

* Ch :- 2 A

Measures of Central Tendency

- Average Discover Uniformity in Variability

Average

Mathematical
[Based on all obs..]

Positional
[Not affected by
Extreme value]

A.M. G.M. H.M.

- best

- Avg. of

- Avg.

Median

Mode

- Commonly
used

~~to~~, %

Speed

Qualitative
data

Quickest

Growth rate,
Ratio

- ogive

- Histogram

- Area
Diagram

*

A.M.

$$\bar{x} = \frac{\sum x_i}{n} \quad \bigg| \quad \bar{x} = \frac{\sum f_i x_i}{n}$$

$$x: x_1, x_2, \dots, x_n \quad \bigg| \quad x: x_1, x_2, \dots, x_n$$

$$f: f_1, f_2, \dots, f_n$$

Class	f
$a_1 - b_1$	f_1
$b_1 - c_1$	f_2
1	1
1	1

$$\bar{x} = \frac{\sum f_i x_i}{n}$$

Class Mark
Mid..value

$$\underline{\underline{\bar{x}}}$$

$$\sum x_i + \sum f_i x_i = \bar{x}$$

$$x: 10, 13, 14, 18, 20 \quad \bigg| \quad \bar{x} = 15$$

$$x - \bar{x}: -5, -2, -1, 3, 5 \rightarrow 0$$

$$(x - \bar{x})^2: 25, 4, 1, 9, 25 \rightarrow 64$$

$$x - 14: -4, -1, 0, 4, 6$$

$$(x - 14)^2: 16, 1, 0, 16, 36 \rightarrow 69$$

$$|x - 14|: 4, 1, 0, 4, 6 \rightarrow 15$$

$$|x - \bar{x}|: 5, 2, 1, 3, 5 \rightarrow 16 \text{ minimum}$$

Note :-

- ① Sum of deviation from A.M. is zero
- ② Sum of square of deviation from A.M. is Minimum
- ③ Sum of absolute deviation from Median is Minimum.

* Combine Mean

$$\bar{x}_c = \frac{h_1 \bar{x}_1 + h_2 \bar{x}_2}{h_1 + h_2}$$

Weighted Mean

$$\bar{x} = \frac{\sum w_i x_i}{\sum w_i}$$

<u>Ex:</u>	<u>x</u>	<u>w</u>
	1	1
	2	2
	3	3
	4	4
		<u>10</u>

$$\bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{120}{10} = 12$$

Trick $\bar{x} = \frac{2 \times 4 + 1}{3} = 3$

Trick :- 1

X

w/f

1	1
2	2
3	3
4	4
1	1
1	1
h	h

$$\bar{x} = \frac{2h+1}{3}$$

Trick :- 2

If nos. are in A.P

$$\bar{x} = \frac{a+l}{2}$$

X : 1, 2, 3, ..., 10

$$\bar{x} = \frac{1+10}{2} = 5.5$$

Imp

Affected by Change of Origin and Scale.

If $ax + by = C$ then
 $a\bar{x} + b\bar{y} = C$

Ex:- $\bar{x} = 30$

$$y = 2x - 10$$

$$\bar{y} = 2\bar{x} - 10$$

$$\begin{aligned}\bar{y} &= 2 \times 30 - 10 \\ &= 50\end{aligned}$$

$$\bar{x} = 5$$

$$2x - 4y = 40$$

$$2\bar{x} - 4\bar{y} = 40$$

$$2 \times 5 - 4\bar{y} = 40$$

$$10 - 4\bar{y} = 40$$

$$10 - 40 = 4\bar{y} = 0$$

$$-30 = 4\bar{y}$$

$$\frac{-30}{4} = \bar{y} = -7.5$$

*

G.M

→ Special Average used only for % , Growth Rate or Ratio.

$$\rightarrow G.M = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n}$$

$$G.M = (\text{Product})^{1/n}$$

//_

— Note :-

$$n \times A.M = \text{Sum}$$
$$G.M^n = \text{Product}$$

— Most difficult to Calculate.

— Cannot be asure if any value is Zero or Negative

— Combine G.M

$$\begin{array}{lcl} \text{group of} & : & n_1 \quad n_2 \\ \text{G.M} & : & G.M_1 \quad G.M_2 \end{array}$$

$$G.M_c = \left(G.M_1^{n_1} \times G.M_2^{n_2} \right)^{\frac{1}{n_1 + n_2}}$$

Ex If $G.M(x) = 2$, $G.M(y) = 3$
then $G.M(x, y) = G.M(x) \cdot G.M(y)$
 $= 2 \times 3 = 6$

— If $z = x \cdot y$ then $G.M_z = G.M(x) \cdot G.M(y)$

— If $z = \frac{x}{y}$ then $G.M_z = \frac{G.M(x)}{G.M(y)}$

If nos. are in G.P

then $G.M = \sqrt[n]{a \times x \times \dots}$

//_

* H.M.

— Special type of average

— $H.M.(a, b) = \frac{2}{\frac{1}{a} + \frac{1}{b}}$

$$H.M.(a, b, c) = \frac{3}{\frac{1}{a} + \frac{1}{b} + \frac{1}{c}}$$

Ex:- $H.M.(2, 3, 5) = \frac{3}{\frac{1}{2} + \frac{1}{3} + \frac{1}{5}}$

$$\left. \begin{array}{l} 2 \div = M + \\ 3 \div = M + \\ 5 \div = M + \end{array} \right\} MRC \div = \times 3 = 2.90$$

— It is used to find

- (i) Avg. Speed when distance is Constant
- (ii) Avg. Rate when total Interest is Constant
- (iii) Avg. Price when total Expenditure is Constant

//_

— Cannot Calculate if any value is Zero.

— Combine H.M.

Group of: h_1, h_2

H.M : $H.M_1, H.M_2$

$$H.M_c = \frac{h_1 + h_2}{\frac{h_1}{H.M_1} + \frac{h_2}{H.M_2}}$$

* Note

① $AM \geq GM \geq H.M$

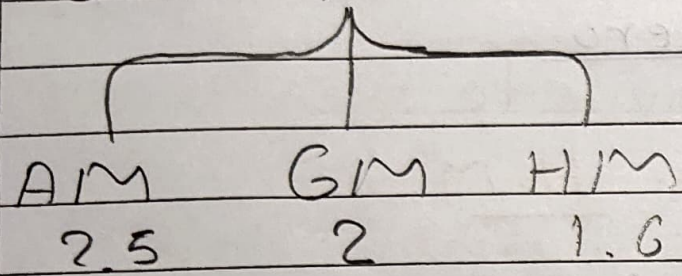
If all value are same
 $AM = GM = H.M$

If values are distinct
 $AM > GM > H.M$

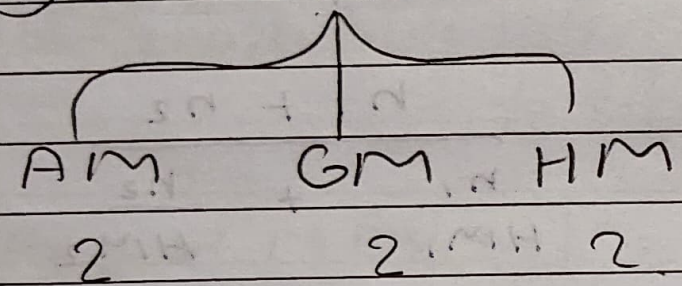
② $G.M = \sqrt{A.M \times H.M}$

$$G.M^2 = A.M \times H.M$$

Ex:- ① 1, 4



② 2, 2



* Median

- 2nd best average after A.M
- Qualitative
- not affected by extreme value
- Only MCT for open-ends
- Graphically located using ogive
- Affected by change of origin and Scale

$$\text{If } ax + by = c \text{ then} \\ aM_x + bM_y = c$$

Ex (1) 3, 5, (10), 19, 23

|
M

(2) 3, 5, 10, 19, 23, 51

$$\frac{10 + 19}{2} = 14.5$$

(3)

<u>x</u>	<u>f</u>
9	3
5	8
2	1
4	2

<u>x</u>	<u>f</u>	<u>cf</u>
2	1	1
4	2	3
5	8	11
9	3	14 = N

$$\frac{n+1}{2} = 7.5$$

(4)

<u>Class</u>	<u>f</u>	<u>cf</u>
0-5	3	3
5-10	2	5
10-15	10	15
15-20	5	20 = N

$$\frac{N}{2} = \frac{20}{2} = 10$$

$$M = l + \frac{\frac{N}{2} - cf}{f} \times c$$

$$M = 10 + \frac{10-5}{10} \times 5 = 12.5$$

l = lower class boundary
 C = class width

//_

* Quartiles :

$$Q_i = i \left(\frac{N+1}{4} \right)^{\text{th}} \text{ Rank}$$

$$Q_i = l + i \frac{\frac{N}{4} - cf}{f} \times c$$

* Deciles :

$$D_i = i \left(\frac{N+1}{10} \right)^{\text{th}} \text{ Rank}$$

$$D_i = l + i \frac{\frac{N}{10} - cf}{f} \times c$$

* Percentiles :

$$P_i = i \left(\frac{N+1}{100} \right)^{\text{th}} \text{ Rank}$$

$$P_i = l + i \frac{\frac{N}{100} - cf}{f} \times c$$

— Note : All Partition values (M, Q, D, P) can be located graphically using Ogive.

