-Measure of Central Tendancy

Arithmetic Mean

$\bar{x} = \dfrac{\text{Total sum of observation}}{\text{No. of observation}}$ Simple

In discrete series $= \dfrac{\Sigma f_i x_i}{\Sigma f_i} = \dfrac{\Sigma f_i x_i}{\Sigma f_i}$

- f_i = frequency
- x_i = variable. • Σ = summation
- Σf_i = sum of frequency.

In continuous series

$\bar{x} = \dfrac{\Sigma f_i x_i}{\Sigma f_i}$ (Here x_i = midvalue of classes & than multiply to f_i & Total of $\Sigma f_i x_i / \Sigma f_i$)

Measure of Central Tendancy & Dispersion

Step deviation Method $A + \dfrac{\Sigma u_i f_i}{\Sigma f_i} \times h$

- A = Assumed mean =
- $\Sigma u_i f_i = \Sigma \dfrac{(x_i - A)}{h}$ — height of class
- h = height of class = U.C.B − L.C.B.

* Tendancy *

* Properties of A.M. *

1. If All value of observation same than $\bar{x}$ is Also same to constant variable

Ex $k, k, k, k = \dfrac{4k}{4} = k$

2. If All observation are +, −, ÷ or × by same no. Than Mean is Also have same effect

3. $|x_i - \bar{x}|$ or Sum of deviation = 0 All observation Are subtracted from Mean & After Sum of all Ans is (0).

4. $\boxed{y = a + bx}$ origin shift, Scaling, Linear relation ship.

If we put x Mean in this than we Also put y mean & calculate from this Also

$\boxed{\bar{y} = a + b\bar{x}}$ In same eq. x & y replace by their mean.

5. Combine Mean

$\boxed{\bar{x} = \dfrac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}}$

Ans will come b/w both of mean value

Correcting Incorrecting Mean *

Correct Mean = Incorrect Mean + $\dfrac{(\text{correct value} - \text{wrong value})}{n}$

(1) $\bar{x}_c = \bar{x}_{ic} + \dfrac{(x_{i_c} - x_{i_w})}{n}$

(2) $\bar{x}_{Ic} = \dfrac{\Sigma x_i}{n}$

$\Sigma x_i(\text{sum}) = \bar{x}(\text{Incorrect}) \times n$ = Total sum

$\dfrac{\text{Total sum} - \text{Incorrect value} + \text{correct value}}{\text{No. of observation of}}$

Ans is Corrected Mean

A.M $\dfrac{}{\Sigma w_i}$

w_i = weight

Σw_i = sum of weight

x_i = observations

$w_i \cdot x_i$ & than Σ of or Sum of $w_i \cdot x_i = \Sigma w_i x_i$

* $\boxed{\text{weighted Arithmetic Mean}}$ *

$$\boxed{\text{weighted A.M}} = \frac{\Sigma w_i x_i}{\Sigma w_i}$$

w_i = weight

Σw_i = sum of weight

x_i = observations

$w_i . x_i$ & than Σ of or

sum of $w_i . x_i$ = $\Sigma w_i x_i$

Merits

- Properly defined or Jo definition me likha hua hai vhi hai (unambiguously)
- easy to comprehend.
- simple to calculate.
- Based on all observation
- should have certain Mathematical Properties
[It should be least affect by presence of Extreme observation.] X

Demerits

- It is very much affecting by extreme values.
- It can't be calculated by open-end class series.
- It can't be calculated graphically
- lead to wrong conclusion if one value of observation is missed.

Median

Tells centermost value.

[Calculate Median]

① 1st Arrange observation in Ascending or descending order

② For odd no. of observation

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

③ For even no. of observation

② $M = \frac{\left(\frac{N}{2} + \frac{N}{2}+1\right)}{2}$

Median in Discrete Series

① $M = \left(\frac{n+1}{2}\right)^{th}$ Term, or $\left(\frac{\frac{n}{2} + \frac{n}{2}+1}{2}\right)$ Term

After this

(x) value $=$ CF value $> \left(\frac{n+1}{2}\right)$

(x) value जो CF के सामने है और CF जो $\left(\frac{n+1}{2}\right)$ से बड़ा है।

Med $x =$ $\left[CF > \frac{n+1}{2}\right]$

Median in continuous series

① $\left(\frac{N}{2}\right)^{th}$ term = Find Above or equal value in CF for $\left(\frac{N}{2}\right)^{th}$ term

② After find the value of CF the class in which CF lies that is Median class

$$\boxed{F = L + \frac{\left(\frac{N}{2} - CF_{-1}\right)}{F} \times h}$$

- L = Median class ki LCB
- CF_{-1} = CF se phle vali CF ki value
- h = Height of class.

Properties of Median

① $Y = a + bX$ (Linear relationship)

Put Median value or median

$$\boxed{\text{Median } Y = a + b(\text{Median } X)}$$

② Sum of Absolute deviation (Ignore − sign) & sum of All Absolute deviation is taken from Median is least or minimum.

Merits

① Rigidly defined

② Easily calculate & Simple to understand.

③ Can be computed with open-end classes.

④ Being a Positional Avg., Not much Affected by extreme values

⑤ dealing with Qualitative data.

⑥ It can be located sometimes by Inspection or by ogives (Less than / More than) graphically.

Demerits

② Can't determine exact value while ungrpd. dist. consisting even no. of observation

③ It is much Affected by Sampling Fluctuation.

④ It is necessary to Arrange observation of Data in order either Ascending or Descending.

(5) It is most Appropriate Avg. to be used while dealing with Qualitative data.

(6) It can be located sometimes by Inspection or by ogives (Less than / More than) graphically.

* Demerits

(1) Being a positional Avg., Not based on All observation.

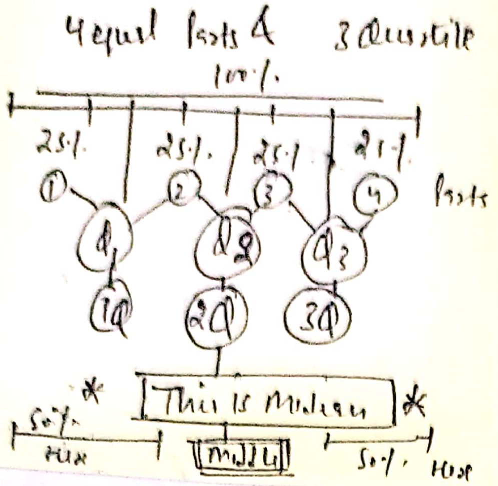
(2) Can't determine exact value while ungrpd. dist. consisting even no. of observation

(3) It is much Affecting Sampling ~~Fluctuation~~.

(4) It is necessary to Arrange observation of Data in order either Ascend-ing or Descending.

