ term $\rightarrow$ • $(3.25)^{th}$ term $\Rightarrow 3^{rd}$ term $+ 0.25$ ($4^{th} - 3^{rd}$)

$Q_2 = \frac{1}{2}(n+1)^{th}$ term • $(9.75)^{th}$ term $\Rightarrow 9^{th}$ term $+ 0.75$ ($10^{th} - 9^{th}$)

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

Q: 2, 3, 5, 7, 8, 10, 11, 13, 15, 16, 19 → Find Quartile.

→ First find Q_2 (as it is the median of data)
→ Here 10 is the median.

$$\underset{\underset{Q_2}{\downarrow}}{2, 3, 5, 7, 8} \;\Big|^{10}_{Q_2}\; \underset{\underset{Q_3}{\downarrow}}{11, 13, 15, 16, 19}$$

→ Another way is to find using percentile.
→ As $Q_1 = P_{25}$

→ $\boxed{P_K = \frac{K}{100}(n+1)}$ ↳ total no. of data

$P_{25} = \frac{25}{100}(11+1) = \frac{1}{4}(12) = 3$ → at 3^{rd} position from left

$\boxed{P_{25} = Q_1 = 5}$

$Q_2 = P_{50} = 10$

$P_{50} = \frac{500}{100}(11+1) = 6^{th}$

Q: 5, 6, 8, 9, 11, 12, 14, 15, 16, 17, 19, 20, 22, 23, 25
→ Median is 15.5 $\left(\because \frac{15+16}{2}\right) = 15.5$

$$\underset{}{5, 6, 8, 9} \;\Big|^{10}_{Q_1}\; 11, 12, 14, 15 \;\Big|^{15.5}_{Q_2}\; 16, 16, 17, 19 \;\Big|^{19.5}_{Q_3}\; 20, 22, 23, 25$$

For Deciles : $D_K = \frac{K}{10}(N+1)^{th}$ → group data ⇒ $D_K = l + \frac{h}{f}\left(\frac{K\Sigma f}{10} - CF\right)$

For Percentile : $P_K = l + \frac{h}{f}\left(\frac{K}{100}\Sigma f - C.F\right)$ ⇒ group data.

$L_{25} = \frac{25}{100}(16+1) = \frac{1}{4}(17) = 4.25$ [now we will take avg of 4th & 5th value]

→ 4 + (0.25)(5th − 4th)

$X_4 = 9$, $X_5 = 11$ ⇒ $\frac{9+11}{2} = 10$

$\boxed{P_{25} = 10} = Q_1$

Q_3: $L_{75} = \frac{75}{100}(16+1) = 12.75^{th}$

$\frac{X_{12} + X_{13}}{2} = \frac{19+20}{2} = 19.5 = Q_3 = P_{75}$

$D_6 = P_{60}$

$L_{60} = 60/100\,(16+1) = 10.2$ — here we will consider 16 as 10.2 is close to 10 so at 10th position it's 16

$P_{60} = \frac{X_{10} + X_{11}}{2} = (16+17)/2 = 16.5$

Calculating Percentile for "12":

⇒ $\boxed{\frac{x + 0.5y}{n}(100)}$ where x = no. of elements less than 12

y = freq of 12

n = total elements

$\frac{5 + 0.5(1)}{16}(100) = 34.375$

= 34th percentile, so $\boxed{P_{34} = 12}$

⇒ $P_? = 16$

$\frac{8 + 0.5(2)}{16}(100) = 56.25$

= 56th $\boxed{P_{56} = 16}$

Q: Create a cummulative freq. table using the data & use it to calc val of 4th, 7th, 3rd & 6th decile.

⇒ 3, 4, 7, 9, 2, 12, 3, 9, 6, 2, 9, 10, 3, 9, 4, 10, 3, 9, 4, 6

Value	f	RF	CRF
2	2	0.10	0.10
3	4	0.20	0.30
4	3	0.15	0.45
6	2	0.10	0.55
7	1	0.05	0.60
9	5	0.25	0.85
10	2	0.10	0.95
12	1	0.05	1.00
	20	1.00	

$D_4 = P_{40} = 4$

P_{40} matlab 0.4 CRF par jo value hogi. 0.3 par 3 ki values khtam hojayengi. 0.3 k aage se 4 ki shuroo hojayegi to laazmi P_{40} 4 hoga.