*

Mode

— Highest frequency

— Quickest

— Ill-defined

Ex (1) 1, 1, 1, 2, 2, 3

$$Z = 1$$

(2) 1, 1, 1, 2, 2, 2, 3, 3

$$Z = 1, 2$$

— Graphically located using Histogram

— If $ax + by = C$ then
 $az_x + b.z_y = C$

$$Z = l + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times c$$

$$\begin{array}{l} \boxed{\begin{array}{c} f_1 \\ f_m \end{array}} \quad \begin{array}{l} f_m - f_1 = d_1 \\ f_m - f_2 = d_2 \end{array} \end{array}$$

$$Z = \frac{l + \frac{d_1}{d_1 + d_2} \times c}{d_1 + d_2}$$

$$\begin{aligned} 2f_m - f_1 - f_2 \\ = (f_m - f_1) + (f_m - f_2) \\ = d_1 + d_2 \end{aligned}$$

— Symmetric:

$$\bar{O} = M = Z$$

— Moderately Skewed:

$$\begin{aligned} Z &= 3M - 2\bar{O} \\ \text{Mean} - \text{Mode} &= 3(\text{Mean} - \text{Median}) \end{aligned}$$

— Note:

- ① Based on all observation: AM, GM, HM
- ② Rigidly defined (Unique): Mean, Median
- ③ Affected by extreme values:
A.M, G.M, H.M
- ④ Algebraic Properties (Combine):
AM, GM, HM

* Ch :- 2 A

Measures of Dispersion

→ Dispersion discovers Variability in Uniformity

→ 2nd Order average

→ Not affected by change of Origin but affected by change of Scale

→ MCT Can take any value but MD is always- Non Negative

Measure of Dispersion (Scatterness)

Absolute (Used to <u>Measure</u> .)	Relative (Used to <u>Compare</u>)
--	---------------------------------------

(1) Range = $H - L$

(1) Coeff. of Range
 $= \frac{H - L}{H + L} \times 100$

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

(2) Coeff. of QD
 $= \frac{(Q_3 - Q_1)}{(Q_3 + Q_1)} \times 100$
 $= \frac{Q.D}{Q_2} \times 100$

(3) $MD_A = \frac{\sum |x - A|}{n}$

$A = \bar{x}, M, Z$

(3) Coeff. of MD_A
 $= \frac{MD_A}{\bar{A}} \times 100$

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

Best

(4) Coeff. of Variance
 $= \frac{S.D}{\bar{x}} \times 100$

//_

Imp Ex 15, 18, 20, 25, 30, 32, 35.

→ $\text{Range} = H - L = 35 - 15 = 20$

$$\text{Coeff of Range} = \frac{H - L}{H + L} \times 100 = \frac{35 - 15}{35 + 15} \times 100 = 40.$$

→ $Q_1 = \frac{7+1}{4} = 2^{\text{nd}} \text{ rank} = 18$

$$Q_3 = 3\left(\frac{7+1}{4}\right) = 6^{\text{th}} \text{ rank} = 32$$

$$QD = \frac{32 - 18}{2} = 7$$

$$\text{Coeff. of QD} = \frac{32 - 18}{32 + 18} \times 100 = 28$$

→ $\bar{x} = 25$

$$MD_{\bar{x}} = \frac{\sum (x - 25)}{7}$$

$$15 - 25 = -10 \quad M-$$

$$18 = -7 \quad M-$$

$$20 = -5 \quad M-$$

$$30 = 5 \quad M+$$

$$32 = 7 \quad M+$$

$$35 = 10 \quad M+$$

$$MRC \div 7$$

$$= 6.29 \text{ MD}$$

$$\text{Coeff of MD}$$

$$= \frac{6.29}{25} \times 100$$

$$= 25.14$$

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

$$15 - 25x = M +$$

$$18 - 25x = M +$$

$$20 - 25x = M +$$

$$30 - 25x = M +$$

$$32 - 25x = M +$$

$$35 - 25x = M +$$

$$1120 \div 7$$

$$\sqrt{\quad}$$

$$7.05$$

$$= SD$$

$$7.05 \div 25 \times 100 = 28.20$$

Coeff. of variance

* Range

- Quickest
- Not based on all observation
- If $ax + by = C$ then
 $|a| R_x = |b| R_y$

Ex :- $-3x + 4y + 6 = 0$

$$3R_x = 4R_y$$

* QD

- Only one which can be used in Open end
- Not based on all observation (Central 50%)
- If $ax + by = C$ then
 $|a| QD_x = |b| QD_y$

* MD

$$MD_A = \frac{\sum |x - A|}{n}$$

$$A = \bar{x}, M, Z$$

- Minimum for $A = M$
- Based on all observation
- If $ax + by = c$ then

$$|a| MD_x = |b| MD_y$$

* SD

- Best, Most Commonly Used
- Based on all observation
- If $ax + by = c$ then

$$|a| SD_x = |b| SD_y$$

- $V(x) = SD^2$

— 1 Root mean Square of Deviation from mean'

$$SD = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

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

$$\sigma^2 = \frac{\sum x^2}{n} - \bar{x}^2$$

$$\sum x^2 = (\sigma^2 + \bar{x}^2) \times n$$

$$\sum x = \bar{x} \times n$$

— for two nos. a, b

$$SD = \frac{|b-a|}{2} = \frac{\text{Range}}{2}$$

SD is half of Range

Ex' - 3, 11

$$\bar{x} = 7, \begin{array}{l} 3-7 \times = M+ \\ 11-7 \times = M+ \end{array} \quad \left. \begin{array}{l} \text{MIRC} \div 2 = \sqrt{} \\ SD = 4. \end{array} \right\}$$

Trick $SD = \frac{11-3}{2} = 4$

$$\rightarrow SD(3, 3, 11, 11)$$

$$\bar{x} = 7$$

$$3 - 7 \quad x = 14 + 14 + \sqrt{144 \div 4} = \sqrt{}$$

$$11 - 7 \quad x = 14 + 14 + \sqrt{144} = 4$$

- If all frequency are same then to find SD Consider each values only once.

$$\rightarrow x = 1, 2, 3, 4, 5$$

$$SD = \sqrt{\frac{5^2 - 1}{12}} = 1.4142$$

- If $x = 1, 2, 3, 4, \dots, n$

$$SD = \sqrt{\frac{n^2 - 1}{12}}$$

$$V(x) = \frac{n^2 - 1}{12}$$

*

Combine S.D.

Group of	n_1	n_2
Mean	$\bar{x}_1$	$\bar{x}_2$
SD	s_1	s_2

$$SD_c = \sqrt{\frac{n_1(s_1^2 + d_1^2) + n_2(s_2^2 + d_2^2)}{n_1 + n_2}}$$

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

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

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

* ch :- 3A

Correlation Analysis

→ * Correlation :

— Degree of association between two or more variable

— r = Coefficient of Correlation
 $-1 \leq r \leq 1$

— $r = 0$ No relation

$r = 1$ Perfect Positive Relation

$r = -1$ Perfect Negative Relation

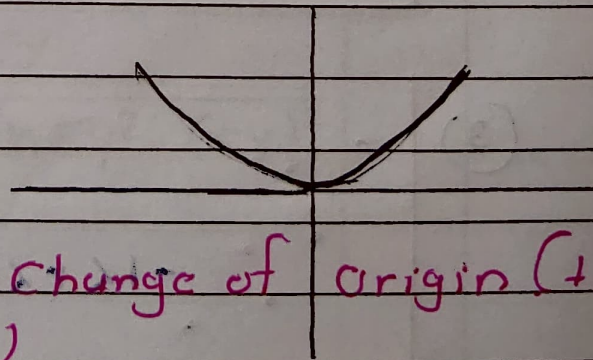
— In Correlation we have only Linear relation.

