Statistics

Measures of Central Tendency (Arithmetic mean

Most of the data is clustered around a single central value.

→ Measures of Central Tendency

44 kg ⟵ ⋯ ⟶ 73 kg

Ex :→ Average.

Measures of Central Tendency

1. Arithmetic Mean   A·M
2. Median
3. Mode
4. Geometric mean   G·M
5. Harmonic mean

Arithmetic Mean   (Arithmetic = + −)

$$A \cdot M = \frac{\text{Sum of all Obs.}}{\text{no. of observation}}$$

Set of obs. x

Σ Sigma = Summation

$$x_1, x_2, x_3, \dots, x_n$$

$$A \cdot M = \frac{x_1 + x_2 + x_3 \cdots x_n}{n}$$

$$\frac{\Sigma x_i}{n}$$

formula No
Discrete obs.

frequency data

Grouped frequency distribution

4, 4, 7
4 + 4 + 4 = 12
4 × 3 = 12

x	f	X_m (mid value)	fx_m
10-20	6	15	15×6
20-30	4	25	25×4
30-40	3	35	35×3
40-50	7	45	45×7
			= 610

$N = \Sigma f$ $\Rightarrow \frac{\text{Sum of obs.}}{\text{no. of obs.}} = \frac{\Sigma f x}{N}$

$\Rightarrow \frac{610}{20} = 30.5$

Ungrouped frequency distribution

x	f	fx
2	2	4
3	1	3
6	8	48
7	3	21
9	2	18
	16	= 94

$\bar{x} = \frac{94}{16} = 5.875$

discrete obs $\rightarrow \bar{x} = \frac{\Sigma x}{n}$

ungrouped F.D $\rightarrow \bar{x} = \frac{\Sigma fx}{N}$

grouped F.D $\rightarrow \bar{x} = \frac{\Sigma fx}{N} \Rightarrow x =$ Mid value of class

Properties :-

① If all the observation assumed by a variable are constant say K, then the A.M is also K.

6, 6, 6, 6 A.M = 6

Class	f	x_m	fx_m
0 - 19	6	9.5	57
20 - 39	1	29.5	29.5
40 - 59	8	49.5	396
60 - 79	9	69.5	625.5
	24		1108

find A·M = ?

$$\bar{x} = \frac{\Sigma f x}{N} = \frac{1108}{24} = 46.16$$

(i) if all the observations assumed by a variable are constant, say K, then the A·M is also k.

$$6,\ 6,\ 6,\ 6 = 6$$

ii The alzebraic sum ~~observation~~ of deviations of a set of observation from their A·M is zero

i.e for unclassified data $\sum (x_i - \bar{x}) = 0$
for grouped F·D $\sum f_i (x_i - \bar{x}) = 0$

Ex

x	$x - \bar{x}$
4	-2
6	0
3	-3
5	-1
12	6
	⇒ 0

$$\bar{x} = \frac{\Sigma x}{n} = \frac{30}{5} = 6$$

Ex → obs 3, 6, 4, 9, 12
Sum of deviation from A·M = 0

Change of origin $(-, +)$ & change of scale $(\times, \div)$

x	$Y = x+3$	$Z = 2x$
3	6	6
5	8	10
7	10	14
15	18	30
10	13	20
40	55	80

$\bar{y} = \bar{x} + 3$
$\bar{y} = 8 + 3 = 11$
Chang of origin

$\frac{80}{5} = 16$

$\bar{x} = \frac{40}{5} = \bar{x} = 8$ $\quad = \frac{55}{5} = 11$

iii A·M is affected due to a change of origin / scale which implies that if the original variable x is changed to another variable y by effecting a change of origin, say a, & scale by b, of x i.e. $y = a + bx$, then the A·M of y is given by $= \bar{y} = a + b\bar{x}$.

$y = 2x + 3$
$\bar{x} = 9$, $y = 2 \times 9 + 3 = 21$

Combined A·M

Ques mean Salary of group of 40 female workers is 5,200 per month & group of 60 male worker is 6800 / month. Find Combined A·M.