Quartile

4 equal parts & 3 Quartile



* Calculation of Quartile
* In Individual observ.

$Q_K = \left(K\left(\frac{n+1}{4}\right)\right)^{th}$ Term

If $1Q = \left(1\left(\frac{n+1}{4}\right)\right)^{th}$ term

$2Q = \left(2\left(\frac{n+1}{4}\right)\right)^{th}$ Term

$3Q = \left(3\left(\frac{n+1}{4}\right)\right)^{th}$ term

* In Descrete Series *

$Q_K =$ CF value $> K\left(\frac{n+1}{4}\right)^{th}$ Term

Means CF ki vo value Jo $K\left(\frac{n+1}{4}\right)$ se badi Hai Jis variable ka or (X) ki Jo CF value Jis Hai vo X ki value Qk ki value Hai.

* In continuous Series *

$Q_K =$ CF value $> K\left(\frac{N}{4}\right)^{th}$ term

$$Q_K \rightarrow \boxed{LC + \frac{\frac{K(N)}{4} - CF_{-1}}{\text{Frequency}} \times h}$$

* Decile

- Individual Series $\rightarrow D_K = K\left(\frac{N+1}{10}\right)^{th}$ Term
- Discrete Series $\rightarrow$ $CF > K\left(\frac{N+1}{10}\right)^{th}$

Iske [same vali (X) ki value Ans.]

- Continuous

$$D_K = LC + \frac{\frac{K(N)}{10} - CF_{-1}}{F} \times h$$

* Percentile *

- Individual Series $\rightarrow P_K = K\left(\frac{N+1}{100}\right)$
- Discrete Series $\rightarrow$

value corresponding To (X)

$(X) = CF > K\left(\frac{N+1}{100}\right)$

- Continuous Series $= P_K = LC + \frac{\frac{K(N)}{100} - CF_{-1}}{F} \times h$

MODE

Inspection	Distribution
which freq. is High. that is Mode	Jisme 2 Mode Hote Hai, ek hi Individual Series me.

Multimodal distribution

2 से Jyada Mode Hote hai ek hi individual Series.

* Mode in descrete Series *

value of (X) Jiski Freq. Sabse Jyada Ho vo Mode Hoti Hai (X ki value)

$\llcorner$ This is Mode.

* Mode

Individual Series

By Inspection

which freq. is High. that is Mode

Bimodal Distribution

Jisme 2 Mode Hote Hai. ek hi Individual Series Me.

Multimodal distribution

2 से Jyada Mode Hote hai ek hi individual Series.

MODE

* **No Mode** → Jab koi mode nhi hota. ek individual series me.

* **Mode in descrete series** *

value of (x) Jiski Freq. Sabse Jyada Ho vo Mode Hoti Hai (X ki value)
This is mode.

Mode for continuous series

$$\text{Mode} = LC + \frac{(f_1 - f_0)}{2f_1 - f_2 - f_0} \times h$$

i.e.

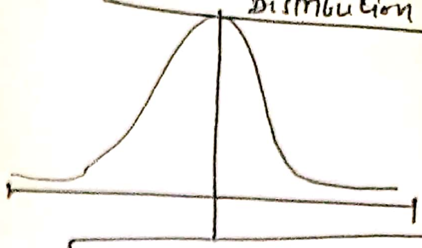
5-10 —	9	f_0
10-15 —	15 →	f_1
15-20 —	11	f_2

Assumptions of this formula

(1) The freq. dis. must be continuous with exclusive class type without any gaps.

(2) The class interval must be same. size of All C.I. same.

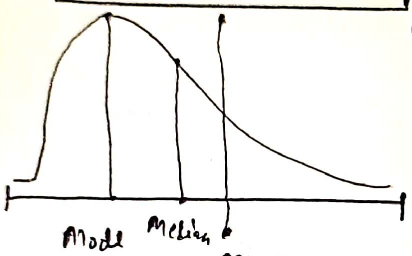
Symmetrical Distribution



Mean = Mode = Median

* Skewed distribution *

(1)



(1) Skewness Toward right

(2) Positive skewness

(3) Mean > median > mode.

(2)

(1) Skewness Toward left

(2) Negative skewness

(3) Mean < median < Mode.

* Properties of Mode *

(1) For grp. freq. dis. we consider emperical formula b/w Mean Mode Median.

empirical formula = (1) Mean − Mode = 3(Mean − Median)

Or (2) 3 Median = 2 Mean + Mode

* It hold for basically Moderately skewed Distribution.

(2) If $y = a + bx$ (linear relationship) than

Mode y = a + b(mode x)

no Mode).

* In some case its only located by inspection.

* can be determined graphically by Histogram.

* It is not Affect by extreme value of observation & can be calculated when extreme value not known

* Not based on all observation

* Not suitable for further Mathematical observation.

* As compare to mean, Median, Mode is more Affect by sample fluctuation.

Merits

* Simple to understand & Easy to calculate.
* In some case its only located by inspection.
* can be determined graphically by Histogram.
* It is not Affect by extreme value of observation & can be calculated when extreme value not known

Demerits

* It is not rigid defined (Some time 2 Mode & Sometime no Mode).
* Not based on all observation
* Not suitable for Further Mathematical observation.
* As compare to mean, Median, Mode is more Affect by sample fluctuation.

* Geometric mean *

* Product of set of observation & their nth root of their nth product

i.e. $\left[(X_1) \times (X_2) \times (X_3) \cdots (X_n)\right]^{\frac{1}{n}}$

* [nth root by calculator Trick]

* This is for Individual Series

* Discrete Series *

use freq. as power of (X) $\frac{1}{n}$

i.e. $\left[(X_1^{f_1}) \times (X_2^{f_2}) \cdots (X_n^{f_n})\right]$

* continuous Series

Same, but (X) is here mid value of class-Intervals

Uses

- For which ratio of consecutive terms remain constant Approximately.
  ex → Rates of changes, %, in sales, population sizes over diff. Time Period.

* Most Appropriate Avg. to be used in construction of Index No.

* It is suitable Avg. when it is desired to give more

Properties.

* If All observation are constant like variable (K) the G.M. is Also (K).

* if $Z = xy$

$\boxed{GM \text{ of } Z = GM \text{ of } x \times GM \text{ of } y}$

* if $Z = \frac{x}{y}$

than $GM \text{ of } Z = \frac{GM \text{ of } x}{GM \text{ of } y}$

* Harmonic mean *

$= \frac{\text{Total observation or } (n)}{\text{Sum of reciprocal of All observation}}$

i.e $\frac{n}{\frac{1}{X_1} + \frac{1}{X_2} \cdots \frac{1}{X_n}}$

Individual series.

number is constant.

