॥ ॐ नमः शिवाय ॥

Unit - I

★ **Discrete and continous rate** :- Discrete data is the information than can take only certain values.
Eg:- shoes size
These type of data is often represented using tally charts, bar charts or pie charts.
Continous data is the data that can take any value
Eg:- Speed, height, weight, temperture.
This data is shown on, line graph, histogram & scatter plot

★ **Measure of central tendency** :- One of the most important objective of statical analysis is to get one single value that describes the characteristics of the entire mass of data such a value is called a central value or average

★ **Objectives** :- 1) Measures of central value by condensing the mass of data into one single value gives the words view of entire data.
2) Measure of central value by recording the marks data to one single figure enables comparisions to be made

★ **Requisities of good average** :-
1) Easy to understand
2) Should be simple to compute so that it can be used widely
3) It should be based on all the items.
4) should be well defined.
5) should be capable for further treatment
6) It should have sampling stability.

★ **Types of central tendency** :-
1) **Arithematic mean** :- Arithematic mean of the set of observations is there sum by the no of observations.
if $x_1, x_2, x_3, ---, x_n$ given n observations then the mean is given by $\bar{x} = \frac{x_1 + x_2 + --- + x_n}{n}$

problem 1) The monthly income 12 families is given below
280 180 96 98 104 85 80 94 100 75 600 200
find the arithematic mean

$\Rightarrow \bar{x} = \frac{\Sigma x_i}{n} = \frac{1992}{12} = 166$ //

In case of freq distribution

$x : x_1\ x_2\ x_3 - - - x_n$

$f : f_1\ f_2\ f_3 - - - f_n$

$\bar{x} = \frac{f_1x_1 + f_2x_2 + - - + f_nx_n}{f_1 + f_2 + - - f_n}$

$\therefore \boxed{\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f}} \quad \boxed{N = \Sigma f}$

2) find the arithematic mean of

x	1	2	3	4	5	6	7
f	5	9	12	17	14	10	6

$\Rightarrow$

$\bar{x} = \frac{5 + 18 + 36 + 68 + 70 + 60 + 42}{73} \quad \bar{x} = 4.095$ //

3) calculate mean marks for the foll. data

Marks :	0-10	10-20	20-30	30-40	40-50	50-60
No of stu:	12	18	27	20	17	6

$\Rightarrow$

Marks	No of stu (f)	Mid pt (x)
0-10	12	5
10-20	18	15
20-30	27	25
30-40	20	35
40-50	17	45
50-60	6	55

$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} \quad \bar{x} = \frac{60 + 270 + 675 + 700 + 765 + 330}{100}$

$\bar{x} = \frac{2800}{100} \quad \bar{x} = 28$ //

★ Step deviation method :- when x & f values are large the previous used formulae become time consuming hence in case of group or continous freq. distribution the operation can be reduced to a greater extend by taking $d = \frac{x - A}{h}$ // $d = \frac{x - A}{h}$ where A - assumed mean / aribary value

h - common magnitude of class interval & arithematic mean is given by

$\boxed{\bar{x} = A + \frac{\Sigma f \cdot d}{\Sigma f} \times h}$

1) C.I	freq	Mid. pt (x)	f.x
0-8	8	4	32
8-16	7	12	84
16-24	16	20	320
24-32	24	28	672
32-40	18	36	576
40-48	7	44	280

$\Rightarrow \bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{32 + 84 + 320 + 672 + 576 + 280}{80} = \frac{1964}{80} = 24.55$ //

2) C.I	f	Mid. pt (x)	$d = \frac{x - A}{h}$	f.d
0-8	8	4	-3	-24
8-16	7	12	-2	-14
16-24	16	20	-1	-16
24-32	24	28	0	0
32-40	18	36	1	18
40-48	7	44	2	14
	80			-22

Assumed mean (A) = 28

h = 8

$\bar{x} = A + \frac{\Sigma f \cdot d}{N} \times h \quad \bar{x} = 28 + \frac{(-22)}{80} \times 8$