Page No.
Date

$\bar{x} = \frac{\Sigma x}{n}$ $n = 40$, $\bar{x} = 5200$
$\Sigma x = 40 \times 5200 = 208$

	Male	female	Total
n	60	40	100
$\bar{x}$	6800	5200	=
Σx	408000	208000	616000

Combined mean Salary $= \frac{616000}{100} = 6160$

Combined A.M $= \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$

① Male ② felmale

$n_1 = 60$ $n_2 = 40$
$\bar{x}_1 = 6800$ $\bar{x}_2 = 5200$

$= \frac{(60 \times 6800) + (40 \times 5200)}{60 + 40}$

$= 6160$

Median $\rightarrow$ Location Average

5, 6, 7, 8, 12
↓ Median
2, 6, 3, 8, 9

2, 3, 6, 8, 9 → 6
↓ Median

before Calculating median, arrange observation in ascending order.

$n =$ odd

$n =$ even

1, 2, 3, 4, 7, 8

$\frac{3+4}{2} = 3.5$

2, 8, 3, 5, 4, 9, 12, 14

2, 3, 4, 5, 8, 9, 12, 14 $\rightarrow n =$ even

$\frac{13}{2} = 6.5$

Formula of median (discrete obs.

In Case of odd

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

In Case of even $= \left(\frac{n+1}{2}\right)^{th} +$ Preceding term $\Big/ 2$

$$\left(\frac{n+1}{2}\right)^{th} \text{term} = \frac{6+1}{2} = 3.5$$

3.5

3^{rd} term $\quad (4^{th} - 3^{rd}) \times 0.5$

8 $\quad (10 - 8) \times 0.5$

$8 + 1 = \underline{9}$

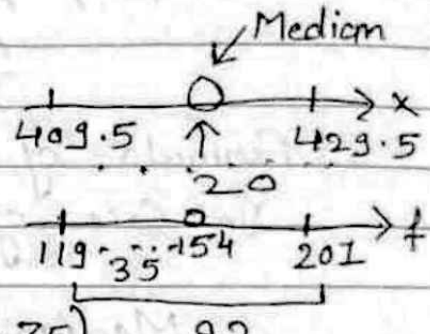
Median in case of grouped frequency distribution

Class	fre	Cum freq.
350 - 369	23	23
370 - 389	38	61
390 - 409	58	119 → N_l
410 - 429	82	201 — Median c
430 - 449	65	266 → N_u
450 - 469	31	297
470 - 489	11	308
	308	

$\frac{20}{82} \times 3$

$N = 308$

$N/2 = 154$



$$\text{Median} = 409.5 + \left(\frac{20}{82} \times 35\right)$$

$$\Rightarrow 418.04$$

$$\text{Median} = l_1 + \left[\frac{N/2 - N_l}{N_u - N_l}\right] \times C$$

l_1 = L.C.B of median class
N_l = cum freq of pre-median class
N_u = Cum freq of med Class
C = UCB − L.C.B

$$M = 409.5 + \left[\frac{\frac{308}{2} - 119}{201 - 119}\right] \times 409.5 - 429$$

$$= 418.04$$

Page No.
Date / /

parition value

Divide a paper in two parts = Me → 2
" " " 10 parts = Decile → 10
" " " 100 parts = percentile → 100
" " " 4 parts = Quartile → 4

Ques wages of labourers :- ₹ 82, ₹ 56, ₹ 90, ₹ 50, ₹ 120, ₹ 75, ₹ 75, ₹ 80, ₹ 130, ₹ 65. Find Q_1, D_6, P_{82}.

50, 56, 65, 75, 75, 80, 82, 90, 120, 130

$n = 10$

$$Q_1 = \left(\frac{n+1}{4}\right) \times 1 = \left(\frac{10+1}{4}\right) \times 1 = 2.75 \text{ Term}$$

$$2nd \text{ term} + (0.75) \times [3^{rd} - 2^{nd}]$$

$$= 56 + (65 - 56) \times 0.75$$

$$= 62.75$$

$$D_6 = \frac{(n+1)}{10} \times 6 = \frac{10+1}{10} \times 6 = 6.6 \text{ Term}$$

$$6^{th} \text{ term} + (7^{th} - 6^{th}) \times 0.6$$

$$80 + (82 - 80) \times 0.6 = 81.2$$

Page No.
Date

Mode discrete obs.

2, 6, 2, 4, 5, 6, 2, 2, 5, 1 → Unimodal distrib
Mode : obs. repeated for max no. of times.

