- Bhumika Ma'am

FUNDAMENTALS OF STATISTICAL ANALYSIS

19/6/23

* Tally (frequency):-

0-9 ← Inclusive
10-19
20-29

0-10 ← Exclusive
10-20
20-30 (more preferable)
30-40
Qualitative →
Quantitative →

* Frequency Distribution :-

→ A survey was conducted for 20 person. Manager what's to report. The frequency difference distribution of the respondent. references were the following age years:
52, 34, 32, 29, 63, 40, 46, 54, 36, 36, 24, 19, 45, 20, 28, 29, 38, 33, 49, 37.

A: Frequency Distribution – min 19, max 63

nos.	Tally	Total
10-20	I	1
20-30	IIII	5
30-40	IIII II	7
40-50	IIII	4
50-60	II	2
60-70	I	1
		20

26/6

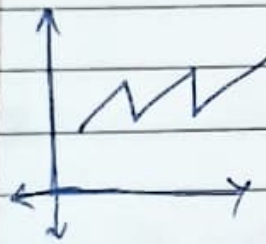
* Overview of Statistics :-

→ It is Mathematical Science pertaining to data collection, analysis, interpretation & presentation.

28/6

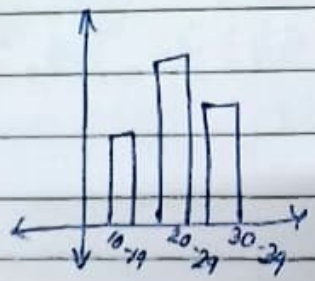
* no. of Chart / Graph :-

1) Line Graph :- when simple / discrete data is given



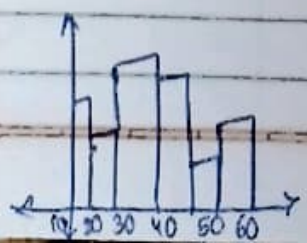
x
10
15
20
25

2) Bar Graph :- It is used when inclusive class series is given. (0-9, 10-19)



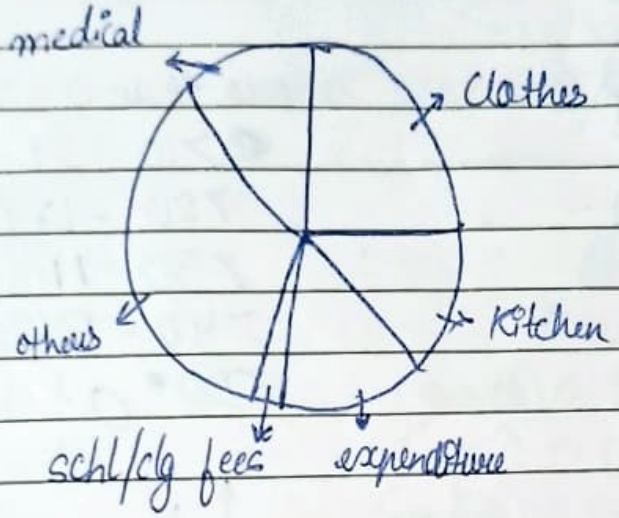
Class
10-19
20-29
30-39

3) Histogram Graph :- It is used when exclusive series is given (10-20, 20-30).



Class
10-20
20-30
30-40
40-50

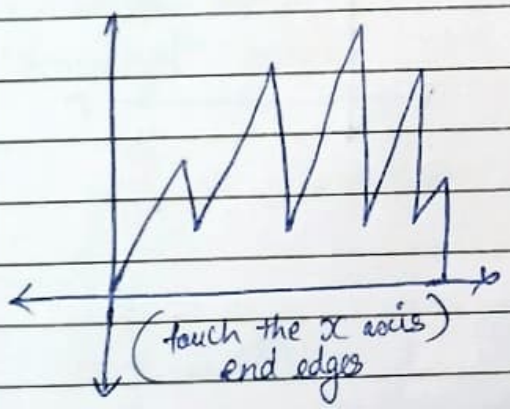
4) Pie - Chart :- (Household)



5) Frequency Polygon :- (Sheeno market)

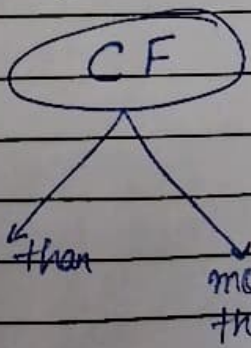
Class	
10-20	(15)
20-30	(25)
30-40	(35)