$\bar{x} = 28 - \frac{176}{80} \quad \bar{x} = 25.8$ //

★ Properties of mean :- i) Algebraic sum of the deviation of the set of values from there arithematic mean is zero

2) if $\bar{x}_i$ for $i = 1, 2 \dots k$ are the means of k component series of sizes n_i for $i = 1, 2 \dots k$ then the mean $\bar{x}$ is obtained by combining the component series to obtain the formula

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3 + \dots + n_k\bar{x}_k}{n_1 + n_2 + n_3 + \dots}$$

q:- 1) The mean age of 100 workers in a factory running two ships of 60 & 40 workers with the mean of 38 ₹, the mean age of 60 workers is ₹40 find the mean age of 40 workers in the night shift

⇒ $n_1 = 60$, $n_2 = 40$

$\bar{x} = 38$

$\bar{x}_1 = 40$, $\bar{x}_2 = ?$

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} \qquad \bar{x} = \frac{60(40) + 40(\bar{x}_2)}{60 + 40} \qquad \bar{x}(100) = 2400 + \bar{x}_2(40)$$

$38 \times 100 = 2400 + 40\bar{x}_2 \qquad 1400 = 40(\bar{x}_2) \qquad \therefore \bar{x}_2 = \frac{1400}{40}$ //

2) The avg weakly salary of male employee in a firm was 5200 & that of female of 4200 the mean salary of all the emp was 5000. find the percentage of male & female employee

⇒ $\bar{x}_1 = 5200$, $\bar{x}_2 = 4200$, $\bar{x} = 5000$

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} \; ; \quad 5000 = \frac{n_1(5200) + n_2(4200)}{n_1 + n_2}$$

$5000n_1 + 5000n_2 = 5200n_1 + 4200n_2 \qquad \frac{n_1}{n_2} = \frac{800}{200} = \frac{4}{1}$

$800n_2 = 200n_1$

$n_1 = \frac{4}{4+1} \times 100 = 80\%$ //, $n_2 = \frac{1}{4+1} \times 100 = 20\%$ //

★ Median :- Median is the middle item of the series arranged in ascending or descending order of magnitude

(i) if the no. of observation is odd the median is the middle value.

(ii) If the no of observations are even then median is obtained by taking arithematic mean of middle terms

Eg:- 1) 25 20 15 45 18 78 10 38 12

⇒ 10 12 15 18 20 25 38 45 78 ⇒ median = 20 //

2) 8 20 50 25 15 30

⇒ 8 15 20 25 30 50 median = 22.5 //

b) For continous frequency distribution :- The median is obtained by using formula :-

$$\text{Median} = l + \frac{h}{f}\left[\frac{N}{2} - c\right]$$ where l ⇒ lower limit of the median class

f - frequency of the median class, h ⇒ magnitude of median

c → Is cumulative freq. of the class preceding the median class

$N = \Sigma f$

1) find the median of the foll. distribution of class interval

C.I	2000-3000	3000-4000	4000-5000	5000-6000	6000-7000
f	3	5	20	10	5

⇒ $N/2 = 43/2 = 21.5$

C.I	f	C.f
2000-3000	3	3
3000-4000	5	8
4000-5000	20	28
5000-6000	10	38
6000-7000	5	43

median = 4000 - 5000

$l = 4000$, $h = 1000$, $f = 20$, $c = 8$

$\text{median} = l + \frac{h}{f}\left[\frac{N}{2} - c\right]$

$= 4000 + \frac{1000}{20}[21.5 - 8] = 4675$ //

2) find the median of the foll.