2, 6, 2, 6, 6, 6, 2, 2, 4, 1 → Bimodal distribu
Mode = 2, 6
→ Multimodal "

x = 2, 6, 1, 9, 7 → mode not defined

Ungrouped F · D

	x	F
Mode	2	1
	6	7
	4	5
	9	2

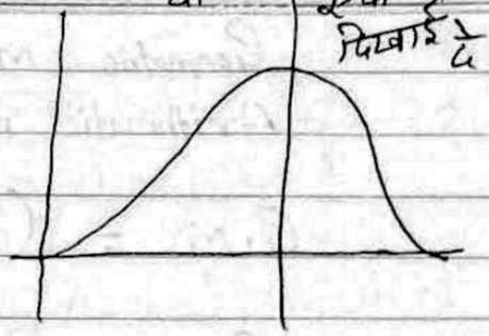
Ex. Grouped Frequency distribution

$$\text{Mode} = l_1 + \left[\frac{f_0 - f_1}{2f_0 - f_{-1} - f_1}\right] \times C$$

Modal class → Class with higher frequency
l_1 = L·C·B of Modal class
f_0 = freq of modal class
f_{-1} = freq of premodal class
f_1 = " " post modal Class

Relation b/w mean, median, mode

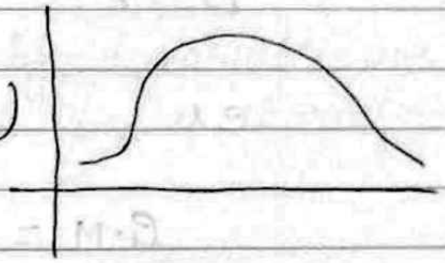
दोनों ही तरफ से एक जैसा दिखाई दे।



Symmetric distribution

Mean = Median = mode

SKew symmetric distribution



Mean − Mode = 3 (Mean − Median)

Ques moderately skewed distribution of marks in statistics for a ~~student~~ group of 200 students, the mean mark & median mark were found to be 55.60 & 52.40 what is modal mark.

Ans Mean = 55.60 Median = 52.40, mode = ?

Mean − mode = 3 (Mean − Median)

55.60 − mode = 3 (55.60 − 52.40)

55.60 − mode = 3 × 3.2

55.60 − 9.6 = mode

m = 46.

Property of mode :- Mode is also affected by both change of origin & change of scale.

$x_{mode} = 5$ $y = 2x + 6$

$y_{mode} = ?$ $y = 2 \times 5 + 6 = 16 = y_{mode}$

Geometric Mean

Arithmetic mean = + − Geometric = × ÷

$$G.M = (x_1 \times x_2 \times x_3 \times x_4)^{1/4}$$

→ discrete observ

$$G.M = \sqrt[4]{x_1 \times x_2 \times x_3 \times x_4}$$

ex. 2, 6, 5, 10 ⟹ A.M = 5.75

$$G.M = \sqrt[4]{2 \times 6 \times 5 \times 10}$$

$$\Rightarrow \sqrt[4]{600} = 4.949$$

Frequency distribution

x	f
2	2
4	1
7	3
9	4

2, 2, 4, 7, 7, 7, 9, 9, 9

$(2 \times 2) \times 4 \times (7 \times 7 \times 7) \times (9 \times 9 \ldots$

$$[2^2 \times 4^1 \times 7^3 \times 9^4]^{1/10}$$

$$G.M = [x_1^{f_1} \times x_2^{f_2} \times \ldots \times x_n^{f_n}]^{1/N}$$

Properties of G.M

① Logarithm of G for a set of observations is the A.M of the logarithm of the observation; i.e.