* Discrete Series. * Total of freq.

$= \frac{\Sigma f_i}{\Sigma \frac{f_i}{x_i}} = \frac{[f_1 + f_2 \cdots f_n]}{\left[\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots \frac{f_n}{x_n}\right]}$

* continuous Series → Same / But here (x_i) is mid value of class.

... HM is Also (K).

(2) Combine HM $= \frac{n_1 + n_2}{\frac{n_1}{HM_1} + \frac{n_2}{HM_2}}$

Use

1. Avg. of speed.
2. when the ratio of the Qty. are given in which number is constant.

* Properties *

1. All variable are taken from or All obser. are (k) then H.M. is Also (k).
2. $\boxed{\text{Combine HM}} = \dfrac{n+n_2}{\dfrac{n_1}{HM_1}+\dfrac{n_2}{HM_2}}$

Weighted Mean

A.M. $\frac{\sum w_i x_i}{\sum w_i}$

G.M. $\left[x_1^{w_1} \times x_2^{w_2} \cdots x_n^{w_n}\right]^{\frac{1}{\sum w_i}}$

H.M. $\frac{\sum w_i}{\sum \frac{w_i}{x_i}} = \left[\frac{w_1 + w_2 \cdots w_n}{\frac{w_1}{x_1} + \frac{w_2}{x_2} \cdots \frac{w_n}{x_n}}\right]$

* Sum of Natural no. $(1+2 \cdots n)$ $(n) = \frac{n(n+1)}{2}$

* Sum of sq. of Natural no $(1^2+2^2 \cdots n^2)$ $(n^2) = \frac{n(n+1)(2n+1)}{6}$

* Sum of cube of the observation $(1^3+2^3 \cdots n^3)$ $(n^3) = \left(\frac{n(n+1)}{2}\right)^2$

* Relation b/w AM | GM | HM *

* If All observation are same than $\boxed{AM = GM = HM}$.

* If All data are not equal than $\boxed{AM > GM > HM}$

* For 2 no. or 2 observation [a, b]

AM $= \frac{a+b}{2}$ | GM $(a \times b)^{\frac{1}{2}}$

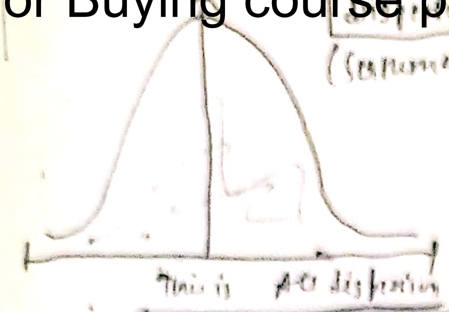
HM $= \frac{2}{\frac{1}{a}+\frac{1}{b}}$

$= \boxed{GM^2 = AM \cdot HM}$ equation Imp.

If one value in observation is 0 than GM is Always 0 & H.M can't calculate & Not defined.

If Any observation is −ve than can't calculate GM.

-Measure of Dispersion



D I S P E R S I O N

Absolute	Relative
Range	coff. of Range
M.D.	coff of m.D.
S.Deviation	coff of variation
	Deviation

Range = max - min

or

When Range in freq. dis: High C.B - L.C.B

or Largest value - Smallest value *

* coff. of Range

$$\frac{L-S}{L+S} \times 100$$

* Properties of Range *

① if $y = a + bx$ (linear relationship)

than $Range_y = Range_x$

Forex: $2x + 5y = 10$ — Ignore these values or signs also

$Range_x = 5$

$Range_y = \frac{2 \times 8}{8} = \boxed{2}$

If $3x + 2y + 10 = 0$ then x & y Range

$\boxed{3R_x = 2R_y}$ Ans.

* Ideal Measure Should Be

① Properly defined.

② Easy to comprehend.

③ Simple to compute

④ Based on all observations

⑤ Not affected by sample fluctuation. & desirable for further Mathematical Treatment

* Merits of Range *

① Simple to understand & easy to calculate.

② It require min. time to calculate value of Range.

Demerits

- Not Based on all observation
- Range is poor measure of dispersion. only consider extreme value.
- More Affected by sample fluctuations.
- cant calculate with open end series.
- Not suitable for further Mathematical Treatment

Individual

$$= \frac{\sum |x_i - A|}{n}$$ (A = Mean / Median)

* **Discrete Series** *

$$\frac{\sum f_i |x_i - A|}{\sum f_i}$$ (A = Mean / median)

$\sum f_i$

* **coff. of Mean Deviation** →

① $\bar{X} = \frac{M.D.}{\bar{X}} \times 100$

② Median $= \frac{M.D.}{Median} \times 100$

* **Mean deviation**

Mode is not use in M.D.

Absolute Avg. of deviation from (~~For~~ Median use)

$$= \frac{\sum |x_i - A|}{n}$$

either Mean

Median

Individual

* **Discrete Series:** *

$$\frac{\sum f_i |x_i - A|}{\sum f_i}$$

Mean

median

* **Continuous Series** *

Here X_i = Class value. (Mid value)

$$\frac{\sum f_i |X_i - A|}{\sum f_i}$$

Mean / median

* **coff of Mean Deviation** →

(1) $\bar{X}$ = $\frac{M.D.}{\bar{X}} \times 100$

(2) Median = $\frac{M.D.}{Median} \times 100$

(2)

* Properties of M.D.

- M.D. takes its min value when deviation are Taken from median.
- if $y = a + bx$

~~and of x = y~~

$\boxed{M.D \text{ of } y = b \times M.D. \text{ of } x}$

Ex> $3x + 4y + 7 = 0$ — ignore this.

M.D x = 6.40

than M.D of y = $\frac{3(6.40)}{4}$

$= \boxed{4.8}$

Merits of M.D.

- easy to understand & Simple to calculate.
- It is based on each & every item of data.
- It is rigidly defined.
- less Affected by extreme observation

Demerits

- Algebric signs (- or +) are ignored.
- Not suitable for Further Mathematical elements treatment
- can't calculated with open end class distribution.

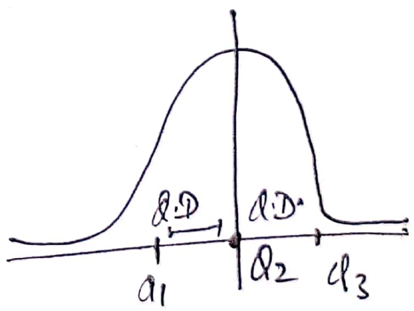
* **Quartile Deviation** → **Symmetrical Distribution**

* ① **Inter Quartile Range.**

$\boxed{Q_3 - Q_1}$

② **Semi - Inter Quartile range**

$\boxed{\frac{Q_3 - Q_1}{2}}$



① $Q.D = Q_2 - Q_1$

② $Q.D = Q_3 - Q_2$ (Subtract)

$Q_2 - Q_3 + Q_2 - Q_1$

$\boxed{2Q_2 = Q_1 + Q_3}$

Imp formula.