C.I	0-7	7-14	14-21	21-28	28-35	35-42	42-49
f	19	25	36	22	51	43	28

$\Rightarrow \frac{N}{2} = \frac{224}{2} = 112$

C.I	f	c.f
0-7	19	19
7-14	25	44
14-21	36	80
21-28	22	102
28-35	51	153
35-42	43	196
42-49	28	224

median class = 28-35

$l = 28$, $h = 7$, $f = 51$, $c = 80$

$\text{Median} = l + \frac{h}{f}\left[\frac{N}{2} - c\right]$

$= 28 + \frac{7}{51}[112 - 80]$

median = 42·27 //

3) In a factory employing 3000 persons in a day 5% work less than 3 hrs, 580 work from 3·01 to 4·50 hrs, 30% work from 4·51 to 6·00 hrs, 500 work from 6·01 to 7·50 hrs 20% work from 7·51 to 9·00 hrs & the rest work 9·01 or more what is the median hrs at work.

$\Rightarrow$

C.I	f	c.f	class boundaries
less than 3	150	150	less than 3·005
3·01-4·50	580	730	3·005 - 4·505
4·51-6·00	900	1630	4·505 - 6·005
6·01-7·50	500	2130	6·005 - 7·505
7·51-9·00	600	2730	7·505 - 9·005
more 9·01	270	3000	more than 7·005

$\frac{N}{2} = \frac{5f}{2} = \frac{3000}{2} = 1500$

median class = 4·505 - 6·005

$\text{median} = l + \frac{h}{f}\left[\frac{N}{2} - c\right]$

$l = 4·505$, $h = 1·5$, $f = 900$

$c = 730$

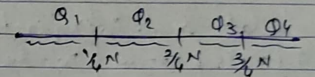
$\text{median} = 4·505 + \frac{1·5}{900}[1500 - 73]$

median = 5·788 //

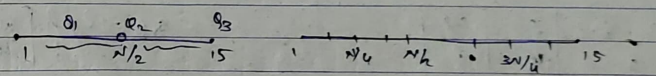
★ Measure of dispersion :- consider scores of 2 criteria A & B ; A = {34, 6, 23, 81, 76, 2, 12} ; B = {28, 41, 36, 33, 38, 27, 35} for batsman A the score differ 34 (mean) and differ large but batsman B the score differ from 34 & diff is small just based on this we can divide who is more consistent ie the property which denotes the extend to which the indivually values of the variables are dispersed about central tendency is called dispersion

★ Quartail deviations (QD) :- It is also called as semi-interquartail deviation

$\left[QD = \frac{Q_3 - Q_1}{2}\right]$



$\text{Median} = l + \frac{h}{f}\left[\frac{N}{2} - c\right]$



$Q_1 = l + \frac{h}{f}\left[\frac{N}{4} - c\right]$ $\quad Q_3 = l + \frac{h}{f}\left[\frac{3N}{4} - c\right]$

quartile deviation $(QD) = \frac{Q_3 - Q_1}{3}$

where $Q_1 \to l + h/f\left[\frac{N}{4} - c\right]$ $\quad Q_3 \to l + h/f\left[\frac{3N}{4} - c\right]$

This gives only the absolute measure of this version but for the comparative study of variability b/w two distribution, relative measure known as coefficient of quartile means

$\left[\text{coeff. } Q.D = \frac{Q_2 - Q_1}{Q_2 + Q_1}\right]$

1) Calculate the quartile deviation & coeff. quartile deviation for the foll. ungrouped items 20 28 40 12 30 15 50

⇒ 12 15 20 28 30 40 50

Q_1 = size of $\left(\frac{n+1}{4}\right)^{th}$ item = size of 2nd item = 15

Q_3 = size of $\left(\frac{3(n+1)}{4}\right)^{th}$ item = size of 6th item = 40

$Q.D = \frac{Q_3 - Q_1}{2} = \frac{40-15}{2} = \frac{25}{2} = 12.5$ //

coeff of $Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{40-15}{40+15} = \frac{25}{55} = \frac{5}{11} = 0.45$

2) calculate semi-interquartile range & coeff of the foll data

C.I	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
f	11	18	25	28	30	33	22	15

⇒

C.I	f	C.f
0-10	11	11
10-20	18	29
20-30	25	54
30-40	28	82
40-50	30	112
50-60	33	145
60-70	22	167
70-80	15	182