$$\log G = \frac{1}{n} \sum \log x$$

G.M of $x = (2\times4\times8)^{1/3} = (64)^{1/3} = 4$

x	$\log_2 x$
2	$\log_2 2 = 1$
4	$\log_2 4 = 2$
8	$\log_2 8 = 3$
	$\Rightarrow 6$

$$\frac{\Sigma \log x}{n} = \frac{6}{3} = 2$$

$$\log GM = \log_2 4 = 2$$

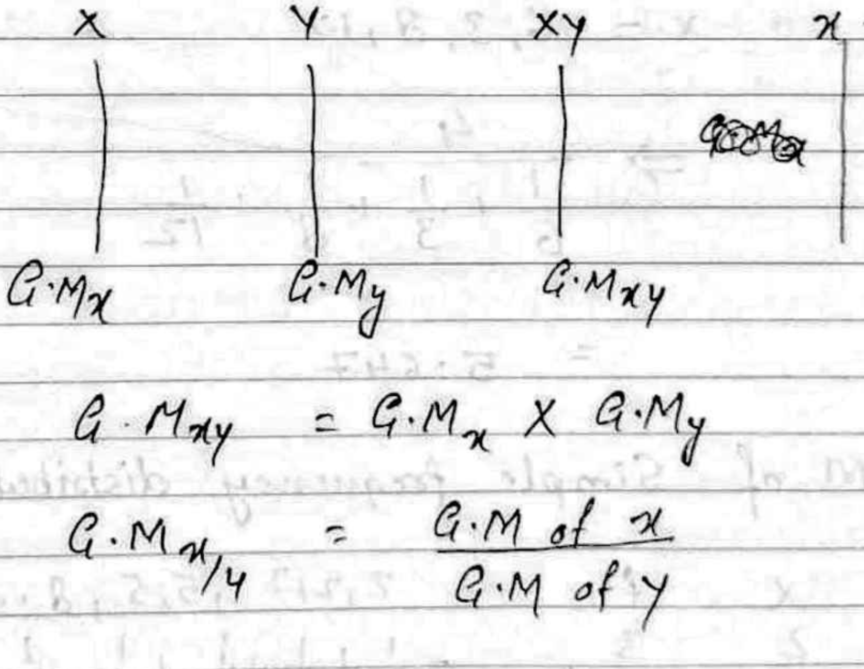
$$\log G.M = \frac{1}{n}\Sigma \log x \rightarrow \text{Arithmetic mean of } \log x$$

ii) if all the observations assumed by a variable are constants, say $k > 0$, then the G.M of the observation is also k.

$$6, 6, 6, 6 \Rightarrow G.M = \sqrt[4]{6\times6\times6\times6} = 6$$

iii G.M of the product of two variables is the product of their G.M's i.e. if $z = xy$, then

$$G.M \text{ of } z = (G.M \text{ of } x) \times (G.M \text{ of } y)$$



$$G.M_{xy} = G.M_x \times G.M_y$$

$$G.M_{x/y} = \frac{G.M \text{ of } x}{G.M \text{ of } y}$$

iv) G.M of the ratio of two variables is the ratio of the G.M's of the two variables i.e. if $z = x/y$ then

$$G.M \text{ of } z = \frac{G.M \text{ of } x}{G.M \text{ of } y}$$

Harmonic Mean

$$\text{obs. } x \Rightarrow \quad \overset{x_1}{2}, \overset{x_2}{3}, \overset{x_3}{9} \qquad n = 3$$

reciprocal of $x \Rightarrow \frac{1}{2}, \frac{1}{3}, \frac{1}{9}$

A.M of reciprocal of $x \Rightarrow \dfrac{\frac{1}{2} + \frac{1}{3} + \frac{1}{9}}{3}$

Reciprocal of A.M of reciprocal of $x = \dfrac{3}{\frac{1}{2} + \frac{1}{3} + \cdot}$

$$H.M = \frac{1}{\frac{1}{x_1} + \frac{1}{x_3} + \frac{1}{x_3} + \cdots + \frac{1}{x_n}}$$

$$= \frac{1}{\Sigma\left(\frac{1}{x}\right)}$$

$\rightarrow$ discrete observ

Ex. $x = 6, 3, 8, 12$

$$\Rightarrow \frac{4}{\frac{1}{6} + \frac{1}{3} + \frac{1}{8} + \frac{1}{12}}$$

$$= 5.647$$

H.M of Simple frequency distribution

x	f
2	3
5	2
8	4
11	6
	15

$2, 2, 2, 5, 5, 8 \ldots, 11 \ldots$

$= \frac{1}{2} + \frac{1}{2} + \frac{1}{2}, \frac{1}{5} + \frac{1}{5}, \frac{1}{8} + \frac{1}{8} + \frac{1}{8}$