Lower limit + Upper limit / 2



6) Ogive Graph or (Cumulative frequency) Chart :-

Class	fi	Cf
0-10	2	2 (=)
10-20	4	6 (2+4)
20-30	6	12 (6+6)
30-40	8	20 (12+8)
40-50	3	23 (20+3)
	23	



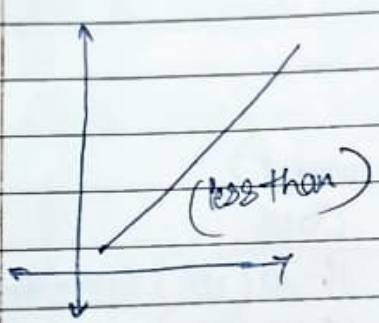
5/7

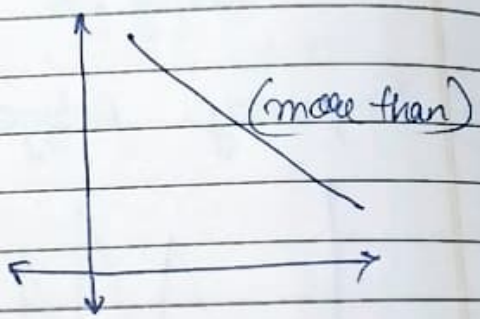
less than 4

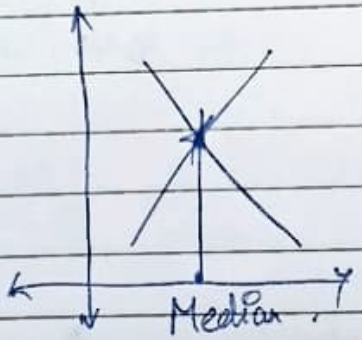
less than 10 = cf 2
less than 20 = 6
less than 30 = 12
< 40 = 20
< 50 = 23

more than 10 -

more than 0 = 23 (23-2)
> 10 = 21 (21-4)
> 20 = 17 (17-6)
> 30 = 11 (11-8)
> 40 = 3 (3)
> 50 = 0







BLOCK
* CENTRA

BLOCK 2 & 3

CENTRAL TENDENCY :- (DISPERSION)

CENTRAL TENDENCY / Dispersion

- Mean
- Median
- Mode
- Quartiles
- Deciles
- Percentiles
- Geometric Mean
- Harmonic Mean

- Range
- Mean Deviation
- (quartile) Q.D
- (Decile) D.D
- (percentile) P.D

All co-efficient

- Standard Deviation

CENTRAL TENDENCY

- Mathematical
  - Mean (A.M)
  - H.M
  - G.M
- Positional
  - Median
  - Mode
  - Quartiles
  - Percentiles
  - Deciles

(1) Mean :-

When simple observation is given (Raw Data).

x
1
2
3
4
5
15

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

$= \frac{\text{Sum of ob}^{n}}{\text{no. of ob}^{n}}$

$= \frac{15}{5}$

$\therefore \bar{x} = 3$ Ans.

x
50
75
A = 85
87
92
389

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

$= \frac{389}{5}$

$\therefore \bar{x} = 77.8$ Ans

Assumed mean $= \bar{x} = A + \frac{\Sigma d}{n}$

$d = (x - A)$

x	d
50	-35
75	-10
A = 85	0
87	2
92	7
389	-36

$\bar{x} = 85 + \frac{(-36)}{5}$

$= 85 - \frac{36}{5}$

$= 85 - 7.2$

$\therefore \bar{x} = 77.8$

x	$d=(x-A)$
1	-1
A: 2	0
3	1
4	2
5	3
15	5

$$\bar{x} = A + \frac{\Sigma d}{n}$$

$$= 2 + \frac{5}{5}$$

$$= 2 + 1$$

$$\therefore \boxed{\bar{x} = 3}$$

* Properties of mean :-

1) $\Sigma(x - \bar{x}) = 0$ → always zero

x	$x-\bar{x}$	$(x-\bar{x})^2$	$d=(x-A)$	$(x-A)^2$ (3rd property)
1	-2	4	-1	1
2	-1	1	0	0
3	0	0	1	1
4	1	1	2	4
5	2	4	3	9
	0	10	5	15

2) $\Sigma(x - A)$:- Sum of diviations from any other value are either positive, negative or zero.