$Q_1 = l + \frac{h}{f}\left[\frac{N}{4} - c\right]$

$Q_3 = l + \frac{h}{f}\left[\frac{3N}{4} - c\right]$

for Q_1

$\frac{N}{4} = \frac{182}{4} = 45.5$

median class = 20-30, $f = 25$, $h = 10$, $l = 20$, $c = 29$

$\therefore Q_1 = 20 + \frac{10}{25}[45.5 - 29]$

$Q_1 = 26.6$

for Q_3 ⇒ $\frac{3N}{4} \Rightarrow 136.5$

median class = 50-60, $f = 33$, $h = 10$, $l = 50$, $c = 112$

$Q_3 = 50 + \frac{10}{33}[136.5 - 112]$ $\quad Q_3 = 57.42$

$Q.D = \frac{Q_3 - Q_1}{2} = \frac{57.42 - 26.6}{2} = 15.41$

coeff $Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{57.42 - 26.6}{57.42 + 26.6} = 0.3668$

Hw

3) find Q.D for data

C.I	40-49	50-59	60-69	70-79	80-89	90-99
f	306	192	144	96	42	34

⇒ Q.D = 10.31, c.Q.D = 0.189

★ Standard Deviation :- this gives the +ve sq. root of the arithmetic mean of the sq's of deviations of given observation from there arithmetic mean.

$$\left[S.D = \sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{N}} \Big/ \sigma = \sqrt{\frac{\Sigma x_i^2}{N} - \left(\frac{\Sigma x}{N}\right)^2}\right]$$

1) calculate the standard deviation the ungrouped data

8 9 15 23 5 11 19 8 10 12

$\bar{x} = \frac{\Sigma x}{N} = 12$

x	$x - \bar{x}$	$(x - \bar{x})^2$
8	-4	16
9	-3	9
15	3	9
23	11	121
5	-7	49
11	-1	1
19	7	49
8	-4	16
10	-2	4
12	0	0

$S.D = \sqrt{\frac{\Sigma(x - \bar{x})^2}{N}} = \sqrt{\frac{274}{10}} = 5.23$ //

2) given the foll. information find S.D $n = 10$, $\Sigma x = 60$, $\Sigma x^2 = 1000$

⇒ $\sigma = \sqrt{\frac{\Sigma x^2}{N} - \left(\frac{\Sigma x}{N}\right)^2} = \sqrt{\frac{1000}{10} - \left(\frac{60}{10}\right)^2} = \sqrt{100 - 36} = 8$ //

2) Calculate the S.D for the foll. data

Size of i'th	6	7	8	9	10	11	12
freq	3	6	9	13	8	5	4

⇒

C.I	f	$x-\bar{x}$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
6	3	-3	9	27
7	6	-2	4	24
8	9	-1	1	9
9	13	0	0	0
10	8	1	1	8
11	5	2	4	20
12	4	3	9	36
				124

$\bar{x} = \frac{18+42+72+117+80+55+48}{48}$ $\bar{x} = 9$

$\therefore \sigma = \sqrt{\frac{\Sigma (x-\bar{x})^2 f}{\Sigma f}} = \sqrt{\frac{124}{48}} = 1.607$ //

3) find the S.D for foll. data.

C.I	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
f	15	15	23	22	25	10	5	10

⇒

C.I	f	mid pt (x)	$x-\bar{x}$	$(x-\bar{x})^2 f$
0-10	15	5	-30.16	13644.384
10-20	15	15	-20.16	6096.384
20-30	23	25	-10.16	2374.1888
30-40	22	35	0.16	0.5632
40-50	25	45	9.84	2420.64
50-60	10	55	19.84	3936.256
60-70	5	65	29.84	4452.125
70-80	10	75	39.84	15872.25
				49801.78

$\bar{x} = \frac{\Sigma xf}{\Sigma f} = \frac{75+225+375+770+1125+550+325+750}{}$