$\frac{1}{11} + \frac{1}{11} \cdots + \frac{1}{11}$

$3/2$

Date

Formula $H.M = \frac{N}{\Sigma\left(\frac{f_i}{x_i}\right)}$

$\Rightarrow \frac{15}{3/2 + 2/5 + \frac{4}{8} + \frac{6}{11}}$

$\Rightarrow 5.0926$

Properties of H.M

i) if all the observations taken by a variable are constants, say k, then the H.M of the observations is also k.
Ex :- 4, 4, 4, 4 => H.M = 4

ii) if there are two groups with n_1 & n_2 observations & H_1 & H_2 as respective HM's than the combined HM is given by $\rightarrow \frac{n_1 + n_2}{\frac{n_1}{H_1} + \frac{n_2}{H_2}}$

Relationship b/w AM GM HM

In General

AM ≥ G.M ≥ HM

if observations are Same
A.M = GM = HM

if observations are distinct
AM > GM > HM

Ex → 2, 4, 8

$AM = \frac{2+4+8}{3}$
$= 4.66$

$GM = (2 \times 4 \times 8)^{1/3}$
$= (64)^{1/3}$
$= 4$

$HM = \frac{3}{\frac{1}{2} + \frac{1}{4} + \frac{1}{8}}$
3.428

A.M > GM > HM

Date / /

Uses of GM & HM

When observations are in the form of rates.
Ex. speed, wages rate etc
we use H·M

When observation are in the form of % or ratios. Ex marks %, Int. rates etc
we use G·M

In all other cases A·M used.

drawback of AM

open end classification

	f	x_m
Below 100		not av.
100 - 200		
200 - 300		
Above - 300		not av.

AM Can Calculate

Page No.
Date / /

Measures of Dispersion | Basics & Range

why dispersion?

Two set of data can have same measure of central tendency but actually the level of Centralisation is different.

Range = Tell us about the difference b/w Largest value & smallest value.

P	Q
2	4
4	6
6	6
8	6
10	8

A.M of $P = 30/5 = 6$ | A.M of $Q = 30/5 = 6$

Range of $P = 10-2 = 8$ | Range of $Q = 8-4 = 4$

$R \to P > R \to Q$

higher the dispersion weak is Central Tendency.
lower " " Strong " "

Defination of Dispersion

Two distributions may be identical in respect of its first important characterstic i.e. central tendency & yet they may differ on account of scatteriness.

Dispersion for a given set of observations may be defined as the amount of deviation of the observations usually, from an appropiate measure of Central tendency.

Types of Dispersion

① Absolute measure / Dispersion
② Relative " "

Range

In a class room

x (weight)	y (Height)
60	150
58	162
48	155
52	143
70	164
43	172
50	168
62	140
68	150
50	158

Rang = 70 - 43 kg $\Rightarrow$ 172 - 140 cm

= 27 kg $\Rightarrow$ 32 cm

we ~~cannot~~ Can'not compare two different units so for solving this problem we use 'Relative measure of dispersion' which is not depend on units.

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

$C.O.R = \frac{kg\ 70-43}{kg\ 70+43} \times 100$ $\qquad \frac{cm\ 172-140}{cm\ 172+140} \times$

$= 23.89$ $\qquad = 10.25$

data of x is scattered as compare to y.

Conclusion :- if we have to compare two sets of observations, we need relative measure of dispersion

Type of Dispersion

Absolute	Relative
Range	Coefficient of Range
Mean deviation	Coefficient of mean deviation
Standard deviation	Coefficient of variation
Quartile deviation	Coefficient of Quartile "

① Absolute measures are dependent on the unit of the variable under consideration whereas the relative measures of dispersion are unit free.

② for Comparing two or more distributions, relative measures & not ~~absow~~ absolute measures of dispersion are Considered.

Range

→ discrete data = $L - S$

grouped freq = $L - S$

where L = UCB of last Class

S = LCB of first class

Grouped

Quest

weight in kg	50-54	55-59	60-64	65-69	70-74
No. of Students	12	18	23	10	3

→ L.C.B = 49.5 UCB = 74.5

Range = $74.5 - 49.5 = 25$ Kgs

Coefficient of Range = $\frac{74.5 - 49.5}{74.5 + 49.5} \times 100 = 20.16$

discrete obs.