Ex :-

x	0	-2	-1	1	2
y	0	4	1	1	4

$$y = x^2$$

No relation

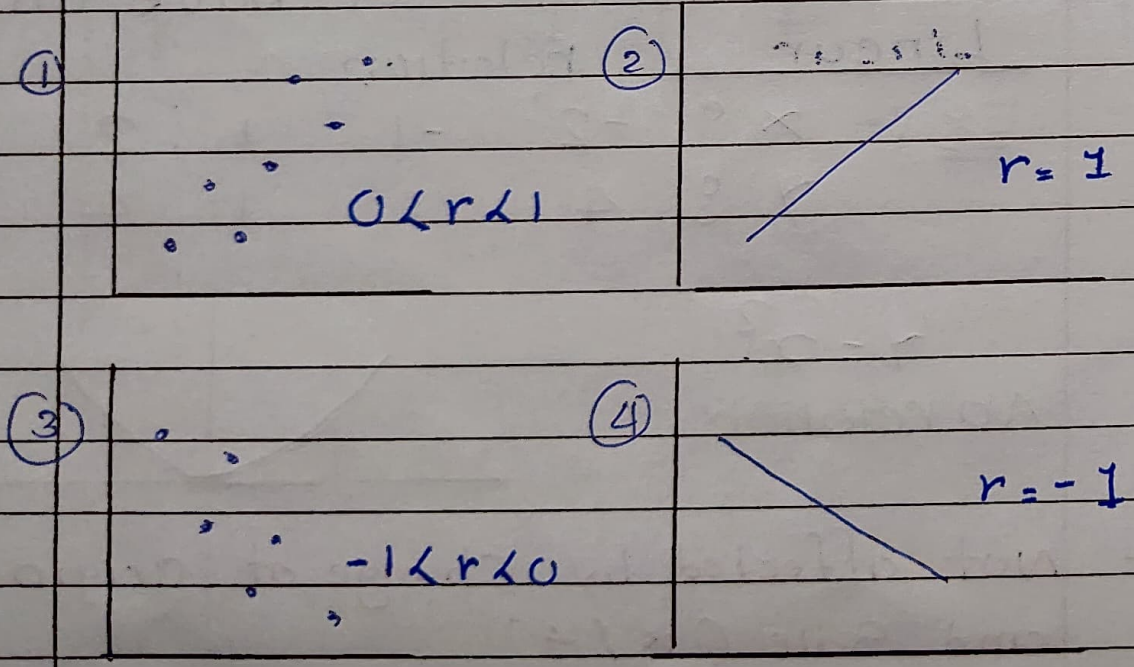


— Not affected by change of origin (+/-) and Scale ($\times / \div$)

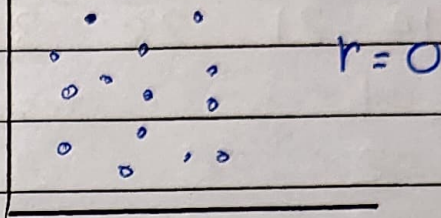
* Method to Calculate Correlation

- (1) Scatter Diagram: Cannot Measure value of r
- (2) Karl Pearson's Method: Best (Product moment)
- (3) Rank Correlation: For Qualitative data
- (4) Concurrent Deviation: Quickest Method

(1) Scatter Diagram



(5)



(2) Karl Pearson's Method (Covariance method)

$$r = \frac{\text{Cov.}(x, y)}{S_x \cdot S_y}$$

$$\begin{aligned} \text{Cov}(x, y) \\ = \frac{\sum (x - \bar{x})(y - \bar{y})}{n} \end{aligned}$$

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}}$$

$$\begin{aligned} \text{Cov}(x, y) \\ = \frac{\sum xy - n\bar{x}\bar{y}}{n} \end{aligned}$$

$$r = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{\sum x^2 - n(\bar{x})^2} \sqrt{\sum y^2 - n(\bar{y})^2}}$$

$$\begin{aligned} \text{Cov}(x, y) \\ = \frac{n\sum xy - \sum x \sum y}{n^2} \end{aligned}$$

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2} \sqrt{n\sum y^2 - (\sum y)^2}}$$

<u>Ex:-</u>	x	0	5	10	15	20	25
	y	0	7	14	21	28	35

$$\Sigma x = 75$$

$$\Sigma y = 105$$

$$\Sigma xy = 1925$$

$$\Sigma x^2 = 1375$$

$$\Sigma y^2 = 2645$$

$$r = \frac{5 \times 1925 - 75 \times 105}{\sqrt{5 \times 1375 - (75)^2} \sqrt{5 \times 2645 - (105)^2}}$$

$$= \frac{1750}{\sqrt{1250} \sqrt{2450}}$$

$$= \frac{1750}{\sqrt{1250} \sqrt{2450}} = 1$$

* Shift of origin

Given - $\sum (x-a)^2$, $\sum (y-b)^2$
 $\sum (x-a)(y-b)$ when

$$a \neq \bar{x}$$

$$b \neq \bar{y}$$

$$\text{let } u = x - a$$

$$v = y - b$$

$$r_{xy} = r_{uv} = \frac{n \sum uv - \sum u \sum v}{\sqrt{n \sum u^2 - (\sum u)^2} \sqrt{n \sum v^2 - (\sum v)^2}}$$

$$\sum u = \sum x - na$$

$$\sum v = \sum y - nb$$

(3) Rank Correlation

— Qualitative data

— Formula for non-repeated Rank

$$R = 1 - \frac{6 \sum d^2}{n(n^2-1)}$$

where $d = R_x - R_y$

$$\boxed{\text{Note : } \sum d = 0}$$

Ex:

X	Y	R_x	R_y	d	d^2
10	3	1	3	-2	4
9	5	2	2	0	0
7	9	3	1	2	4

$$R = 1 - \frac{6 \sum d^2}{n(n^2-1)} = 1 - \frac{6(8)}{3(9-1)}$$

$$= 1 - \frac{48}{24} = 1 - 2 = -1$$

→ For repeated rank

$$R = 1 - \frac{6(\sum d^2 - CF)}{n(n^2-1)}$$

when CF = (Correction Factor)

$$CF = \frac{\sum m^3 - m}{12}$$

$$\text{If } m=2, CF = \frac{2^3 - 2}{12} = 0.5$$

$$\text{If } m=3, CF = \frac{3^3 - 3}{12} = 2$$

(4) Concurrent Deviation Method

- Quickest Method but Cannot take magnitude of Correlation Seriously

$$r = \pm \sqrt{\frac{+ 2C - n}{n}}$$

where, C = no. of Concurrent Deviation

n = Pair of deviation Compared.

Ex. x : 150 + 154 + 160 + 172
 y : 200 - 180 - 170 - 160
C = 0, n = 3

$$r = \sqrt{\frac{2(0) - 3}{3}} = \sqrt{-1} = -1$$

Imp Note

(1) $n = \text{no. of obs.} \neq$

(2) If $C = 0$, $r = -1$

If $C = n$, $r = 1$

If $C = \frac{n}{2}$, $r = 0$

→ * Correlation Coefficient is independent of Change of Origin and Scale but

If $u = ax + c$, $v = by + d$

(1) $r_{uv} = r_{xy}$, If a, b are of Same Sign

(2) $r_{uv} = -r_{xy}$, If a, b are of different Sign

Ex:- $r_{xy} = -0.2$, $u = 3x - 5$

$v = -2y + 100$

$$r_{uv} = 0.2$$

$$r_{bc,y} = 0.8$$

$$300 + 2u + 4 = 0, \quad -3v - 8v + 100 = 0$$

$$2u = -300 - 4$$

$$v = \frac{-3}{8} + \frac{100}{8}$$

$$u = \frac{-300}{2} - \frac{4}{2}$$

$$r_{uv} = 0.8$$

* Imp Note

① If $y = a + bxc$ where $b > 0$ then
 $r = +1$

② If $y = a + bxc$ where $b < 0$ then
 $r = -1$

//_

→ * Imp Formula

(1) Coefficient of determination
 $= r^2 \quad 0 \leq r^2 \leq 1$

(2) Coefficient of non-determination
 $= 1 - r^2$

(3) Coefficient of alienation
 $= \sqrt{1 - r^2}$

(4) Percentage of explained Variance
 $= r^2 \times 100$

(5) Percentage of unexplained Variance
 $= (1 - r^2) \times 100$

(6) Standard Error
 $= \frac{1 - r^2}{\sqrt{n}}$

(7) Probable Error
 $= \frac{1 - r^2}{\sqrt{n}} \times 0.6745$

(8) Probable Error and Standard Error both are use for -
determining the Reliability
of Correlation Coefficient

(9) Probable limit
 $= r - PE, r + PE$

(10) For x, y Correlated Variable.

$$V(x+y) = V(x) + V(y) + 2(\text{cov}(x, y))$$

$$V(x-y) = V(x) + V(y) - 2(\text{cov}(x, y))$$

(11) For x, y Uncorrelated Variables

$$V(x \pm y) = V(x) + V(y)$$

//_

→ * Bivariate Data

— Also known as Bivariate -
frequency distribution or Joint
frequency distribution or Two
way distribution.