$\bar{x} = 35.16$

$\sigma = 19.75$

$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}}$ $\sigma = \sqrt{\frac{49801.78}{125}} = 19.7588$ //

3) The analysis of the result of a ~~by~~ survey of 150 families gave an average monthly expenditure of ₹120 on food items with SD of ₹150 after the analysis completed it was noted that fig. recorded for the household was wrongly taken as ₹15 instead of ₹105 determine correct value of avg expenditure & its std. deviation

⇒ $n = 150$, $\bar{x} = 120$, $\sigma = 15$ (wrong value) correct value = 105

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

$120 = \frac{\Sigma x}{150}$ $\Sigma x = 18000$ //

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

$15 = \sqrt{\frac{\Sigma x_i^2}{150} - (120)^2}$ $225 + 14400 = \frac{\Sigma x_i^2}{150}$

$\Sigma x_i^2 = 2193750$ //

$\Sigma x = 18000 - 15 + 105 = 18090$

$\bar{x} = \frac{\Sigma x}{n} = \frac{18090}{150} = 120.6$ //

$\Sigma x_i^2 = 2193750 - 15^2 + 105^2 = 2204550$

$\sigma = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$ $\sigma = \sqrt{\frac{2204550}{150} - \left(\frac{18090}{150}\right)^2}$

$\sigma = 8.97$ // (12.35)

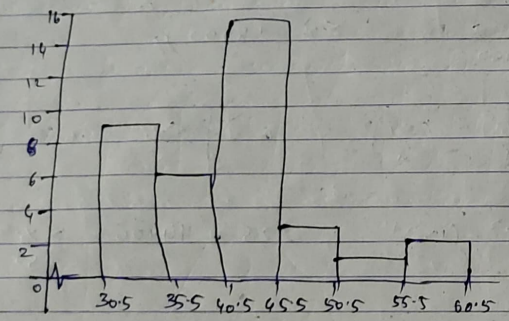
★ Graphical Representation of Data :-

★ Histogram :-

1) Draw the Histogram for the following data

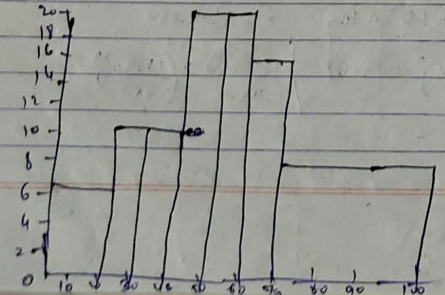
weight (kg)	No of students
30.5 - 35.5	9
35.5 - 40.5	16
40.5 - 45.5	15
45.5 - 50.5	3
50.5 - 55.5	1
55.5 - 60.5	2

⇒



2) A teacher wanted to analyze the performance of 2 section of students in a maths test of 100 marks the following data was formed

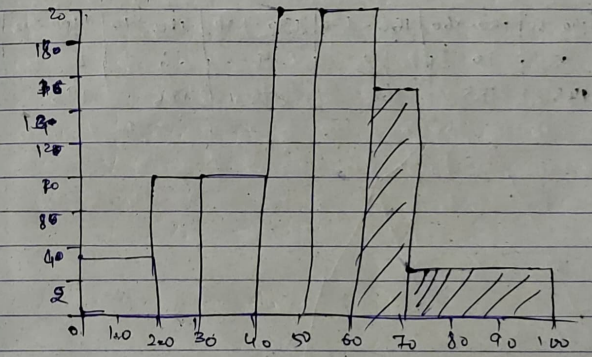
marks	0-20	20-30	30-40	40-50	50-60	60-70	70 & above
No of st	7	10	10	20	20	15	8



Since the intervals are not same here

(i) select a class interval of min class size here the class size is 10. the length of the rectangle are modified to be proportion to the class size.