12, 18, 23, 10, 3

Range = L − S = 23 − 3 = 20

Coefficient / Range = $\frac{23-3}{23+3} \times 100 = \frac{20}{26} \times 100$

$= 76.92$

Range = −6 − (−2

= 14

chang of origin change of scale

x	y = x+3	= Z = 2x	b = −2
3	6	6	−7
6	9	12	−11
8	11	16	−16
9	12	18	−18
10	13	20	−20

→ Range = 13 − 6 = 7

20 → R = 20 − 6

Range = 10 − 3 = 7 → double of the value in change of scale

Measures of dispersion are not affected by change of origin

Measures of dispersion Affected by change of scale but no impact of sign → Range of 6 12

All MOD are always positive

Ques − x & y = 2x + 3y = 10 range of x is 15 range of y

$2x + 3y = 10$

write in standard form

$Y = \frac{10}{3} - \frac{2}{3}x$

origin $+\frac{10}{3}$ α

Scale $\left|-\frac{2}{3}\right| = \frac{2}{3}$

$R_y = \frac{2}{3}x = \frac{2}{3} \times 15 = 10$

Page No.
Date

Range which section data is Centralize

X	Y
2	2
6	8
8	8
10	8
12	8
14	14

In this data of x is disperse but the range of x & y is same So, by the range we cannot define which is scattered. for this we use mean deviation

$R \to x \Rightarrow 14-2 = 12$ | $R \to y = 14-2 = 12$

M.O.D can never be negative

distance is only distance it can never be positive or negative

Mean deviation

x	X−A	|x−A|
2	−7	7
6	−3	3
8	−1	1
10	1	1
12	3	3
14	5	5
		20

Absolute deviation

Median of x

$\frac{8+10}{2} = 9 - A$

$\frac{\Sigma |x-A|}{n} = \frac{20}{6} = 3.33$

y	|Y−A|
2	6
8	0
8	0
8	0
8	0
14	6
	12

Median of y

$\frac{8+8}{2} = 8 - A$

$\frac{\Sigma |Y-A|}{n} = \frac{12}{6} = 2$

Page No.
Date / /

Defination :- Mean deviation is defined as the A.M of absolute deviation taken from any appropriate central tend

$$MD_x = \frac{\Sigma |x-A|}{n}$$

Coefficient of mean deviation $= \frac{M.D}{A} \times 100$

Standard deviation

x	$x-\bar{x}$	$(x-\bar{x})^2$
4	-5	25
6	-3	9
8	-1	1
10	1	1
12	3	9
14	5	25
		70

$\bar{x} = \frac{54}{6} = 9$

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

$= \frac{70}{6} = 11.66$

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

$= \sqrt{11.66}$

$= 3.415$

def :- S.D is the root mean sq deviation & deviation is taken from A.M

Main formula $\sqrt{\frac{\Sigma (x-\bar{x})^2}{n}} = f.D = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{n}}$

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

Page No.
Date / /

Relative measure

Coefficient of variation = $\frac{SD}{AM} \times 100$

if you have two observation ~~data~~ only

$$S.D = \frac{a-b}{2}$$

x	x^2
8	64
10	100
18	164

$\bar{x} = 9$

$$S.D = \sqrt{\frac{164}{2} - (9)^2} = 1$$

$$\frac{10-8}{2} = 1$$

S.D of first n natural no. = $SD = \sqrt{\frac{n^2-1}{12}}$

$$\sqrt{\frac{25-1}{12}} = \sqrt{2}$$

x	x^2
1	1
2	4
3	9
4	16
5	25
15	55

$\bar{x} = \frac{15}{5} = 3$

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

$$= \sqrt{\frac{55}{5} - 9}$$

$$= \sqrt{2}$$

Find SD

Class	f	x_m	fx^2
50-52	17	51	
52-54	35	53	
54-56	28	55	
56-58	15	57	
58-60	5	59	
	100		293372