$D_7 = P_{70} = 9$

$D_3 = P_{30} = \frac{3+4}{2} = 3.5$

0.3 par left mein 3 hoga or right par 4 to avg lelia

$D_6 = P_{60} = \frac{7+9}{2} = 8$

9

2 2 | 3 3 | 3 3 | 4 4 | 4 6 | 6 7 | 9 9 | 9 9 | 9 10 | 10 12

D_0 D_1 D_2 D_3 D_4 D_5 D_6 D_7 D_8 D_9 D_{10}

D_3 = 3.5, D_4 = 4, D_6 = 8

"INTER QUARTILE RANGE & OUTLIERS"

0 — 25 — 50 — 75 — 100

Q_0 — Q_1 — Q_2 — Q_3 — Q_4

Q_1 ← IQR → Q_3

$IQR = Q_3 - Q_1$

outliers → $[Q_1 - 1.5\,IQR,\ Q_3 + 1.5(IQR)]$

numbers which are not in this range are outliers.

~~11~~, ~~31~~, 21, ~~19~~, ~~8~~, 54, 35, 26, 23, ~~13~~, 29, ~~17~~

	Q1		Q2		Q3	
8, 11, 13	|	17, 19, 21	|	23, 26, 29	|	31, 35, 54
	15		22		30	

IQR = 30 − 15 = 15

$[Q_1 - 1.5(IQR),\ Q_3 + 1.5(IQR)] = [-7.5,\ 52.5]$

So, $\boxed{54}$ is an <u>outlier</u>.

Box & Whisker Plots

→ Three parameters needed ⇒ Min, Q_1, Q_2, Q_3, Max

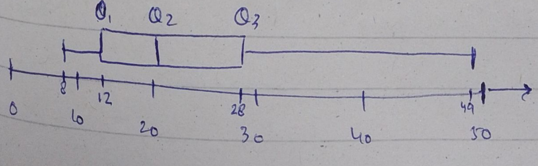
→ But Min & Max value should not be an outlier.

Q: 4, 22, 20, 14, 29, 8, 35, 27, 13, 49, 10, 24, 17

		12		20		28		
$\boxed{8}$	10, 11	|	13, 14, 17	|	22, 24, 27	|	29, 35	$\boxed{49}$
min		Q_1		Q_2		Q_3		max

IQR = 16 ⇒ $[12 - 1.5(16),\ 28 + 1.5(16)] = [-12,\ 52]$

→ both min & max lie in range



① First make a no line acc to our data.
② Now take 2 points one Q_1 & other Q_3 make a box
③ Dividing them into half gives Q_2.
③ Now make two lines for max & min values.

Q: 18, 34, 76, 29, 15, 41, 46, 25, 54, 38, 20, 32, 43, 22

33

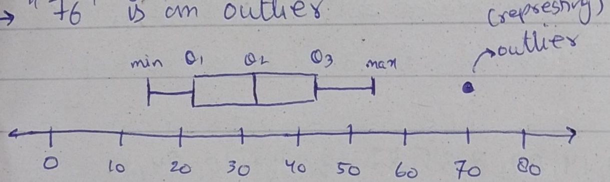
15 18 20 22 25 29 32 | 34 38 41 43 46 54 76

Q_1 Q_2 Q_2 outlier

$IQR = Q_3 - Q_1 = 43 - 22 = 21$

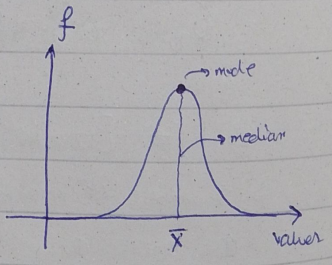
$[22 - 1.5(21),\ 43 + 1.5(21)] = [-9.5,\ 74.5]$

→ "76" is an outlier



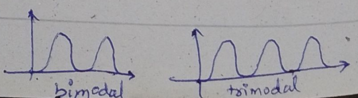
" SKEWNESS " → means lack of symmetry

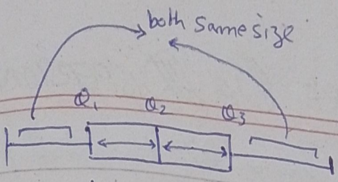
• Symmetric :



→ Median = Mode = Mean
→ Left side & right side of graph identical.