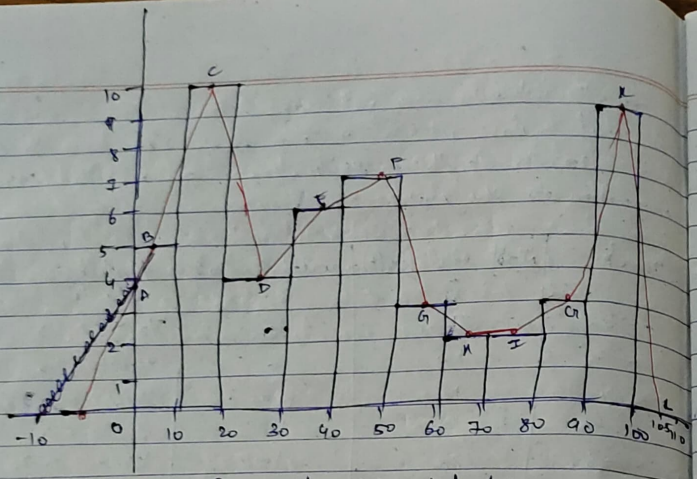
Marks	No of students (f)	width of class	length of rectangle
0-20	7	20	$7/20 \times 10 = 3.5$
20-30	10	10	$10/10 \times 10 = 10$
30-40	10	10	$10/10 \times 10 = 10$
40-50	20	10	$20/10 \times 10 = 20$
50-60	20	10	$20/10 \times 10 = 20$
60-70	15	10	$15/10 \times 10 = 15$
70 & above (70-100)	8	30	$8/30 \times 10 = 2.67$



★ Frequency polygon :-

q) 1) Draw the freq. polygon for the foll. data.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No of st	5	10	4	6	7	3	2	2	3	9

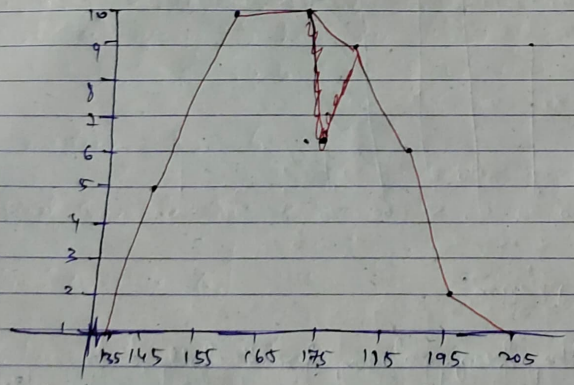


⇒

Now draw freq. poly for 1st problem of histogram

2) draw freq. poly without using histogram

Cost of living index	140-150	150-160	160-170	170-180	180-190	190-200
No of weeks	5	10	10	9	6	2
mid pt	145	155	165	175	185	195



★ Steam leaf graph :-

1) Draw the steam leaf graph for the foll data.

15 27 8 17 13 22 24 25 30 36 38 32 32 28 43 7.

⇒

		steam	leaf
7 8			
13 13 15 17	⇒	0	7 8
22 24 25 27 28		1	3 3 5 7
32 32 36 38		2	2 4 5 7 8
43		3	2 2 6 8
		4	3

2) Draw the steam leaf graph for the foll data.

1.2, 2.3, 2.5, 2.4, 2.6, 1.8, 2.7, 3.2, 4.1, 2.9, 4.5, 7.6, 5.8, 9.3, 10.6, 12.4, 10.9

⇒

		steam	leaf
1.2 1.8		1	0.2 0.8
2.3 2.4 2.5 2.6 2.7 2.9		2	0.3 0.4 0.5 0.6 0.7 0.9
3.2		3	0.2
4.1 4.5	⇒	4	0.1 0.5
5.8		5	0.8
7.6		7	0.6
9.3		9	0.3
10.6 10.9		10	0.6 0.9
12.4		12	0.4

★ box plot

1) draw the box plot of the foll. data 10, 12, 11, 15, 11, 14, 13, 17, 12, 22, 14, 11.