$\boxed{Q_2 = \frac{Q_1 + Q_3}{2}}$

* **coff. of Quartile Deviation**

$\boxed{\frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100}$

or

$\frac{Q_3 - Q_1/2}{Q_3 + Q_1/2} \times 100 \rightarrow \frac{Q.D.}{\text{Median or } Q_2} \times 100$

① $y = a + bx$

Ignore this Always.

$$\boxed{Q.D \text{ of } y = b(\text{absolute Dev. of } x)}$$

Demerits of Q.Dev

- Not base for All observation
- Not suitable for further Mathematical Treatment
- It is Affected by considerable Sample fluctuation.

Merits of Q.Dev.

- Can be calculate with open end series.
- Less affected by sample fluctuation
- Desirable to study Half of underhalf part of data.

Standard Deviation = σ (Sigma)

Always = +ve

root mean sq. deviation.

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

or

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

Individual series

Discrete Series

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

or

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

Same in continuous series

These x_i = mid value.

Variance

$$\sigma^2 = \frac{(x_i - \bar{x})^2}{n} \quad \text{or} \quad \frac{(\Sigma x_i)^2}{n} - (\bar{x})^2$$

Denote

Individual

$$\sigma^2 = \frac{\Sigma f_i x_i^2}{n} - (\bar{x})^2 \quad \text{or} \quad \sigma^2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i}$$

Discrete Series

Same in continuous

Correcting ... S.D.

Steps

1. find $\boxed{\text{correct mean}}$
2. find wrong $(\Sigma x_i)^2$ by $\sigma_w = \sqrt{\frac{\Sigma (x_i)^2}{n} - (\bar{x})^2}$
3. find correct $(\Sigma x_i)^2$ by subtract wrong $(x_i)^2$ & + correct$(x_i)^2$
4. find correct S.D or σ

Combined S.D.

$$\boxed{\bar{X} = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2}} \to \text{combine mean}$$

$$\sigma = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$

Deviation of mean.

$$\boxed{d_1 = \bar{X}_1 - \bar{X}}$$

$$\boxed{d_2 = \bar{X}_2 - \bar{X}}$$

Coff. of variation.

$$= \boxed{\frac{\sigma}{\bar{X}} \times 100}$$

If more uniform / more consistent → value of variation is **Less**

If less " / less " → ———→ is **More**

* **Properties of S.Dev.**

- Less further math. Treatment
  bcz of combine S.Dev.
- $y = a + bx$ — Ignore this

$$\boxed{\text{SD of } y = |b| \text{ Sd. of } x}$$

- If value of constant or (K) then
  S.D, M.D, Q.D
  = 0

S.D. of first (n) natural no. is

$$\boxed{\sigma = \sqrt{\frac{(n)^2 - 1}{12}}}$$

Relation b/w SD, MD, QD

4SD = 5MD = 6QD

Applicable only in symmetrical distribution

Probability

→ Probability, chance, odds in favour, odds against are too familiar now a days. & they have their origin branch of Mathematics known as Probability.

- Event = Result (1) coin Toss ⇒ Possible event

Head	Tail
1/2	1/2

* Some Terms *

- **Experiment** → A process with result (we know result definitely happen)

- **Random Experiment** → A process in which result is not known in Advance

- **Event** →
  - **Simple** → Ya Toh ye Hoga | ya fir ye Hoga → Single result
    - ex → Dice through → only one result.
  - **Composite** → Kuch Kam aise kie Jiska 2 possible Ans Ho sakte Hai Jaise 2 coin Toss & 2 Dice through.

- **Mutually Exclusive Events** → Kuch bhi common nhi Hai result mee.
  - ex → Single coin Toss
    Ans is either Head or Tail
    only 1 is possible Both are not possible.

- **Incompatible Events** → 1 se result Jyada Ho sakte Hai lekin order matter karega.

For Ex 2 coin Toss → Possible events

Coin 1	2
H	T
T	H
T	T
H	H

Here it is counted by 2 result order matter here.

In composite events Both are same (1 result)

- **Exhausted Event** → means सारे possible events

1 coin Toss either Head or Tail
Total event 2 — Head 1 + Tail 1 = 2

→ Dice through → All possible events, 1, 2, 3, 4, 5, 6
= no. of all possible event is 6

Events → (unbiased event) like → coin Toss

P of Head = $\frac{1}{2}$ — same

P of Tail = $\frac{1}{2}$

Dice through → All no. Have equal Probability to come = $\frac{1}{6}$

- For finite Elementary Events → (countable or simple events) → coin Toss, Dice Through

$$P(A) = \frac{\text{no. of equally likely events fav. to } A}{\text{Total no. of equally likely events}}$$

Ex: Dice through → Possibility of even no. = 2, 4, 6 = $\frac{n(A)}{n} = \frac{3}{6} = \frac{1}{2}$

- For finite composite Events → $P(A) = \frac{mA}{m}$ = $\frac{\text{no. of mutually Exc, Exh, equally likely fav. to } A}{\text{Total no.}}$

like → 2 Dice through, 2 coin Toss. Here no. of possibility is more.

- Imp. Points

1. Probability of events lies b/w 0 & 1 & both inclusive means = $\boxed{0 \le P \le 1}$; 0 = impossible events in A, 1 = Sure events in B forese

2. Non-occurrence of event denote by A' or A^c complimentary event of A mutually Exclusive & Exhaustive events.

→ $\boxed{P(A) + P(A') = 1}$

Hone ki P + Na Hone ki (P) = 1

3. = $\frac{\text{No. of fav. events}}{\text{no. of unfavourable events}}$ → Odds in Favour ①

$\frac{\text{No. of unfavourable events}}{\text{No. of fav. events}}$ → Odds in Against ②

$$\boxed{P(A) = \frac{n(A)}{n}}$$

wages	No. of workers
50-60	15
60-70	23
70-80	36
80-90	42
90-100	17
100-110	12

1) Probability less than 70/-wages $= \frac{38}{145}$

$$\boxed{P(A) = \frac{38}{145}}$$ etc.