— Ex:

<u>Stat</u> <u>Math.</u>	0-10	10-20	20-30	Total
0-10	5	3	8	16
10-20	0	1	5	6
20-30	3	2	8	13
Total	8	6	21	35

— In m rows, n Columns: $n \times m$
Cells.

— Univariable distribution

① Marginal → 2

② Conditional → $m+n$

*

Ch: 3 B

Regression Analysis

Regression Analysis

"Estimate or Predict"

2 Types of variables

(1) Dependent

(2) Independent

Correlation is a relative measure

Regression is an absolute measure

Regression Line

Regression Equation
of y on x

$$\rightarrow y - \bar{y} = b_{yx}(x - \bar{x})$$

$$\rightarrow \hat{y} = a + b_{yx}x$$

Regression Equation
of x on y

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$\hat{x} = a + b_{xy}y$$

→ b_{yx} = Coefficient of Regression of y on x ,

Slope of Regression of line of y on x

b_{xy} = Coefficient of Regression of x on y

Slope of Regression of line of x on y

→ Regression Eq. of y on x is obtain by Scatter diagram by taking sum of square of vertical distance minimum by the method of least square

Regression Eq. of x on y is obtain by Scatter diagram by taking sum of square of Horizontal distance minimum by the method of least square

//_

→ Regression Coefficient of y on x

$$b_{yx} = \frac{\text{Cov}(x, y)}{S_x^2}$$

$$b_{yx} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

$$b_{yx} = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n(\bar{x})^2}$$

$$b_{yx} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$b_{yx} = r \cdot \frac{S_y}{S_x}$$

— Properties

— Correlation is Symmetric but regression is Not.

— $b_{xy} = r \cdot \frac{S_x}{S_y}$, $b_{yx} = r \cdot \frac{S_y}{S_x}$

$$b_{xy} \cdot b_{yx} = r^2$$

— $r = \pm \sqrt{b_{xy} \cdot b_{yx}}$

$$r_{xy} = r_{yx}$$

$$b_{xy} \neq b_{yx}$$

Ex :- $b_{xy} = 4$, $b_{yx} = 0.14$

$$r = \sqrt{4 \times 0.14}$$

$$r = 0.7483$$

— r , b_{xy} , b_{yx} are of Same Sign

— $-1 \leq r \leq 1$ | $-\infty < b_{xy} < \infty$

$0 \leq r^2 \leq 1$ | $-\infty < b_{yx} < \infty$

- If one regression Coefficient is Greater than 1 then other must be less than 1 Since $0 \leq r^2 \leq 1$

— Note

$$\begin{array}{c|c} y \text{ on } x & x \text{ on } y \\ (y - \bar{y}) = b_{yx}(x - \bar{x}) & (x - \bar{x}) = b_{xy}(y - \bar{y}) \end{array}$$

$$\bar{x} = \bar{x}, \bar{y} = \bar{y}$$

$(\bar{x}, \bar{y})$ is a Common Solution i.e. Point of intersection of two regression line. Therefore, Required regression Equation will be Satisfied by $(\bar{x}, \bar{y})$

— Note :-

In regression analysis, the difference between the observed value and the estimated value is known as Error or Residue.

//_

Standard Error of estimate of
 x on $y = S_{xx} \sqrt{1-r^2}$

Standard Error of estimate of
 y on $x = S_y \sqrt{1-r^2}$

If $r = \pm 1$ then $SE = 0$

— Regression Coefficient is not
affected by Change of origin
by affected by Change of Scale

If $u = ax + b$, $v = cy + d$

then $b_{uv} = b_{xy} \times \frac{a}{c}$

$b_{vu} = b_{yx} \times \frac{c}{a}$

Ex :- $u = 3x + 10$, $v = -4y - 20$

$b_{xy} = 0.8$

$b_{uv} = b_{xy} \times \frac{3}{-4} = 0.8 \times \frac{3}{-4}$
 $= -0.6$

Q) = Reg. Eq. are

$$\underline{x + 2y = 5}$$

$$\underline{2x + 3y = 8}$$

Y on x

X on Y

$$b_{yx} = -\frac{1}{2}$$

$$b_{xy} = -\frac{3}{2}$$

$$r^2 = b_{xy} \times b_{yx} = -\frac{1}{2} \times -\frac{3}{2}$$

$$\frac{3}{4} = 0.75 < 1$$

$$r^2 = \frac{3}{4}$$

$$r = \sqrt{\frac{3}{4}}$$

$$\underline{x + 2y = 5}$$

X on Y

$$b_{xy} = -2$$

$$\underline{2x + 3y = 8}$$

Y on X

$$b_{yx} = -\frac{2}{3}$$

$$r^2 = b_{xy} \times b_{yx} = -2 \times -\frac{2}{3} = \frac{4}{3} > 1$$

wrong

Correct $b_{yx} = -\frac{1}{2}$, $b_{xy} = -\frac{3}{2}$

$$r^2 = \frac{3}{4} \therefore r = \sqrt{\frac{3}{4}}$$

//_

* Steps to Identify Regression Equation

Step 1 Consider any one equation of x on y and other of y on x and find b_{xy} and b_{yx} .

Step 2 (1) If b_{xy} , b_{yx} are of different Sign then data is in-Correct

(2) If of same Sign then find $r^2 = b_{xy} \cdot b_{yx}$.

Step 3 (1) If $r^2 \leq 1$ then assumption is Correct

(2) If $r^2 > 1$ then assumption is in-Correct.

- To get Correct value $b_{xy} = \frac{1}{b_{yx} \text{ (wrong)}}$

$$b_{yx} = \frac{1}{b_{xy} \text{ (wrong)}}$$

$$r^2 = \frac{1}{r^2 \text{ (wrong)}}$$

//_

Note :-

(1) If $r=0$ then $\text{Cov}(x,y)=0$, $b_{yx}=b_{xy}=0$

--> Regression equation are
 x only $\rightarrow x = \bar{x}$ and y on $x \rightarrow y = \bar{y}$

--> Regression line becomes -
Perpendicular. $(90^\circ, \perp)$

(2) If $r = \pm 1$ then

(0°)

--> Regression Line are identical

$r=0$		$r = \pm 1$
Angle = 90°		Angle = 0°

(3) In Regression analyses we get a functional relation between two Variable using average.

*

ch :- 4

Index Numbers

—

Index Number —

— Special type of Avg.

— Expressed as ratio

— Calculated as Percentage

— Unit Free (Pure Number)

*

Type of Index Numbers

(1)

Price Index Number

(2)

Quantity Index Number (Volume)

(3)

Value Index Number

(Value = Price $\times$ Quantity)

//_

* Notation :-

0 — Base year

1 — Current year

I_{01} — Index Number for year 1 with base as 0

I_{10} — Index Number for year 0 with base as 1

I_{11} — 100.

* Price Index Number

(1) Price Relative :- PR :-

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

//_

(2) Method of Aggregate :-

(i) Simple Aggregate of P.I.N.

$$I_{01} = \frac{\sum P_1}{\sum P_0} \times 100$$

- Unweighted
- With Unit rest all are Unit Free

(ii) Weighted Aggregate.

$$I_{01} = \frac{\sum P_1 w}{\sum P_0 w} \times 100$$

- Weight is most Important in I.N

//_

Laspeyres's P.I.N = $L_{01} = \frac{\sum P_1 q_0}{\sum P_0 q_0} \times 100$

— $w = q_0$

Pasche's P.I.N = $P_{01} = \frac{\sum P_1 q_1}{\sum P_0 q_1} \times 100$

— $w = q_1$

Fisher's P.I.N. = $F_{01} = \sqrt{L_{01} \times P_{01}}$

— Best Index Number

— F_{01} is G.M. of L_{01} and P_{01}

Bowley's P.I.N = $B_{01} = \frac{L_{01} + P_{01}}{2}$

— B_{01} is A.M of L_{01} and P_{01}

Marshall + Edgeworth P.I.N.

$$= \frac{\sum P_1 \left(\frac{q_0 + q_1}{2} \right)}{\sum P_0 \left(\frac{q_0 + q_1}{2} \right)} \times 100$$

$$= \frac{\sum P_1 q_0 + \sum P_1 q_1}{\sum P_0 q_0 + \sum P_0 q_1} \times 100$$

— Best approximation to Fisher's I.N.