⇒ 10, 11, 11, 11, 12, 12 | 13, 14, 14, 15, 17, 22

(1st quartile | 2nd | 3rd)

Median = Q_2 = 12.5

First quartile = $Q_1 = \frac{11+11}{2} = 11$

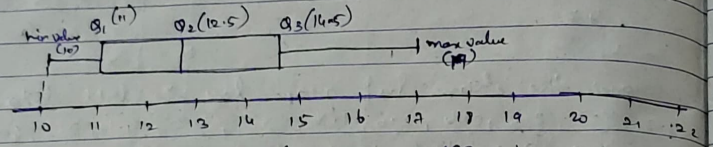
Third quartile = $Q_3 = \frac{14+15}{2} = 14.5$

Inter Quartile range (IR) = $Q_3 - Q_1 = 14.5 - 11 = 3.5$ //

Range $[Q_1 - (1.5 \times IR),\ Q_3 + (1.5 \times IR)]$

$[11 - (1.5 \times 3.5),\ 14.5 + (1.5 \times 3.5)]$

$[5.75,\ 19.75]$ //



min value = 10, max value = 17, outlier is = 2

2) The owner of the restaurant wants to find more about were his workers are coming from. He decided together data to gathere about distances that people comuted to the rest. people reported the foll. dist. travelled

14, 6, 3, 2, 4, 15, 11, 8, 1, 7, 2, 1, 3, 4, 10, 12, 20

Create a graph to help him spread of dists

⇒ 1, 1, 2, 2, 3, 3, 4, 4 | 6 | 7, 8, 10, 11, 12, 14, 15, 20

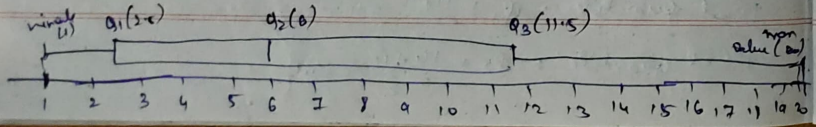
median = Q_2 = 6

First Quartile Q_1 = 2.5

third Quartile Q_3 = 11.5

Inter Quartile range (IR) = $Q_3 - Q_1 = 11.5 - 2.5 = 9.0$

Range $[Q_1 - (1.5 \times 9),\ Q_3 + (1.5 \times 9)] = [-11,\ 25]$



min value = 10 max value = 20 outlier is None

(3 types → less than, more than, Both combination).

★ Ogive curves :-

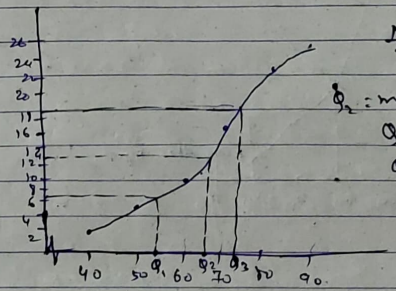
1) plot the Ogive curve for the foll. data.

C.I	30-40	40-50	50-60	60-70	70-80	80-90
f	2	3	5	7	6	2

⇒

a) less than curve

C.I	f	less than	C.f
30-40	2	40	2
40-50	3	50	5
50-60	5	60	10
60-70	7	70	17
70-80	6	80	23
80-90	2	90	25



$\frac{N}{2} = \frac{\Sigma f}{2} = \frac{25}{2} = 12.5$

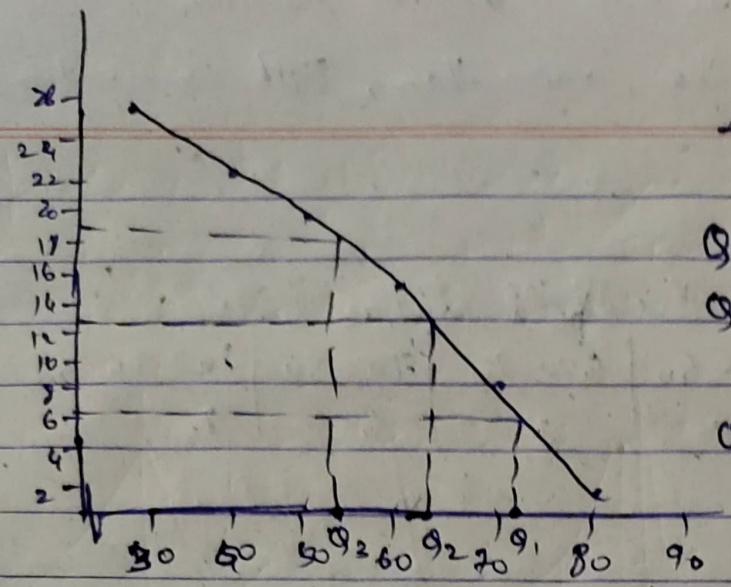
Q_2 : median = 64 (approx).