3) $\Sigma(x-\bar{x})^2 < \Sigma(x-A)^2$

$\boxed{10 < 15}$

4) Mean is dependent on origin $(+, -)$ and scale $(\div, \times)$.

* Merits (Advantages) of Mean :-

→ It is the best measure of central tendency.

* De-Merits of Mean :-

→

(2) When discretes observation are given (x & f)

x	f	fx
1	2	2
2	4	8
3	6	18
4	8	32
5	10	50
	30	110

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

$$= \frac{110}{30}$$

$$\therefore \boxed{\bar{x} = 3.667}$$

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

mid value $= \frac{L.L + U.L}{2}$

$= \frac{100 + 200}{2}$

$= \frac{300}{2}$

$\therefore$ **M.V = 150** Ans.

Class
100-200
200-300
300-400
400-500
500-600

odd no.

9, 7, 5, 7, 5, 4, 7

ascending order = 4, 5, 5, 7, 7, 7, 9

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

$= \left(\frac{7+1}{2}\right)^{th}$ Obs

$= \left(\frac{8}{2}\right)^{th}$ Obs.

$= 4^{th}$ Obs

$\therefore$ **M = 7** Ans.

even no.

9, 7, 5, 7, 5, 4, 7, 8

ascending order - 4, 5, 5, (7, 7), 7, 8, 9

4th obs.

5th obs.

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

$= \left(\frac{8+1}{2}\right)^{th}$ obs.

$= \left(\frac{9}{2}\right)^{th}$ obs.

$= 4.5^{th}$ obs

Median $= \frac{4^{th} \text{ obs.} + 5^{th} \text{ obs.}}{2}$

$= \frac{7+7}{2}$

$= \frac{14}{2}$

$\therefore M = 7$ Ans.

$n = \Sigma f$

Date
Page

Q: When discrete data is given

x	f	cf	
40	2	2	
45	3	5	(2+3)
M = 50	4	9 (6.5 Obs)	(5+4)
55	1	10	(9+1)
60	2	n = 12	(10+2)
	n = 12	39	

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

$= \left(\frac{12+1}{2}\right)^{th}$ Obs.

$= \left(\frac{13}{2}\right)^{th}$ Obs.

$= 6.5^{th}$ Obs

$\therefore$ M = 50 Ans.

Self :-

(no. of students)

	x	f	cf
20	30	10	10
22	25	14	24
=25	20	8 (26.5 Obs.)	32
28	22	12	44
30	28	8	52
		52	

$\frac{52+1}{2} = \frac{53}{2} = 26.5^{th}$ Obs

$\therefore$ M = 25 Ans

formula → 2 Marks

* When continous series is given :-

Class	f	cf
0-10	2	2
10-20	4	6 = cf
MClass = 20-30	6 = f	12
30-40	8	20

n = 20

$M = \left(\frac{n}{2}\right)^{th}$ Class

$= \left(\frac{20}{2}\right)^{th}$ Class

$= 10^{th}$ Class.

M Class = 20-30

$M = \frac{L + \left(\frac{n}{2}\right) - cf}{f} \times C$

$= \frac{20 + \left(\frac{20}{2}\right) - 6}{6} \times 10$

$= \frac{20 + 10 - 6}{6} \times 10$

$= 20 + \left(\frac{4 \times 10}{6}\right)$

DMAS

$= 20 + \left(\frac{40}{6}\right)$

$= 20 + 6.66$

M = 26.667 Ans.

c = class length

Mode

* When continous series is given.

Class	f	
0-10	2	f = 8 highest frequency
10-20	4 f_0	z Class = 20-30.
20-30	8 f_1	
30-40	6 f_2	

$$z = L + \frac{f_1 - f_0}{2(f_1) - f_0 - f_2} \times C$$

$$= 20 + \frac{8 - 4}{2(8) - 4 - 6} \times 10$$

$$= 20 + \frac{4}{16 - 4 - 6} \times 10$$

$$= 20 + \frac{4}{6} \times 10$$

$$= 20 + 0.67 \times 10$$

$$= 20 + 6.7$$

$$= 26.7$$ Ans.

21/7

$M = \left(\frac{n+1}{2}\right)^{th}$ Obs.