2) **Set theoretical Approach**

Universal set (U)

↳ Sample space (S) or (Ω)

Then $P(A) = \frac{n(A)}{n(S)} = \frac{\text{No. of fav. count in event } A}{\text{Total no. in sample space.}}$

Imp. Points

- If 2 events are Mutually Exclusive, if $P(A \cap B) = 0$

then $n(A \cup B) = n(A) + n(B)$

$\therefore \boxed{P(A \cup B) = P(A) + P(B)}$ (1)

$= \frac{n(A \cup B)}{n(S)} = \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)}$

- If 3 events are Mutually Exclusive, if $P(A \cap B) = 0$, $P(A \cap C) = 0$,

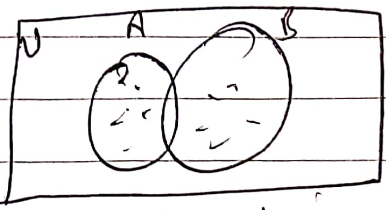
$P(B \cap C) = 0$, $P(A \cap B \cap C) = 0$

then $\boxed{P(A \cup B \cup C) = P(A) + P(B) + P(C)}$ (2)

$= \frac{n(A \cup B \cup C)}{n(S)} = \frac{n(A) + n(B) + n(C)}{n(S)} = \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} + \frac{n(C)}{n(S)}$

- If 2 events are A & B are **Exhaustive** if

$P(A \cup B) = \frac{n(A \cup B)}{n(S)} =$ (1)



All elements in A or B

$$P(A\cup B\cup C) = \frac{n(A\cup B\cup C)}{n(S)} = ①$$

no. of All elements A, B, C

$n(A\cup B\cup C) = n(S)$

- If Three events A, B, C are equally likely events if (Sbke Hone ke chance barabar hai) same hai

$\therefore \boxed{P(A) = P(B) = P(C)}$

- If A & B are not mutually exclusive & $n(A\cap B) \neq 0$ than

$$= P(A\cup B) = \frac{n(A\cup B)}{n(S)} = \frac{n(A)+n(B)-n(A\cap B)}{n(S)}$$

$$= \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} - \frac{n(A\cap B)}{n(S)}$$

$$= \boxed{P(A\cup B) = P(A) + P(B) - P(A\cap B)} \quad ③$$

- If 3 events A, B, C are not Mutually exclusive & $n(A\cap B\cap C) \neq 0$ than

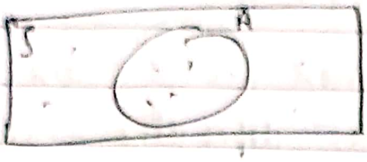
$$P(A\cup B\cup C) = \frac{n(A\cup B\cup C)}{n(S)} = \frac{n(A)+n(B)+n(C)+n(A\cap B\cap C) - n(A\cap B) - n(A\cap C) - n(B\cap C)}{n(S)}$$

④

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

(Extension of set Approach)

1. P(A) is Always lie b/w 0 & 1 for every $\boxed{A \subset S}$



$\boxed{A \text{ is subset of } S}$

* if $n(A) = 0$, then $P(A)$ is Also 0
* $P(A) = 1$, when $n(A) = n(S)$

2. $\boxed{P(S) = \frac{n(S)}{n(S)}}$ = 1 Probability of sample space is Always 1.

3. For Any sequence of $\boxed{\text{Mutually exclusive}}$ events A1, A2, A3 $\boxed{\text{Probability}}$ will

$(P) = P(A_1) + P(A_2) + P(A_3)$ - - - $\boxed{\text{So on}}$

* Individually (P) nikalke sum kar denge.

---

o $\boxed{\text{Addition Theorems (or) Theorm on Total Probability}}$ *

1. * Meaning of +, or is union (∪)

* Meaning of ∩, and, × is Intersection (∩)

→ $P(A \cup B)$ or $P(A+B) = P(A) + P(B)$ [For Any 2 Mutually Exc. Events.]

o If nothing mention in Ques - Assume $\boxed{\text{not Mutually Exclusive}}$.

2. Mutually Exclusive events A_1, A_2, A_3 - - A_K

$P(A_1 \cup A_2 \cup A_3 - - - \cup A_K) = P(A_1) + P(A_2) + P(A_3) - - P(A_K)$

3. For Any 2 events A & B the (P) either A or B occurs is given by

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

4. For Any 3 events A, B, C the (P) that Atleast one of the events occurs

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

- Compound probability

→ (P) of occuring 2 events A & B simultaneously called compound Probability.

= P(A & B) = $\boxed{P(A \cap B)}$

- $\boxed{\text{Dependent events}}$ → If 2 or more event are there & peforming a event is depending on the other.

- $\boxed{\text{Independent event}}$ → If 2 or more event are happening & there is no dependency of event on each other. ~~(Mutually exclusive event)~~ ~~(one of them only occurs)~~

→ $\boxed{\text{Compound (P) of dependent events}}$

If B depend on A

= ① $P\left(\frac{B}{A}\right) = \frac{P(A \cap B)}{P(A)} = \boxed{P(A) > 0}$ $\boxed{\text{not an Impossible event}}$

② $P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}$

(Also $P(B) > 0$, not an Impossible event

* $\boxed{\text{Demorgan's law}}$

Rule ① $\frac{P(A' \cap B')}{P(B')} = \boxed{P(A \cup B)'}$

$\hookrightarrow = \frac{\boxed{P(S) - P(A \cup B)}}{P(B')}$

Rule ② $\frac{P(A' \cup B')}{P(B')} = \frac{P(A \cap B)'}{P(B')} = \frac{P(S) - P(A \cap B)}{P(B')}$

- Compound Probability of Independent Event [Independent]

① $P\left(\frac{B}{A}\right) = P(B)$ | ① $P\left(\frac{A}{B}\right) = P(A)$

~~Formula~~ $\boxed{P(A \cap B) = P(A) \cdot P(B)}$

→ If A & B are independent

∴ ① $P(A \cap B) = P(A) \cdot P(B)$

② $P(A \cap B') = P(A) \cdot P(B')$

③ $P(A' \cap B) = P(A') \cdot P(B)$

④ $P(A' \cap B') = P(A') \cdot P(B')$

① $P(A \cap B') = \boxed{P(A-B)}$

② $P(A' \cap B) = \boxed{P(B-A)}$

- When dependence is not given Assume Independent event.

- Compound (P) of Three events (Independent)

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

- Compound (P) of Three events (dependent)

A, B, C

A = Independent , B dependent on A , C dependent on A & B

$$P(A \cap B \cap C) = P(A) \times \boxed{P\left(\frac{B}{A}\right)} \times \boxed{P\left(\frac{C}{A \cap B}\right)}$$

$$P\left(\frac{B}{A}\right) = \frac{P(A \cap B)}{P(A)} \qquad P\left(\frac{C}{A \cap B}\right) = \frac{P(A \cap B \cap C)}{P(A \cap B)}$$

* Random variable Probability distn

- X = Random variable

Experiment → Sample space → Function (X) = Ans

(Sample set) event

Roll no

Ex → 3 coin Tossed X = Denote no. of Head

Date
Page

coin ①	②	③	No of Head
H	H	H	3
H	H	T	2
H	T	T	1
T	T	T	0
T	T	H	1
T	H	H	2
T	H	T	1
H	T	H	2

X	0	1	2	3
n	1	3	3	1
$P(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Total $\frac{8}{8} = 1$

- If random variable (X) Assume (n) finite values $X_1, X_2, X_3 \ldots X_n$ with corresponding (P) = $P_1, P_2, P_3 \ldots P_n$
  1. $P_i \geq 0$ (For every i) $\boxed{0 \leq P_i \leq 1}$
  2. $\Sigma P_i = 1$ Always.