Note

G.M. is best Average in the -
Construction of I.N. but -
Practically we use A.M., because
G.M. is difficult to compute

(3) Relative Method

(i) Simple A.M. of P.R

$$I = \frac{\sum P.R}{n}$$

(ii) Weighted A.M. of P.R

$$I = \frac{\sum P.R \times w}{\sum w}$$

(iii) Simple G.M. of P.R

$$I = (\prod P.R)^{1/n}$$

$$\log I = \frac{1}{n} \left[\sum \log \frac{P_i}{P_0} + 2 \right]$$

* ($\prod$ = Product)

⑧ Cost of Living Index Number
(CLIN)

- Wholesale P.I.N.,
- Consumer P.I.N.,
- General P.I.N.,

$$CLIN = \frac{\sum IW}{\sum W} \quad \text{or} \quad \frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times 100$$

— Based on Laspeyre's Method

Ex Dearness Allowance (DA)

	Salary	CLIN
2024	50,000	150
2029	55,000	180

150 → 50,000

180 → ? = 60,000

- 55,000

5000

(D.A)

//_

$$* \text{ Real Income (Deflation) } = \frac{\text{Income}}{\text{CLIN}} \times 100$$

<u>Ex:</u>		Salary	CLIN
2000		40,000	150
1990		25,000	100

$$\text{Real Income} = \frac{40,000 \times 100}{150}$$

$$= 26,666.67$$

$$* \text{ Purchasing Power } = \frac{1}{\text{CLIN}} \times 100$$

* * For Quantity IN, In PIV formula,

$$P \leftrightarrow Q$$

//_

— Base Shift $IIV = \frac{\text{Current Yr } IIV}{\text{New Base Yr } IIV} \times 100$

— Fixed and Chain Base Method

— Fixed base Method: Base Year is Fixed

$$FBIIV = \frac{\text{Current Yr Price}}{\text{Base Yr Price}} \times 100$$

— Chain base Method: Base Year is changed Previous Yr.

$$CBIIV = \frac{\text{Current Yr Price}}{\text{Previous Yr Price}} \times 100$$

(Link Related)

* FBI/V $\rightarrow$ CBI/V

Yr	0	1	2	3
FBI/V	80	120	180	200
CBI/V	100	150	150	111.11

Note:- To find Chain Base Method
take 1st yr TN = 100

* CBI/V $\rightarrow$ FBI/V

Yr	0	1	2	3
CBI/V	60	150	180	250
FBI/V	60	90	162	405

Note:- (1) for 1st yr FBI/V = CBI/V
for that yr

(2)

$$\text{FBI/V} = \frac{\text{Previous yr FBI/V} \times \text{Current yr CBI/V}}{100}$$

* CBIV $\rightarrow$ CIN

yr	0	1	2	3
CBIV	60	150	180	250
CIN	100	150	270	675

Note:- (1) 1st yr CIN = 100

(2)

$$CIN = \frac{\text{Previous yr CIN} \times \text{Current yr CBIV}}{100}$$

$\rightarrow$ Splicing

Ex:-

	IN Base 2001		IN Base 2003
2001	100		66.67
2002	120		
2003	150		100
2004	210		140
2005			180

→ * Value Index Number

$$V = \frac{\sum P_1 Q_1}{\sum P_0 Q_0} \times 100$$

Test	Simple Aggregate	L	P	F	B	M.E.	Weight Gm of Price relative	Simple Gm of Price relative
Unit Test	X	✓	✓	✓	✓	✓	✓	✓
Time - reversal test	✓	X	X	✓	X	✓	✓	✓
Factor - reversal test	X	X	X	✓	X	X	X	X
Circular test	✓	X	X	X	X	X	X	✓

* ch:- 5 A

Probability

— $P(A) = \frac{m}{n}$

— * Coin

— 2 Coin (Trick $\frac{11^2}{2^2}$)

$U = \{HH, HT, TH, TT\}$

No. of Head/ Tail	0	1	2
P	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

— 3 Coin (Trick $\frac{11^3}{2^3}$)

No. of Head/ Tail	0	1	2	3
P	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

— 4 Coin (Trick $\frac{11^4}{2^4}$)

No. of Head Tail	0	1	2	3	4
P	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

— * Cards

52 Cards

Black (26)

(26) Red



Club



Spade

Diamond



Heart



A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, k

Face Cards = (J, Q, k) = $3 \times 4 = 12$

Honour Cards = (A, J, Q, k) = $4 \times 4 = 16$

* Dice

— 2 Dice

$(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)$
 $(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)$
 $(3,1), (3,2), (3,3), (3,4), (3,5), (3,6)$
 $(4,1), (4,2), (4,3), (4,4), (4,5), (4,6)$
 $(5,1), (5,2), (5,3), (5,4), (5,5), (5,6)$
 $(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)$

(1)

Sum of
no.

2	3	4	5	6	7	8	9	10	11	12
---	---	---	---	---	---	---	---	----	----	----

times

1	2	3	4	5	6	5	4	3	2	1
---	---	---	---	---	---	---	---	---	---	---

(2)

$F = S$

$F > S$

$F < S$

$\frac{6}{36}$

$\frac{15}{36}$

$\frac{15}{36}$

(3)

Product is 12 = $\frac{4}{36}$

$(6,2), (2,6), (4,3), (3,4)$

→ * Leap Year

366 days = 52 weeks + 2 odd days

M, T	—	
T, W		
W, Th		7
Th, F	—	ways
F, Sat		
Sat, Sun		
Sun, M		

- (1) $P(53 \text{ Sundays}) = 2/7$
- (2) $P(53 \text{ Sundays } \overset{\text{or}}{\text{and}} 53 \text{ Saturdays}) = 3/7$
- (3) $P(53 \text{ Sundays and } 53 \text{ Saturdays}) = 1/2$
- (4) $P(53 \text{ Sundays or } 53 \text{ Fridays}) = 4/7$
- (5) $P(53 \text{ Sundays and } 53 \text{ Fridays}) = 0$

//_

* Rule of addition

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

↓
or
Atleast
one

↓
and
Both

Ex :- From 1 to 200

$P(\text{No. is divisible by } 6 \text{ or } 8)$

$$P(6 \cup 8)$$

$$= P(6) + P(8) - P(6 \cap 8)$$

$$= 33 + 25 - 8$$

$$= \frac{50}{200} = \frac{1}{4}$$

//_

→ * Formula

$$(1) \quad P(A \cap B) = P(A) + P(B) - P(A \cup B)$$

$$(2) \quad P(A') = P(A^c) = 1 - P(A)$$

$$(3) \quad P(A - B) = P(A \cap B') = P(A) - P(A \cap B)$$

$$(4) \quad P(B - A) = P(A' \cap B) = P(B) - P(A \cap B)$$

$$(5) \quad P(A' \cap B') = 1 - P(A \cup B)$$

$$(6) \quad P(A' \cup B') = 1 - P(A \cap B)$$

$$(7) \quad P(A' \cup B) = 1 - P(A \cap B')$$

$$(8) \quad P(A \cup B') = 1 - P(A' \cap B)$$

— Odd in the favor of A is $P:q$ then

$$P(A) = \frac{P}{P+q}, \quad P(A') = \frac{q}{P+q}$$

Ex:- Odd in the favor of A is $2:3$ then

$$P(A) = \frac{2}{5}, \quad P(A') = \frac{3}{5}$$

— odd Against A is $m:n$ then

$$P(A) = \frac{n}{m+n}, \quad P(A') = \frac{m}{m+n}$$

Ex:- Odd against A is $3:5$ then

$$P(A) = \frac{5}{8}, \quad P(A') = \frac{3}{8}$$

//_

* Conditional Probability

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

* Rule of Compound Probability

$$P(A \cap B) = P(A/B) \times P(B)$$

$$P(A \cap B) = P(B/A) \times P(A)$$

- Ex: - $P(A) = 0.7$, $P(B) = 0.4$, $P(A \cap B) = 0.3$

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.7 + 0.4 - 0.3 \\ &= 0.8 \end{aligned}$$

$$\begin{aligned} (1) P(A \cap B') &= P(A) - P(A \cap B) = 0.7 - 0.3 \\ &= 0.4 \end{aligned}$$

$$\begin{aligned} (2) P(A' \cup B) &= 1 - P(A \cap B') \\ &= 1 - 0.4 = 0.6 \end{aligned}$$

_1_1_

$$\begin{aligned} (3) P(A' \cap B) &= P(B) - P(A \cap B) \\ &= 0.4 - 0.3 \\ &= 0.1 \end{aligned}$$

$$(4) P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.3}{0.4} = \frac{3}{4}$$

$$\begin{aligned} (5) P(A'/B') &= \frac{P(A' \cap B')}{P(B')} \\ &= \frac{1 - P(A \cup B)}{1 - P(B)} = \frac{1 - 0.8}{1 - 0.4} \\ &= \frac{0.2}{0.6} = \frac{1}{3} \end{aligned}$$

$$\begin{aligned} (6) P(A/B') &= \frac{P(A \cap B')}{P(B')} \\ &= \frac{0.4}{0.6} = \frac{2}{3} \end{aligned}$$

→ * A and B are Independent events

$$P(A/B) = P(A) = P(A/B')$$

$$P(B/A) = P(B) = P(B/A')$$

$$P(A \cap B) = P(A) \times P(B) \rightarrow \text{Rule of Multiplication for Independent}$$

— Note :- If A and B are Independent then.