* **Quantiles :-**

* When raw data is given :-

$Q_1 = \left(\frac{n+1}{4}\right)^{th}$ Obs.

$Q_2 = 2\left(\frac{n+1}{4}\right)^{th}$ Obs.

$Q_3 = 3\left(\frac{n+1}{4}\right)^{th}$ Obs.

Q_1	25%
Q_2	25%
Q_3	25%
	25%

* When continuous series is given :-

Q_1 * Class $= \left(\frac{n}{4}\right)^{th}$ class

$\rightarrow Q_1 = L + \frac{\left(\frac{n}{4} - cf\right)}{f} \times C$

Q_3 * Class $= 3\left(\frac{n}{4}\right)^{th}$ Obs.

$\rightarrow Q_3 = L + \frac{3\left(\frac{n}{4}\right) - cf}{f} \times C$

Q_2 Class $= 2\left(\frac{n}{4}\right)^{th}$ Class

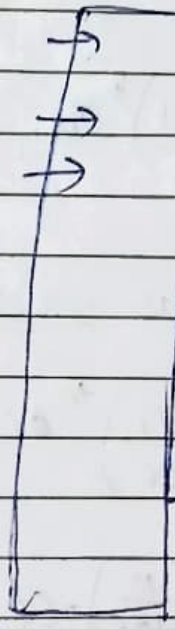
$\rightarrow Q_3 = L + \frac{2\left(\frac{n}{4}\right) - cf}{f} \times C$

* Deciles :-

* When raw data is given :-

$D_1 = \left(\frac{n+1}{10}\right)^{th}$ Obs.

$D_7 = 7\left(\frac{n+1}{10}\right)^{th}$ Obs.



* When class - series continous data is given :-

D_1 class $= \left(\frac{n}{10}\right)^{th}$ class.

$D_1 = L + \frac{\left(\frac{n}{10}\right) - cf}{f} \times C$

D_7 class $= 7\left(\frac{n}{10}\right)^{th}$ class

$D_7 = L + \frac{\left(7\left(\frac{n}{10}\right) - cf\right)}{f} \times C$

24/7

$Z = 3M - 2\bar{x}$

* **Percentile :- 99**

* When raw data is given :-

$$P_1 = \left(\frac{n+1}{100}\right)^{th} \text{Obs.}$$

$$P_{69} = 69\left(\frac{n+1}{100}\right)^{th} \text{Obs.}$$

* When continous series is given :-

$$P_1 \text{ class} = \left(\frac{n}{100}\right)^{th} \text{ Class}$$

$$P_1 = L + \frac{\left(\frac{n}{100}\right) - cf}{f} \times C$$

$$P_{69} = 69\left(\frac{n}{100}\right)^{th} \text{Class}$$

$$P_{69} = L + \left(\frac{69\left(\frac{n}{100}\right) - cf}{f}\right) \times C$$

* Harmonic mean :-

D_5 class $= 5 \left(\frac{n}{10}\right)^{th}$ class

$D_5 = L + \frac{5\left(\frac{n}{10}\right) - cf}{f} \times C$

* Relationship :-

$M = Q_2 = D_5 = P_{50}$

(25) (25) (25) (25)

$M = \left(\frac{n+1}{2}\right)^{th}$ Obs.

$Q_2 = \frac{2}{4}\left(\frac{n+1}{2}\right)^{th}$ Obs	$Q_2 = \left(\frac{n+1}{2}\right)^{th}$ Obs
$D_5 = \frac{5}{10}\left(\frac{n+1}{2}\right)^{th}$ Obs.	$D_5 = \left(\frac{n+1}{2}\right)^{th}$ Obs
$P_{50} = \frac{50}{100}\left(\frac{n+1}{2}\right)^{th}$ Obs	$P_{50} = \left(\frac{n+1}{2}\right)^{th}$ Obs

* ascending order :-

$x = 26, 34, 36, 37, 49, 58, 59, 75, 84, 96$

$n = 10$

D_5 class $= 5\left(\frac{n+1}{10}\right)^{th}$ obs.

$= 5\left(\frac{10+1}{10}\right)^{th}$ obs.

$= 5\left(\frac{11}{10}\right)^{th}$ obs.

$= 5.5^{th}$ obs.