Then (P) distribution =

$$\boxed{P(X) = P_1 + P_2 + P_3 - - - P(X_n)}$$

* Discrete Random Variable → Fix value like → no. of shoe Lace, no. of Accident etc

↳ Probability mass Function

* Continuous Random Variable → In Intervals like → Height, weight etc

↳ Probability density function → ① $F(x) \geq 0$ for $x \in (\alpha, \beta)$

② $\int_{\alpha}^{\beta} F(x)\,dx = 1$

- If random variable (X) assumes (n) values $x_1, x_2, x_3 \ldots x_n$ corresponding with probability $p_1, p_2, p_3 \ldots p_n$ which (1) satisfy (i) $p_i \geq 0$, (2) $\sum p_i = 1$

* then Expected value (u) or $E(x) = \frac{\sum p_i x_i}{\sum p_i} = \sum p_i x_i = \boxed{\sum p_i x_i}$

- Expected value of x^2 given by $E(x^2) = \boxed{\sum p_i x_i^2}$

$E_x = u$

(1) * Variance of (X) → $\boxed{V(x) \text{ or } \sigma^2}$

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

where $E(x^2) = \sum p_i x_i^2$

$E(x) = \sum p_i x_i$

(2) S. Deviation of (X) $\sigma = \sqrt{V(x)}$

$\therefore = \sqrt{E(x^2) - (E(x))^2}$

- Positive sq. root of variance known as $\boxed{\text{S. Deviation}}$

* In particular expected value monotonic fun^n $g(x)$ is given by =

$E(x) = \sum p_i x_i$ $E(x^2) = \sum p_i x_i^2$

$E(g(x)) = \boxed{\sum p_i (g(x_i))}$

* $\boxed{\text{Properties}}$ of $\boxed{\text{Probability dist}^n \text{ Random variable}}$

(1) Mean of (u) of y = (a) + (b) X (mean (u) of x)

$\boxed{\text{Constant}}$ $\boxed{\text{constant}}$

$\boxed{y = a + bx}$

(2) S. Deviation of y = $|b| \cdot$ S.deviation of x

Expected value of Random variable

(1) when variable (x) is $\boxed{\text{discrete}}$ & $\boxed{\text{Probability Mass Fun}^n}$ $= f(x)$ then $E(x)$ i.e μ or Mean is given by

(1) $E(x) = \mu = \boxed{\sum X \cdot f(x)}$ $\boxed{P_i \text{ is replace by } f(x)}$

(2) $E(x^2) = \sum p_i x_i^2 = \boxed{\sum x_i^2 \, (f(x))}$

(3) variance is given by $= E(x^2) - (E(x))^2$

or $\boxed{\sigma_y^2 = |b|^2 \cdot \sigma_x^2}$ $= \boxed{\sum x_i^2 \cdot (f(x))} - \boxed{(\sum X \cdot f(x_i))^2}$

(4) S.D $= \boxed{\sigma = \sqrt{V(x)}}$

(2) $\boxed{\text{In continuous variable}}$ (Probability density function)

* $\mu = E(x) = \int x \cdot f(x)\, dx$

* $E(x^2) = \int x^2 \cdot f(x) \, dx$

* $\sigma^2 = \int x^2 \cdot F(x)\, dx - \left(\int x \cdot f(x)\, dx\right)^2$

$\boxed{\text{Properties of Expected values}}$

(1) Expectation of all value equal to constant (K) is

$X = K, K, K$

$E(K) = \boxed{K}$ $\boxed{\sigma^2 = 0 \quad , \quad \sigma = 0}$ (If all value are same or equal)

(2) Expectation of sum of (2) Random variable is the sum of their expectations →

(ex) $E(x) = 8$, $E(y) = 9$

$\boxed{E(x+y)} = \boxed{E(x) + E(y)}$

$8 + 9 = \boxed{17}$

Expectation of Product of constant & random variable

$= E(Kx) = K(E(x))$ (constant)

$E(x) = 5 \quad E(3x) = ? = 3(E(x)) = 3 \times 5 = 15$ Ans

Expected value of Product of 2 random variable.

$$\boxed{E(xy) = E(x) \cdot E(y)}$$

Theoretical Distribution

- Probability distn → o Discrete random variable & continuous random variable followed.

- o Note → Properties of S.Deviation & Mean (μ) same will followed in Theoretical distn

- As Probability exist only in Theory, that's why it is Theoretical distn

x	x_1	x_2	x_3	— —	x_n
$P(x)$	$P(x_1)$	$P(x_2)$	$P(x_3)$	— —	$P(x_n)$

→ Use case of the distn

1. Life span of electrical item
2. Future Projections for future
3. Hypothesis Test for Decision making on sample.

Discrete Random variable
1. Binomial Distn
2. Poison distn

Continuous Random Variable
3. Normal Distn

Binomial Distn derived from

* Bernoulli Process named After Famous Mathematician Bernoulli

* When Trial (Random exp.) Attempt to produce particular Outcome — (Events) — Success → $\boxed{P}$ = (denotation)

Failure → $\boxed{Q}$ = $\boxed{1-P}$ (denotation)

o neither certain i.e not 100% sure = P(F)

* $\boxed{0 < P < 1}$ *

o nor uncertain (100% Impossible bhi ni hai PFO)

1. Trial is Associated with 2 mutually Exclusive & exhaustive Elements.
2. The occurrence which known as (success)(P) & non-occurrence (failure) (q) or $\boxed{1-P}$
3. (n) no of random exp. or trial (n) should be finite, +ive Integers

$\boxed{P \text{ \& } q \text{ both are } > 0}$

$\boxed{\text{Formula}}$ → $f(x), P(x) = \boxed{{}^{n}C_{x} \cdot P^{x} \cdot q^{n-x}}$

- Here (P) → is for one or single event/experiment
  q → is for single event/experiment.
  by this formula we get Probability with Total no. of events or experiment

4. $\boxed{(X) = \text{fav. no. of success}}$

- $\boxed{\text{Imp. Points of Binomial dist}^{n}}$ ·

→ 0 is Possible when (n, P) are given these (n, P) are Also called $\boxed{\text{Parameters}}$ of Binomial dist^{n}.
If n, P not given we can't calculate.
- n of exp. should be fixed.