→ Uni modal

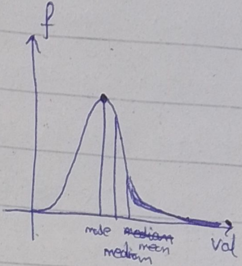




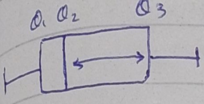
both halfs of B same size

$Q_3 - Q_2 = Q_2 - Q_1$

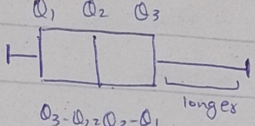
- Skewed Right / Positive Skewed:



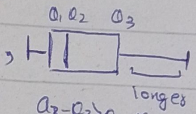
- $\boxed{\bar{X} > \text{Median} > \text{Mode}}$



$Q_3 - Q_2 > Q_2 - 1$



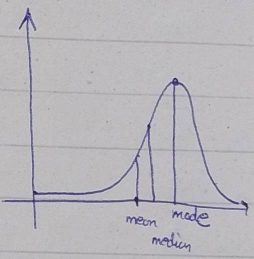
$Q_3 - Q_2 = Q_2 - Q_1$

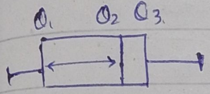


$Q_3 - Q_2 > Q_2 - Q_1$

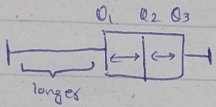
- Skewed Left / Negetaive Skewed:

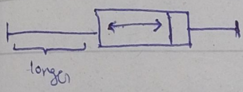
- $\boxed{\bar{X} < \text{Median} < \text{Mode}}$





$Q_2 - Q_1 > Q_3 - Q_2$





"MEASURE OF DISPERSION"

→ Dispersion measures the extent to which the items vary from from some central value (mean, median, mode) central val / central tendency

→ Dispersion also called scatter, spread or variation. a single value which represents the whole items.

→ If we know a center value then also we can't identify how much the items are spread from central val.

	Series A	Series B	Series C
	100	98	1
	100	99	2
	100	100	3
	100	101	4
	100	102	490
Total	500	500	500
$\bar{X}$	$\boxed{100}$	$\boxed{100}$	$\boxed{100}$
Dispersion	0	less	very hight

→ A data having high dispersion is unstable whereas low dispersion data is stable.

Measure of Dispersion

1. Range
2. Quartile Deviation
3. Mean Deviation
4. Standard Deviation
5. Variance

Absolute Measure
(Unit same)

1. Coefficent of Range
2. Coefficent of Qu-Dev-.
3. Coefficent of Mean Dev-
4. Coefficent of Std. Deve
5. Coefficent of Variance

Relative Measure
(Unit free)

These are mostly used to compare two data's having diff units
~~A data in wh~~ To compare data's which have a great diff in central values.

1. Range :

Range = L − S
L = Largest Obse.
S = Smallest Obser.

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

2. QUARTILE DEVIATION :

→ also called semi inter Quartile Range.

• 139, 140, 140, 141, 141, 142, 142, 143, 143, 144, 145

$$Q.D = \frac{Q_3 - Q_1}{2} \rightarrow \frac{IQR}{2}$$

$Q_1 = \frac{1}{4}(n+1)^{th} = \frac{1}{4}(12+1)^{th}$ term = 3.25^{th} term

$= 3^{rd}$ term + $(0.25)(4^{th} - 3^{rd})$

$= 140 + (0.25)(141 - 140)$

$\boxed{Q_1 = 140.25}$

$Q_3 = \frac{3}{4}(n+1)^{th}$ term $= (9.75)^{th}$ term $= 9^{th}$ term $+ 0.75(10^{th} - 9^{th})$

$= 143 + (0.75)(144 - 143)$

$= 143 + 0.75$

$\boxed{Q_3 = 143.75}$

$Q.D = \frac{143.75 - 140.25}{2}$

$\boxed{Q.D = 1.75}$

Q:

class	f	C.F
0 – 10	3	3 C.F
l 10 – 20	9 f	12 — Q_1
20 – 30	12	24 C.F
l 30 – 40	11 f	35 — Q_3
40 – 50	5	40
50 – 60	6	46

$\Sigma f = 46$

For Grouped Data:

$Q_1 = l + \frac{h}{f}\left(\frac{\Sigma f}{4} - C.F\right)$ (class selection), $Q_3 = l + \frac{h}{f}\left(\frac{3\,\Sigma f}{4} - C.F\right)$ (class selection)

$Q_1 \Rightarrow$

class $= \frac{\Sigma f}{4} = 11.5$ (2^{nd} class)

CF mein 11.5 check krenge kaha lie kreg

$Q_1 = 10 + \frac{10}{9}\left(\frac{46}{4} - 3\right)$

$\boxed{Q_1 = 19.44}$

$Q_3 \Rightarrow$ class $= \frac{3(46)}{4} = 34.5$

$= 30 + \frac{10}{11}(34.5 - 24)$

$\boxed{Q_3 = 39.54}$

$Q.D = \frac{Q_3 - Q_1}{2} = \boxed{10.05}$

CF se values ka order pta chal jayega.

x	f
5	2
9	2
2	1
10	5
15	3

→ Rearrange ascending order

x	f	CF	
2	1	1	→ 1st term is 2 val.
5	2	3	→ 2nd se 3rd tak is 5 val
9	2	5	→ 4th se 5th tak is 9 val
10	5	10	→ 6th se 10th tak 10 val
15	3	13	→ 11th se 13th tak 15 val

$Q_1 = \frac{1}{4}(n+1)^{th} = 3.5^{th} = 3^{rd} + 0.5(4^{th} + 3^{rd})$

$= 5 + (0.5)(9-5)$

$\boxed{Q_1 = 7}$

$Q_3 = \frac{3}{4}(n+1)^{th} = 10.5 = 10^{th} + 0.5(11^{th} + 10^{th})$

$= 10 + (0.5)(15-10)$

$\boxed{Q_3 = 12.5}$

$\boxed{Q.D = 2.75}$

③ STANDARD DEVIATION : (σ) → Root Mean Square Deviation

$$\sigma = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n-1}} \qquad \bar{X} = \text{Mean}$$

④ VARIANCE (σ^2):

$$\text{Variance} = \frac{\Sigma(X-\bar{X})^2}{n-1}$$

$$\text{Coefficient of Variance (CV)} = \frac{\sigma}{\bar{X}} \times 100$$

Q: 10, 12, 13, 15, 20

$\bar{x} = \frac{\Sigma x}{n} = \frac{10+12+13+15+20}{n}$ (mean)

$\boxed{\bar{x} = 14}$

x-14

x	$x-\bar{x}$	$(x-\bar{x})^2$
10	-4	16
12	-2	4
13	-1	1
15	1	1
20	6	36
		$\Sigma(x-\bar{x})^2 = 58$

$\sigma = \sqrt{\frac{58}{54}} = 3.$

$\boxed{\sigma = 3.8}$

$\sigma^2 = (3.4)^2$

$\boxed{\sigma^2 = 14.5}$

(variance)

Coefficent of Variance = $\frac{\sigma}{\bar{x}} \times 100$

$\boxed{CV = 27.14\%}$

Q: 48, 43, 65, 57, 31, 60, 37, 48, 59, 78

$\bar{x} = \frac{\Sigma x}{n} = \frac{526}{10}$

$\boxed{\bar{x} = 52.6}$

x	$x-\bar{x}$	$(x-\bar{x})^2$
31	-21.6	466.56
37	-15.6	243.36
43	-9.6	92.16
48	-4.6	21.16
48	-4.6	21.16
57	4.4	19.36
59	6.4	40.96
60	7.4	54.76
65	12.4	153.76
78	25.4	645.16
		$\Sigma = 1758.4$

$$\sigma = \sqrt{\frac{1758.4}{10-1}}$$

$$\boxed{\sigma = 13.977}$$

$$\boxed{\sigma^2 = 195.377}$$

$$\boxed{CV = 26.57\%}$$

- Z-Score:

$$Z = \frac{X-\bar{X}}{S}$$

S → standard deviation

→ The Z-score is the no. of standard deviation a data value is from the mean.

→ A data value is considered extreme outlier if its Z-score is less than -3.0 or greater than +3.0.

→ The larger the absolute value of Z-score, the farther the data value from mean.

The modal class is the class having largest frequency. Sometimes the midpoint is used rather than boundaries.