- (1) A and B' are Independent
- (2) A' and B are Independent
- (3) A' and B' are Independent

//_

— Three events A, B, C are mutually Independent if,

$$P(A \cap B) = P(A) \cdot P(B)$$

$$P(A \cap C) = P(A) \cdot P(C)$$

$$P(B \cap C) = P(B) \cdot P(C)$$

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

= No. of Condition to check for 3 events to be mutually Independent is 4

* No. of Condition to check for n events to be mutually Independent =

$$2^n - 1 - n$$

— Notation

$$A + B = A \cup B$$

$$AB = A \cap B$$

//_

* Theory :-

* Types :-

1) Subjective :- Biased,
Used in management

2) Objective :- Unbiased

* Types of Events :-

1) Elementary Event.

Event which cannot be divided
into smaller event.

2) Compound (Composite) Events :-

Made of two or more -
elementary events

3) Sure Events :-

$$P(A) = 1 \quad ; \quad P(A) = U$$

//_

4) Impossible Events:-

$$P(A) = 0 \quad ; \quad P(A) = \phi$$

5) Chance Events:-

$$0 < P(A) < 1$$

6) Equally Likely Events:-

A and B are equally Likely events if $P(A) = P(B)$

7) Mutually Exclusive Events:-

If an event A and B are mutually Exclusive then,

$$1) P(A \cup B) = P(A) + P(B)$$

$$2) P(A - B) = P(A)$$

$$3) P(B - A) = P(B)$$

$$4) P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

8) Collectively Exhaustive Events

$$P(A \cup B) = 1$$

$$P(A \cup B \cup C) = 1$$

- Imp Notes:-

If A and B are two events with ~~not~~ non-Zero Probability.

Mutually Exclusive and Independent

$$P(A \cap B) = 0, \quad P(A \cap B) = P(A) \times P(B)$$

$$0 = P(A) \times P(B)$$

$$P(A) = 0$$

$$P(B) = 0$$

Contradiction.

Two events with Non-Zero Probability cannot be simultaneously mutually exclusive and independent both.

* Definition of Probability

Classical
or
mathematical

Empirical
or
Statistical

Axiomatic

$$= \frac{m}{n}$$

$$= \lim_{n \rightarrow \infty} \frac{m}{n}$$

No Formula
But we have
Rule of set

$$\begin{aligned} \text{Eg. } P(A \cup B) \\ = P(A) + P(B) - P(A \cap B) \end{aligned}$$

* Limitation of Mathematical/Classical definition.

- All the outcome must be known in advance.
- If $n \rightarrow \infty$ then we cannot use this definition.
- Elementary events are equally likely

* ch:- 5 B

Random Variable

Ex: 3 Coins $\left(\frac{11^3}{2^3} \right)$

<u>X = No. of H</u>	<u>P = f(x)</u>
0	$1/8$
1	$3/8$
2	$3/8$
3	$1/8$

X = Random Variable

f(x) = Probability Function
Frequency Function

Type of Variable

	Discrete variable	Continuous variable
→	Whole numbers	Real numbers in Interval.
→	Finite values	Infinite values
→	Probability function is known as - Probability mass function.	Probability function is known as - Probability density function.
→	Condition of PMF (1) $f(x) \geq 0$ (2) $\sum f(x) = 1$	Condition of PDF (1) $f(x) \geq 0$ (2) $\int f(x) dx = 1$

→ * Expected value = $E(x) = \bar{x}$

- It is also known as mean

- Formula :-

$$E(x) = x_1 P_1 + x_2 P_2 + \dots + x_n P_n$$
$$= \sum x_i P_i$$

- Note :- (1) If all Probabilities are equal then,

$$E(x) = \text{Avg. of } x$$

(2) If all Probability are equal and x values in A.P. then.

$$E(x) = \frac{a + l}{2}$$

Ex :- Dice

<u>x</u>	<u>P(x)</u>
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

$$E(x) = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6}$$

$$= \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$$

Trick

$$E(x) = \frac{1+6}{2} = 3.5$$

Ex:- Game Fees = ₹ 6

$$6 \rightarrow ₹ 30$$

$$x = P$$

$$30 = 1/6$$

$$E(x) = 30 \times 1/6 = 5 - 6 = -1$$

— Note:- (1) For Game $\Rightarrow f(x) = 0$

(2) For Game Fees Probability
First Find $E(x)$ and
at the end subtract game
Fees to get actual value
of $E(x)$.

Ex:

A box contains 12 lamps of which 5 are defective. A man selects 3 lamps at random then what is the expected no. of defective lamps.

X P

$$0 \quad {}^7C_3 = 35/220$$

$$1 \quad {}^5C_2 \times {}^7C_1 = 105/220$$

$$2 \quad {}^5C_1 \times {}^7C_2 = 70/220$$

$$3 \quad {}^5C_3 = 10/220$$

$$E(X) = 1.25$$

Trick

$$12 \rightarrow 5$$

$$3 \rightarrow 9$$

$$= 1.25$$

— Properties of $E(x)$

(1) $E(c) = c$

(2) $E(ax \pm by \pm c) = aE(x) \pm bE(y) \pm c$

$E(x) = 3, E(y) = 5$

$E(2x + 3y - 10)$

$= 2(E(x)) + 3(E(y)) - 10$

$= 2(3) + 3(5) - 10 = 6 + 15 - 10$

$= 11$

(3) If x and y are Independent then,

$$E(x \cdot y) = E(x) \cdot E(y)$$

(4) $E(x - \bar{x}) = 0$

$= E(x) - E(\bar{x}) = 0$

$= \bar{x} - \bar{x} = 0$

Note For Dice

(1) For Sum

$$E(\text{one die}) = 3.5$$

$$\begin{aligned} E(\text{Sum of two dice}) &= E(x+y) \\ &= E(x) + E(y) = 3.5 + 3.5 = 2 \times 3.5 \end{aligned}$$

$$E(\text{Sum of } n \text{ dice}) = n \times 3.5$$

(2) $E(\text{Product of two dice})$

$$\begin{aligned} E(xy) &= E(x) \cdot E(y) \\ &= 3.5 \times 3.5 \\ &= 12.25 \end{aligned}$$

→ * Standard deviation and Variance

$$\sigma^2 = V(x) = \frac{\sum (x_i - \bar{x})^2}{n} = \text{Avg. of } (x_i - \bar{x})^2$$

$$V(x) = E((x - \bar{x})^2)$$

$$V(x) = E(x^2) - (E(x))^2$$

$$\sigma = SD = \sqrt{V(x)} \\ = \sqrt{E(x^2) - (E(x))^2}$$

$$V = SD^2$$

— Properties of SD and V(x)

$$(1) \quad V(c) = 0, \quad SD(c) = 0$$

$$(2) \quad SD(ax + by + c) = |a|SD(x) + |b|SD(y) \\ SD_x = 3, \quad SD_y = 7 \\ SD(-2x - 3y + 4) = 2 \times 3 + 3 \times 7 = 27$$

$$(3) \quad V(ax + by + c) = a^2V(x) + b^2V(y) \\ V(5x - 2y + 100) = 25V(x) + 4V(y) \\ V(x) = 4, \quad V(y) = 9 \quad | \quad = 25 \times 4 + 4 \times 9 \\ | \quad = 181$$

//_

Example:-

<u>X</u>	<u>P</u>	<u>X²</u>
3	0.2	9
9	0.6	81
10	0.2	100

$$E(x) = 8$$

$$\begin{aligned} E(2x+1) &= 2E(x) + 1 \\ &= 2(8) + 1 \\ &= 17. \end{aligned}$$

$$E(x^2) = 70.4$$

$$\begin{aligned} V(x) &= E(x^2) - (E(x))^2 \\ &= 70.4 - 64 \\ &= 6.4 \end{aligned}$$

$$\begin{aligned} SD &= \sqrt{6.4} \\ &= 2.53 \end{aligned}$$

— Uniform Distribution (For Discrete Variable)

$$X = 1, 2, 3, \dots, n$$

$$P.M.F - f(x) = \frac{1}{n}$$

$$E(X) = \frac{1+n}{2}$$

$$V(X) = \frac{n^2-1}{12}$$

$$SD = \sqrt{\frac{n^2-1}{12}}$$

//_

* Ch:- 6

Theoretical Distribution

- Binomial Distribution

→ Bernoulli Trial

(1) Exactly two possible group of Outcome "Success" and "Failure"