1. As $n > 0$ than $f(x) \geq 0$ for every (x).
   $P, q > 0$, $0 \leq x \leq n$
2. Also $\Sigma f(x) = 1$

- If $x \not< n$ = $x \not> 0$ = $\boxed{P(x) = 0}$

ex = $n = 10$
$X = 11$ Heads
$\boxed{\text{Not Possible}}$ → $= \boxed{P(x) = 0}$

3. Binomial dist^{n} known as $\boxed{\text{Bi-Paramatric dist}^{n}}$ As it is characterised by 2 Parameters $\boxed{n \text{ \& } P}$
4. The Mean of Binomial dist^{n} = $E(x)$ = $\boxed{n \cdot P}$
5. variance of " " = σ^2 = nPq
6. S. Dev of " " $\sigma = \sqrt{V(x)}$
   $= \sqrt{nPq}$

Variance (σ^2) < ($n \cdot p$) or Mean

Variance is Always less than its Mean

$\boxed{\sigma^2 < \mu}$ b/c $0 < p < 1$

$0 < q < 1$

npq np

$\boxed{q < 1}$

(7) Mode in Binomial $dist^n$ → (1) for non-Integers

$\boxed{(n+1)p}$

ex → $(4+1) \cdot 25 = 1.25$

$\boxed{\text{Greatest Integers}}$ → (1)

under 1.25 is

(2) For Integers — If $(N+1)p$ = Integer

than same value Also

$[(N+1)p] - 1$

Ex = $(N+1)p = 5$

$[(N+1)p] - 1 = 5 - 1 = 4$

$\boxed{4, 5} = \boxed{2 \text{ Modes}}$

(8) Additive Property in Binomial $dist^n$ → If x & y are 2 independent variable Such that $X \sim B(N_1, p)$, $Y \sim B(n_2, p)$

than $P(x+y) = {}^{n_1+n_2}C_{x+y} \times p^{x+y} \times q^{(n_1+n_2)-(x+y)}$

$\boxed{X + Y \sim B(N_1 + N_2, p)}$

(2) Poison distⁿ — when $n \geq 50$ & $p < 0.05$ — then we estimate Poison distⁿ

→ The probability of finding success in very small & constant
→ Having more than (1) success in this time interval is very low
→ The (P) of success in this time interval is independent of time & as well as earlier

- The (P) of (x) success in a time period given the length of the period & Avg. events per time =

$$\boxed{P(x) = \frac{e^{-m} \cdot m^{x}}{x!}}$$

$\boxed{e = 2.71828}$ * exponent *

$\boxed{m = n \cdot P}$

n — no. of trials
P — (P) of success probability

Poison — $n \geq 50$ & $P \leq 0.05$

Time interval — $P \leq 0.05$ — mee Avg. ye (P) hai

Repetition of success is too low & independent on previous success

- Poison distⁿ is limiting form of Binomial distⁿ where $n \geq 50$, $P < 0.05$ or (P is near to 0) such that $n \times P = m$ a positive or finite the binomial distⁿ / reduces to the Poison with (P) funⁿ

$$f(x) = P(X = x) = \frac{e^{-m} \cdot m^{x}}{x!} \qquad m = n \times P$$

(2) Important Points of Poison distⁿ

classmate

Date
Page

1. $F(0) + F(1) + F(2) + \dots = 1$ $n \geq 100$

2. Poison distⁿ is known as uni-parametric distⁿ because as it characterised by one Parameter (m) = $\boxed{n \times P}$

$\boxed{m \to \text{Poison variate}}$
$(n \cdot P)$

3. The mean of Poison distⁿ given by m i.e $(n \times P)$

5. The S-Deviation $\boxed{\sigma = \sqrt{m}}$

4. The variance of Poison distⁿ $\boxed{\sigma^2 = m}$

P is next to 0

So $q = 1 - P = 1$

$\sigma^2 = nPQ = nP(1) = \boxed{nP = m}$

5. Like Biomodel distⁿ, Poison distⁿ Also Bimodel or unimodel depending on value of Parameter (m)

1. m = Greatestest value in m $\boxed{\text{if } m = \text{Non integer}}$

like → 2.5 = 2 is mode,

2. $\boxed{m-1}$ if m = Integer $= 3 - 1 = 2$ $\boxed{\text{Mode}}$
& 3

7. $\boxed{\text{Additive Property}}$ → $X \sim P(m_1)$, $Y \sim P(m_2)$ $\boxed{X + Y \sim P(m_1 + m_2)}$

* Application of Poison distⁿ *

→ we can Apply Poison distⁿ in Following cases →

1. The distⁿ of the $\boxed{\text{no. of Printing mistake Per Page}}$ of large book

$n \to \infty$ $P = 0$

2. The distⁿ of the no. of road Accident on busy road $\boxed{\text{Per minute}}$

→ Time interval

$\boxed{P < 0.005}$

3. $\boxed{\text{no. of Shoe lace Purchase}}$ Per min

Time interval

exe $P \to 0$

4. $\boxed{\text{no. of demand Per min}}$ for Health centre & so on

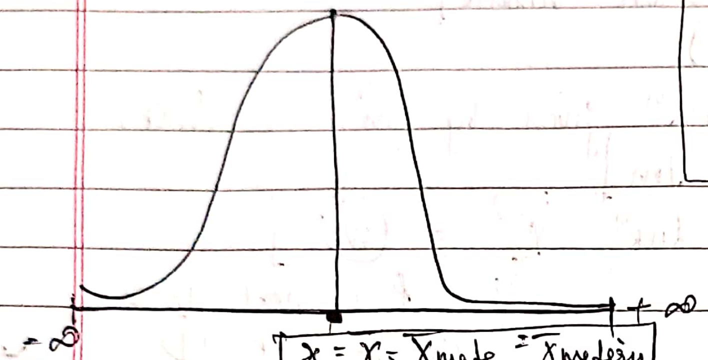
Time interval = $\boxed{P \to 0}$

(3) Normal Distⁿ | Gaussian distⁿ | Symmetric Distⁿ

Always symmetric

= Area under curve = 1

1. $f(x) > 0$
2. $\int_{-\infty}^{+\infty} F_x(dx) = 1 = P(-\infty < x < \infty)$



$x = Y = X_{mode} = X_{median}$

Total Area under curve = 1

$-\infty$ ... 0.5 ... 0.5 ... $+\infty$

$P(x < \bar{x})$ $\quad$ $P(x > \bar{x})$

* $\boxed{P(x < \bar{X})} = 0.5$

$= P\boxed{-\infty < x < \bar{x}} \to 0.5$

* $\boxed{P(x > \bar{x})} = 0.5$

$= P(\bar{X} < x < \infty) = 0.5$

→ Only in continuous variable → sign less than (<) Ho ya less than equal to Ho ($\leq$) koi Farak nhi Pdta.