Q: Suppose mean of a data is 490, with $\sigma^2 = 100$. Calc Z-score of data val 620.

$$Z = \frac{X - \bar{X}}{\sigma^2} = \frac{620 - 490}{100} = \frac{130}{100}$$

$\boxed{Z = 1.3}$ → Score of 620 is 1.3 std. dev above from mean, so not an outlier

Sample Statistics vs Population Parameters

Measure	Population Para.	Sample Stats
Mean	μ	$\bar{X}$
Variance	σ^2	S^2
St. Deviation	σ	S

- VARIANCE & STD. DEVIATION FROM GROUPED DATA:

$$S^2 = \frac{n(\Sigma f.X^2) - (\Sigma f.X)^2}{n(n-1)}$$

where X = midpoint
f = frequency
n = Sum of freq.

$$S = \sqrt{\frac{n(\Sigma f.X^2) - (\Sigma fX)^2}{n(n-1)}}$$

Class	f	X	fX	fX^2
5.5-10.5	1	8	8	64 (1×8^2)
10.5-15.5	2	13	26	338 (2×13^2)
15.5-20.5	3	18	54	972
20.5-25.5	5	23	115	2,645
25.5-30.5	4	28	112	3,136
30.5-35.5	3	33	99	3,267
35.5-40.5	2	38	76	2,888
	$\Sigma=20$		$\Sigma=490$	$\Sigma=13310$

$$s^2 = \frac{n(\Sigma f \cdot X^2) - (\Sigma fX)^2}{n(n-1)}$$

$$= \frac{20(13,310) - (490)^2}{20(20-1)}$$

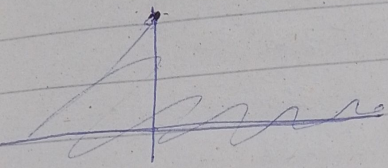
$$\boxed{s^2 = 68.7}$$

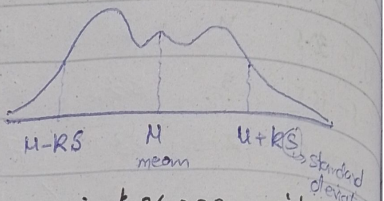
$$\boxed{s = 8.3}$$

(it can be applied to any curve or distribution.)

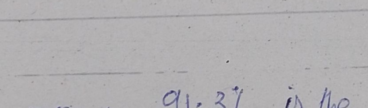
"CHEBYSHEV'S THEOREM"

It gives the proportion of data that is within "K" standard devians of the mean. (where $K > 1$)





Q: The avg price of a new car is \$36,000 with a std dev of \$4,100. What is the min percentage of cars that should sell between \$22,000 & \$50,000



$\Rightarrow$ $\mu + KS = 50{,}000$

$36000 + K(4100) = 5000$

$\boxed{K = 3.4146\ldots}$

$1 - \frac{1}{K^2} = 1 - \frac{1}{(3.416)^2} = .913$

means 91.3% is the ~~mean~~ minimum % of cars that should be sell b/w \$22,000 & \$50,000.

$K = 2 \Rightarrow 1 - \frac{1}{2^2} = 75\%$ [means 75% data lies within 2 standard deviation]

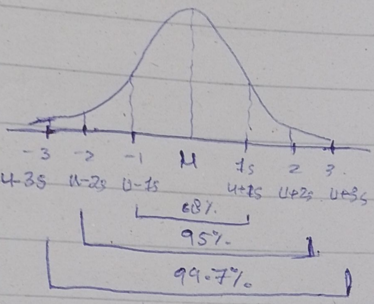
$K = 3 \Rightarrow 1 - 1/3^2 = 89\%$ [means 89% data —]

" EMPERICA RULE "

It states that within 1 standard deviation of mean (both leftside & right side) there is about 68% of the data.

- 2-standard deviation = 95% of data
- 3 - standard deviation = 99.7% of data.

- It is applied to bell-shaped curved or a normal distribution.



: The heights of women follow a bell shaped distribution with a mean of 160 cm & a std. dev of 7.5 cm.
What is the approximate % of women b/w 137.5 cm & 182.5 cm.

$\bar{X} = 160$ cm

$s = 7.5$ cm

$\bar{X} \pm ks \Rightarrow 160 + k(7.5) = 182.5$

$\boxed{k = 3}$

99.7% of women b/w 137.5 —