$D_5 = \frac{5^{th} \text{ class} + 6^{th} \text{ class}}{2}$

$= \frac{49+58}{2}$

$\therefore D_5 = 53.5$ Ans.

* Combined mean :- (2M)

$\overline{x}_{12}$: $n_1, n_2, \overline{x}_1, \overline{x}_2$

$\overline{x}_1$ = avg. age of boys = 20

$\overline{x}_2$ = avg. age of girls = 10

n_1 = no. of boys = 30

n_2 = no. of girls = 20

$$\bar{x}_{12} = \frac{(n_1 \bar{x}_1) + (n_2 \bar{x}_2)}{n_1 + n_2}$$

$$= \frac{(30 * 20) + (20 * 10)}{30 + 20}$$

* Standard deviation :-

population	sample
$S.D. = \sqrt{\frac{\Sigma (x - \bar{x})^2}{N}}$	$S.D = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n-1}}$
when raw data is given.	

Question :-

Last 10 days price of vipro Ltd. are as follows :-

	closing days price
1	150
2	100
3	180
4	200
5	250
6	285
7	180
8	175
9	280
10	210

RETURN.

$$\frac{P_1 - P_0}{P_0} \times 100$$

P_1 = current price

P_0 = previous price

1	150	—
2	100	$\frac{100-150}{150} \times 100$
3	180	$\frac{180-100}{100} \times 100$
4	200	$\frac{200-180}{180} \times 100$
5	250	$\frac{250-200}{200} \times 100$

and so on

* Standard deviation :-

* change in data.
* when raw data is given :-

population sample

N n

Q. $SD = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}} = \sqrt{\frac{10}{5}}$ $= \sqrt{2}$ Ans

$= 1.41$ Ans

x	$(x-\bar{x})$	$(x-\bar{x})^2$
1	-2	4
2	-1	1
3	0	0
4	1	1
5	2	4
15		10

$\bar{x} = \frac{\Sigma x}{n} = \frac{15}{5} = 3$ Ans

Sum. $SD = \sqrt{\frac{10}{5-1}}$

$= \sqrt{\frac{10}{4}}$

$= \sqrt{2.5}$

$= 1.58$ Ans

Date
Page

* Deciles :-

D_1 D_2 D_3 D_4 D_5 D_6 D_7 D_8 D_9

$\frac{1}{10}$ $\frac{2}{10}$

$= 0.10$ $= 0.20$

Percentile (array, 0.10)

* Percentile :-

$\frac{1}{100}$ $\frac{2}{100}$ . . . $\frac{10}{100}$ $P_{10} = D_1$

$= 0.01$ $= 0.02$ $= 0.1$ $D_2 = P_{20}$

STD.P( )^2

. STDS

variance P =

variance S

4/8

* Mean deviation :-

$$M.D = \frac{\Sigma(x-\mu)}{N}$$ for ungrouped data

$$MD(mean) = \frac{\Sigma f|x-\mu|}{N}$$ grouped data

x	$\lvert x-\bar{x}\rvert$
1	$1-5.5=-4.5$
2	$2-5.5=-3.5$
3	$3-5.5=-2.5$
4	$4-5.5=-1.5$
5	$5-5.5=-0.5$
6	$6-5.5=+0.5$
7	$7-5.5=+1.5$
8	$8-5.5=+2.5$
9	$9-5.5=+3.5$
10	$10-5.5=+4.5$

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

mean $(\bar{x}) = 5.5$

= ABS C

160
60
220

Date
Page

* grouped data :-

Class	f	x	fx	$\lvert x-\bar{x} \rvert$
0-10	2	5	2×5 = 10	
10-20	4	15	4×15 = 60	
20-30	6	25	6×25 = 150	
			220	

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

$$MD = \frac{\Sigma f \lvert x-\bar{x} \rvert}{N}$$

* Mean deviation (Median) :-

$$\text{M.D about Median} = \frac{\Sigma \lvert x - \text{median} \rvert}{n}$$

$$= \frac{\Sigma |D|}{n}$$

where $D = x - \text{median}$

n = no. of obs.

Quartile deviation :-

$$Q.D = \frac{(Q_3 - Q_1)}{2}$$

$Q_3 - Q_1$ = interrange quartile

$$\text{Co-efficient of } Q.D = \frac{(Q_3 - Q_1)}{(Q_3 + Q_1)}$$