(2) Probability of Success is independent of Previous Success

→ Binomial is Bernoulli Trial repeated n times

∴ Binomial is an extension of Bernoulli

→ n = no. of times experiment repeated

$X = 0, 1, 2, 3, \dots, n$

—> X is discrete variable

$X \sim B(n, p)$
Follows

—> P. m. f.

$$f(x) = {}^nC_x p^x q^{n-x}$$

x = no. of Success

P = Probability of Success

q = Probability of Failure

$$P + q = 1$$

Sum of Power of $P, q = n$

—> Two Parameters n and P

—> Mean = np

—> Variance = npq

mean > Variance

—> SD = $\sqrt{npq}$

//_

→ Maximum Value of Variance = $n \times \frac{1}{2} \times \frac{1}{2} = \frac{n}{4}$
 ($p = q = \frac{1}{2}$)

Imp $q = \frac{V}{n} \quad n = \frac{n}{p}$

Ex :- mean = 2 $V(x) = 1.2$

$np = 2$ $npq = 1.2$

$2q = 1.2$

$n = \frac{2}{0.4} = 5$ $q = \frac{1.2}{2} = 0.6$

$p = 0.4$

Ex :- mean = 48 $V(x) = 19.2$

$q = \frac{19.2}{48} = 0.4$

$n = \frac{48}{0.6} = 80$

$p = 0.6$

//_

→ * $Mode = (n+1)P$

Uni-Modal

Bi-Modal

$(n+1)P$ is any decimal number

$(n+1)P$ is an integer

Mode = Integ Part of $(n+1)P$

Mode = $(n+1)P$, $(n+1)P + 1$

Ex:- $n=6, P=\frac{1}{4}$

Ex:- $n=7, P=\frac{1}{4}$

$(n+1)P = 7 \times \frac{1}{4}$

$(n+1)P = 8 \times \frac{1}{4} = 2$

$= 1.75$

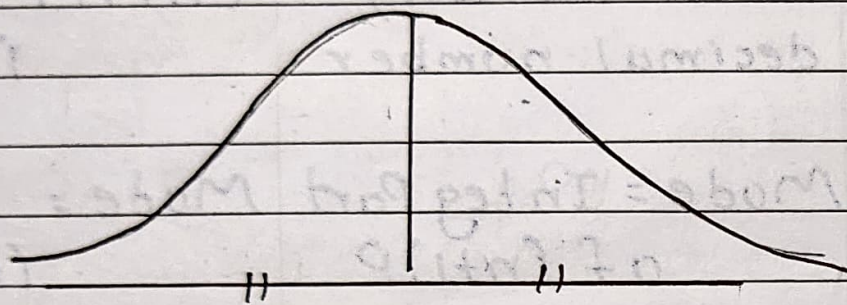
Mode = 2, 1

Mode = 1

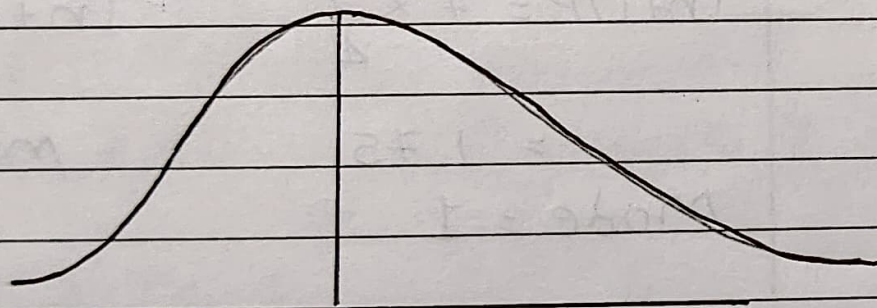
//_

→ Symmetric and Asymmetric

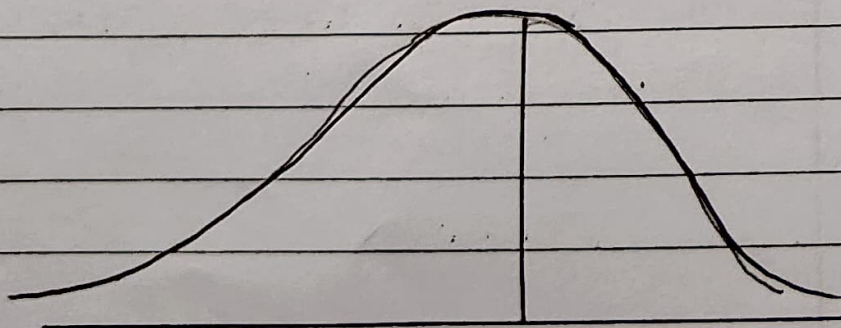
- Symmetric $P = Q = 0.5$
(mean = median = mode)



- Positive Skewed $P < Q$



- Negative Skewed $P > Q$



//_

→ Addition Property

If x, y are Independent variable and

$$x \sim B(n_1, p), y \sim B(n_2, p)$$

$$\text{then } x + y \sim B(n_1 + n_2, p)$$

Ex :- $x \sim B(10, \frac{1}{2})$

$$y \sim B(12, \frac{1}{2})$$

$$x + y \sim B(22, \frac{1}{2})$$

Trick

For Problem of binomial with $p = q = \frac{1}{2}$ then

$$f(x) = \frac{n C x}{2^n}$$

Note:-

$x \sim B(n, p)$ is also written as $(p+q)^n$ or $(q+p)^n$

Bernolli Trial "Success"
"or"
"Failure"

n times ↓

Discrete

Binomial Distribution

$n \rightarrow \infty$
 $p \rightarrow 0, q \rightarrow 1$ ↓

Discrete

Poission Distribution

Continuous Variable
 $p \rightarrow 0.5 \leftarrow q$ ↓

Normal Distribution

Note :-

Parameter $\rightarrow$ Population

Ex Population Mean
 μ

Statistics $\xrightarrow[\text{for}]{\text{use}}$ Sample

Ex Sample Mean
 $\bar{oc}$

Poissonian Distribution

(Rare events / Improbable events distribution)

$$n \rightarrow \infty, p \rightarrow 0, q \rightarrow 1$$

$$m = \text{mean} = np = \lambda$$

$$x = 0, 1, 2, 3, \dots$$

$$x \sim P(m)$$

→ Discrete Variable

→ P.m.f.

$$f(x) = \frac{e^{-m} m^x}{x!}$$

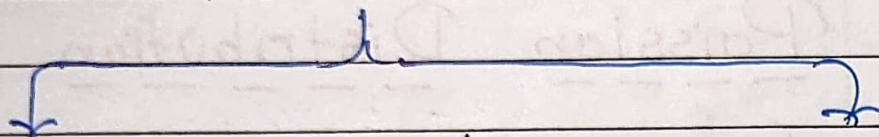
$$e = 2.7183$$

→ Variance = mean = Parameter = m

$$\rightarrow SD = \sqrt{m}$$

→ Mode = m

Mode = m



Uni-Modal

m is not an integer

Bi-modal

m is an integer

Mode = Integer
Part of m

Mode = m, m-1

Ex: m = 3.85
Mode = 3

Ex: m = 4
Mode = 4, 3

→ $P \rightarrow 0$, Always
 $P < 1$

Always Positive Skewed

→ Addition Property

If x and y are Independent
 $x \sim P(m_1)$, $y \sim P(m_2)$ then

$x + y \sim P(m_1 + m_2)$

— Ex :- $n = 400$
1% defective
 $P = 0.01$

$$m = n \times P = 400 \times 0.01 = 4$$

(Given $e^{-4} = 0.0183$)

$$\begin{aligned} \textcircled{1} P(\text{none defective}) &= P(0) \\ &= \frac{e^{-4} 4^0}{0!} \\ &= e^{-4} = 0.0183 \end{aligned}$$

$$\begin{aligned} \textcircled{2} P(1) &= \frac{e^{-4} 4^1}{1!} = 4 \times e^{-4} \\ &= 4 \times 0.0183 \\ &= 0.0732 \end{aligned}$$

$$\textcircled{3} P(2) = \frac{e^{-4} 4^2}{2!} = 0.1464$$

— Note :-

$$\textcircled{1} P(0) = e^{-m}$$

$$\textcircled{2} P(1) = m \cdot e^{-m}$$

//_

Ex: $m = 4$, $e^{-4} = 0.0183$

$$P(\text{At least } 4)$$

$$= 1 - P(0, 1, 2, 3)$$

$$= 1 - \left[e^{-m} + e^{-m}m + e^{-m}\frac{m^2}{2} + e^{-m}\frac{m^3}{6} \right]$$

$$= 1 - e^{-m} \left[1 + m + \frac{m^2}{2} + \frac{m^3}{6} \right]$$

$$= 0.5669$$

— Note :-

Condition to use Poisson Distribution

- (1) The Probability of having Success in this time interval is independent of time as well as earlier Success
- (2) Probability of having more than one Success is very Low.
- (3) Probability of Success in time interval is directly Proportional to time interval length.