$$\bar{x} = \frac{\Sigma f x_m}{\Sigma f} = \frac{5412}{100}$$

$$M = 54.12$$

$$S.D = \sqrt{\frac{\Sigma f x^2}{N} - (\bar{x})^2}$$

$$= \sqrt{\frac{293372}{100} - (54.12)^2}$$

$$= \sqrt{4.7456} = 2.1784 = 2.18$$

Moment, Skewness & Kurtosis

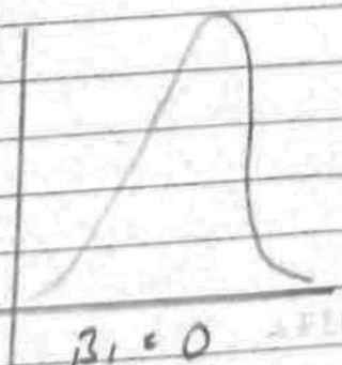
Calculation of Central moment

		Individual Series	F.D
①	μ_1	$\frac{\Sigma d}{N}$	$\frac{\Sigma fd}{N}$
②	μ_2	$\frac{\Sigma d^2}{N}$	$\frac{\Sigma fd^2}{N}$
③	μ_3	$\frac{\Sigma d^3}{N}$	$\frac{\Sigma fd^3}{N}$
④	μ_4	$\frac{\Sigma d^4}{N}$	$\frac{\Sigma fd^4}{N}$

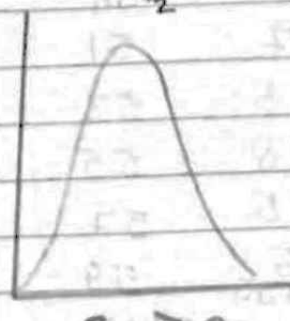
where $d = x - \bar{x}$ Raw $= x - A$

$$\text{Skewness} = \beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

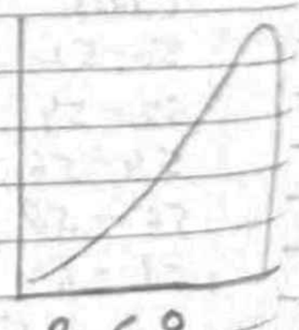
$$\text{Kurtosis} = \beta_2 = \frac{\mu_4}{\mu_2^2}$$



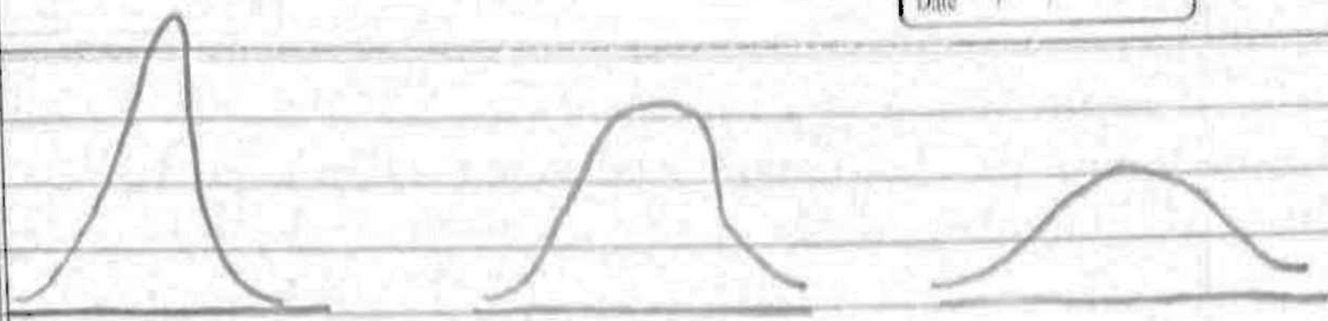
$\beta_1 = 0$
Symmetrical N.D



$\beta_1 > 0$
Positively skewed d.



$\beta_1 < 0$
Negatively Skewed dist.



$\beta_2 > 3$ Leptokurtic | $\beta_2 = 3$ Mesokurtic | $\beta_2 < 3$ Platykurtic

x	f	$d\ (x-\bar{x})$	fd	fd^2	fd^3	fd^4
0	1	-4	-4	16	-64	256
1	8	-3	-24	72	-216	648
2	28	-2	-56	112	-224	448
3	56	-1	-56	56	-56	56
4	70	0	0	0	0	0
5	56	1	56	56	56	56
6	28	2	56	112	224	448
7	8	3	24	72	216	648
8	1	4	4	16	64	256
36	256		0	512	0	2816

$$\text{Mean} = \frac{\Sigma x}{N} = \frac{36}{9} = 4$$

$$\mu_1 = \frac{\Sigma fd}{N} = 0$$

$$\mu_2 = \frac{\Sigma fd^2}{N} = \frac{512}{256} = 2$$

$$\mu_3 = \frac{\Sigma fd^3}{N} = 0$$

$$\mu_4 = \frac{\Sigma fd^4}{N} = \frac{2816}{256} = 11$$