$Q_1 = N/4 = 25/4 = 6.2 = 53N$

$Q_3 = 3N/4 = 18.6 = 72N$

b) More than curve

C.I	f	more than	C.f
30-40	2	30	25
40-50	3	40	23
50-60	5	50	20
60-70	7	60	15
70-80	6	70	8
80-90	2	80	2



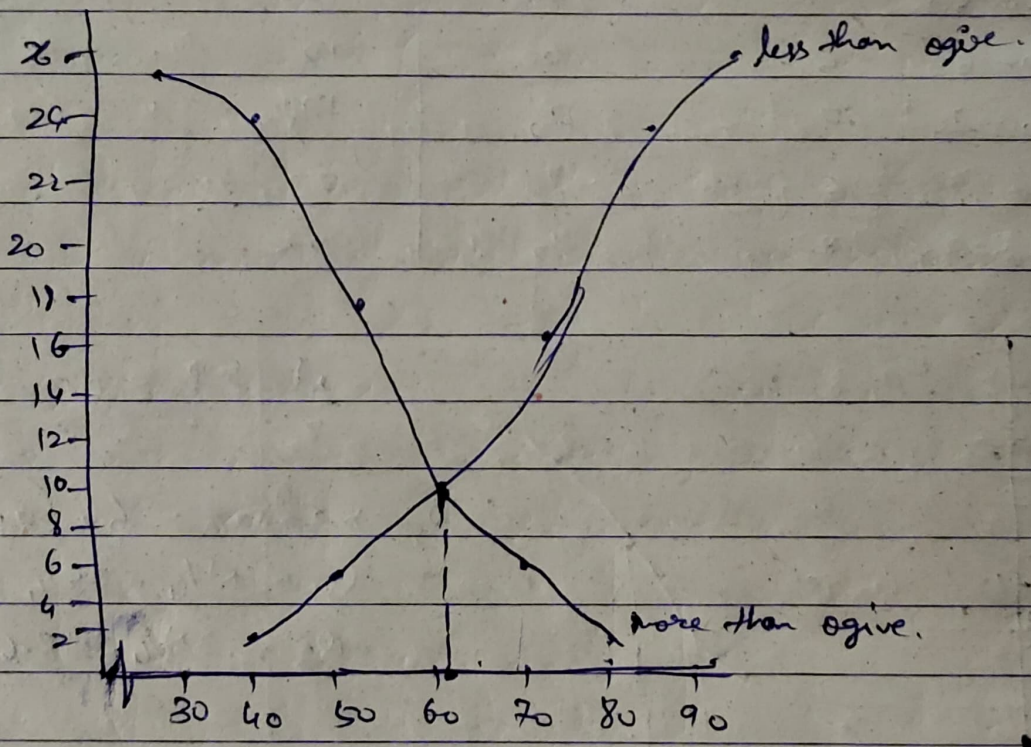
$$\frac{N}{2} = \frac{\Sigma f}{2} = \frac{25}{2} = 12.5$$

Q_2 = median = 62 ≈

$Q_1 = \frac{N}{4} = \frac{25}{4} = 6.2 = 71$ ≈

$Q_3 = 3N/4 = 18.6 = 52$ ≈ //

(iii) Both combination



median ≈ 64