Normal Distribution

→ Continuous Variable

→ $x \sim N(\overset{\text{Mean}}{\mu}, \overset{\text{S.D}}{\sigma})$ or $x \sim N(\overset{\text{Variance}}{\mu}, \sigma^2)$

→ P.d.f.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\rightarrow Z = \frac{x - \mu}{\sigma}$$

$$\begin{aligned}\bar{Z} &= \frac{\bar{x} - \mu}{\sigma} \\ &= \frac{\mu - \mu}{\sigma}\end{aligned}$$

$$\begin{aligned}\sigma_Z &= \frac{\sigma_x}{\sigma} \\ &= \frac{\sigma}{\sigma}\end{aligned}$$

$$\bar{Z} = 0$$

$$\sigma_Z = 1$$

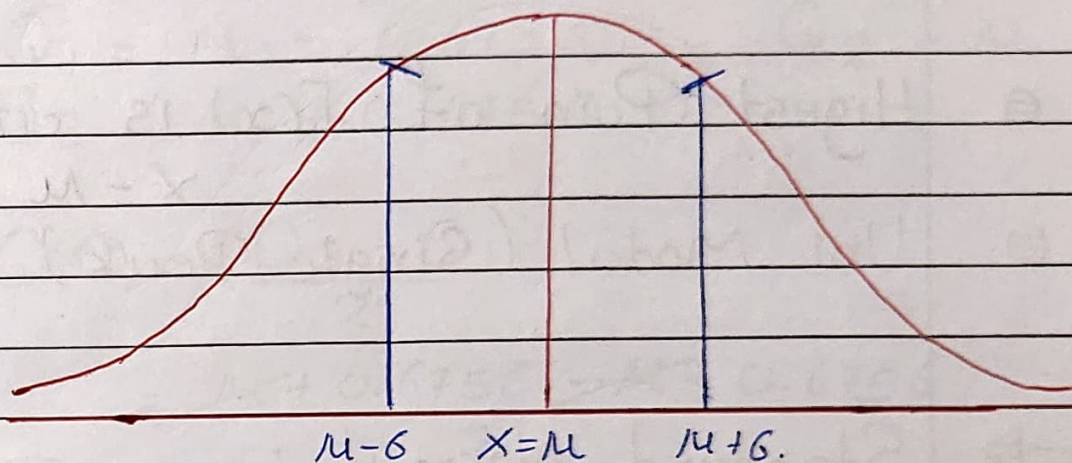
$$Z \sim N(0, 1)$$

Z is a Standard Normal Variable
P.d.f.

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

//_

→ Normal Curve $x \sim N(\mu, \sigma)$



$$-\infty < x < \infty$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Bell Shape
- Symmetric about $x = \mu$
- Total area Under the Curve = 1
- Asymptotic $\Rightarrow$ will never intersect x -axis
- Point of Inflection

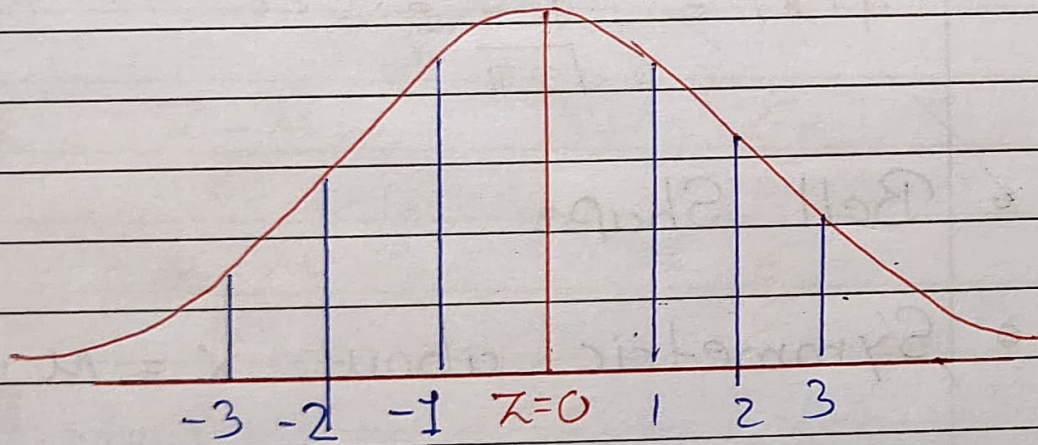
$$\mu - \sigma, \mu + \sigma.$$

⇒ Mean = Median = Mode = μ

⇒ Highest Point of $f(x)$ is at $x = \mu$

⇒ Uni Modal (Single Peak)

→ Standard Normal :



$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

⇒ Symmetric about $z=0$

⇒ Point of Inflection
 $-1, 1$

⇒ Maximum Value at $z=0$.

Formula:

$$Q_1 = M - 0.6756 \quad Q_2 = M$$

$$Q_3 = M + 0.6756$$

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

$$= \frac{M + 0.6756 - M + 0.6756}{2}$$

$$Q.D = 0.6756 \approx \frac{2}{3} 6$$

$$M.D = \frac{4}{5} 6$$

Q.D : M.D : S.D = 10 : 12 : 15

$$5 \times \left(\frac{2}{3} 6 : \frac{4}{5} 6 : 6 \right) = 10 : 12 : 15$$

Ex:- Q.D = 9, M.D = ?

$$Q.D : M.D$$

$$10 : 12$$

$$9 : \quad$$

$$= 10.8$$

Q. In normal distribution, what is the ratio of M.D and S.D?

$$\rightarrow \frac{MD}{SD} = \frac{12}{15} = 0.8$$

Im.P Probability:

$$P(0 < Z < 1) = 0.3413$$

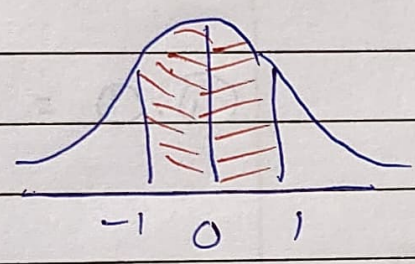
$$P(-1 < Z < 1) = 0.6826$$

$$P(0 < Z < 2) = 0.4772$$

$$P(-2 < Z < 2) = 0.9544$$

$$P(0 < Z < 3) = 0.4987$$

$$P(-3 < Z < 3) = 0.9974$$

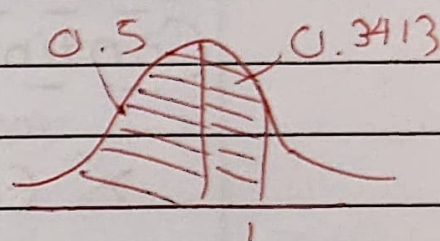


Except
0.0026
= 0.26%

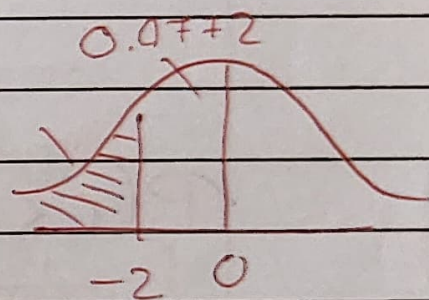
→ Cumulative Distribution Function (C.D.F)

$$F(u) = \phi(u) = P(-\infty < z < u)$$

Ex:- (1) $F(1) = P(Z < 1)$
 $= 0.5 + 0.3413$
 $= 0.8413$



(2) $F(-2) = P(Z < -2)$
 $= 0.5 - 0.4772$
 $= 0.0228$



→ Addition Property

If x, y are Independent variable

$$x \sim N(\mu_1, \sigma_1), y \sim N(\mu_2, \sigma_2)$$

then $x + y \sim N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$

Ex:- $x \sim N(10, 3), y \sim N(12, 4)$

then $x + y \sim N(22, 5)$

//_

→ Normal distribution is also known as Gaussian distribution.

* Uniform Distribution for Continuous Variable

$$X \in [a, b] \text{ P.d.f}$$

$$f(x) = \frac{1}{b-a}$$

$$P(c < x < d) = \frac{d-c}{b-a}$$

— Also known as Rectangular distribution.

* Two Methods of Fitting Normal

(1) Ordinate Method

(2) Area Method

→ For Continuous Variable Probability to take any fixed value is Zero. (since $\frac{1}{\infty} = 0$.)