ex $P(x \leq \bar{X}) = P(-\infty, \bar{X}]$

$= \int_{-\infty}^{\bar{X}} F(x)\, dx = \boxed{0.5}$

* Cumulative distⁿ Funⁿ → or. Cum. Probability Funⁿ or Cum. Prob. density Funⁿ

Means Less than

Mean (X) value less than this etc. or $\Phi(1)$ → this means $P(x<1)$
For ex. 1

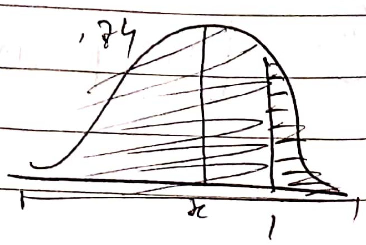
Here (u) or Mean = 0, S. Deviation = 1

* Biometrika Table *

* If $P(x>1) = 1 - P(x<1)$

given $\Phi(1) = .74$ $\quad$ $1 - .74 = \boxed{.26}$

$P(x<1) = 0.74$



$\boxed{\text{Normal Distn}}$ → A continuous random variable defined in terms of its Probability density funn $F(x)$ satisfy following condition :

① * $F(x) \geq 0$, $x \in (-\infty, \infty)$

$$\int_{-\infty}^{\infty} F(x)\,dx = 1$$

② A continuous random variable x is defined to follow normal distn with parameters $\boxed{\mu, \sigma^2}$

③ A density funn of normal distribution called $\boxed{\text{Normal density funn}}$ → (Probability density funn)

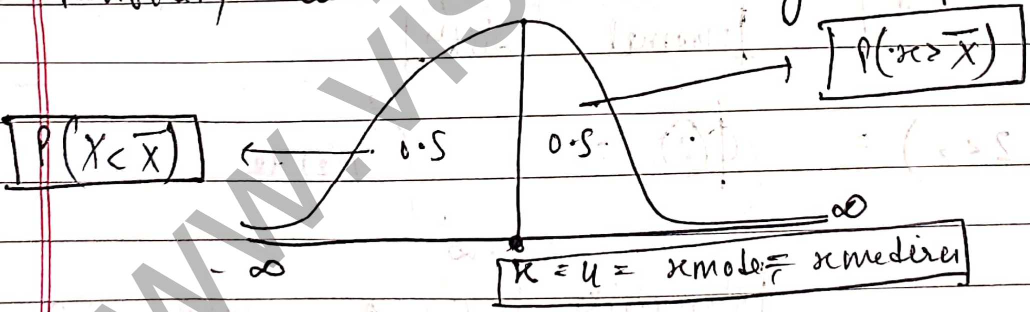
$$\boxed{F(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-(\mu)\text{mean})^2}{2(\sigma^2)\text{ variance}}}} \quad ① \qquad \boxed{-\infty < x < \infty}$$

or

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

$\boxed{\text{Some imp point}}$ of * $\boxed{\text{normal distn}}$ *

① If we plot the Probability funn $y = F(x)$ the curve is called Probability curve, with following shape



Area under $-\infty$ to $\mu = 0.5$ & μ to $\infty = 0.5$ } → Total = ①

Standard normal distribution → A random variable that has a normal dist^n with mean$(\mu) = 0$ & St. dev$(\sigma) = 1$ is said to have standard normal Probability dist^n.

→ So the density fun^n of St normal dist^n of (z) is given by $\mu = 0$, $\sigma = 1$. Here (x) is replaced by (z)

$$\boxed{F(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}} \qquad \boxed{-\infty < z < \infty}$$

(Compare with normal dist^n)

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

In St normal density fun^n

Mean $= 0$, $\sigma = 1$

So $\frac{x-0}{2(1)^2} = \frac{x}{2} = \frac{z^2}{2}$

& x is replaced by z

$\frac{1}{(1)\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}}$

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

* **Cumulative Frequency dist^n in Standard Normal dist^n**

Here $P(Z < z) = \Phi(z) = \int_{-\infty}^{z} F(z)\,dz$

Biometrika Table (where mean $= 0$, S.Dev $= 1$)

* **Properties of normal dist^n** *

① $\int_{-\infty}^{\infty} F(x)\,dx = 1$

② Mean $= (\mu)$, Median & Mode Also given by (μ) (In normal dist^n)

characterized by 2 parameters (μ & σ^2)

$X \sim N(\mu, \sigma^2)$

* 4 SD = 5 MD = 6 QD — In symmetrical distn

(4) * **Mean deviation** → MD = $\frac{4}{5}$ S.D

i.e = $\boxed{MD = 0.8 (S.D)}$

∘ Q. Deviation — Approximate relation:

$QD = \frac{4}{6} SD$

⇒ $\boxed{0.666\ SD}$

(5) * The First & Third Quartile is given by

Q_1: $QD = Q_2 - Q_1$

$\boxed{Q_1 = Q_2 - QD}$ (1)

Q_2 → median = mean = mode or (μ)

Q_3: $QD = Q_3 - Q_2$

(2) $\boxed{Q_3 = QD + Q_2}$

Q_2 → median = mean = mode or (μ)

(3)

(6) * **Quartile deviation** is $\boxed{0.675 (S.Dev.)}$

= $Q_1 = \mu - 0.675 (S.D)$ (4)

= $Q_3 = \mu + 0.675 (S.D)$ (5)

(7) The normal curve (μ) = $F(x)$ has 2 point of inflexion

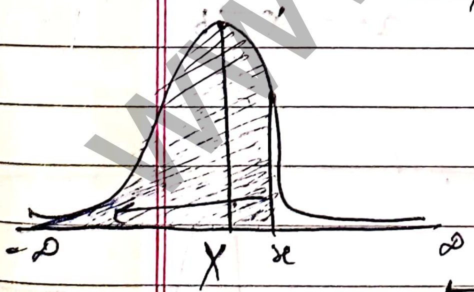
(1) $\boxed{x = \mu - \sigma}$ (2) $\boxed{x = \mu + \sigma}$

this is when going to $-\infty$ from μ

this is when going to ∞ form μ

$-\infty$ $\mu - \sigma$ μ $\mu + \sigma$ ∞

* **cumulative distn funn** *



$\boxed{P(-\infty < X < x)}$

funn converting to (Z) = $\mu = 0, \sigma = 1$

* Imp. value *

$\phi(0) = 0.5$

$\phi(1) = .8413$

$\phi(2) = .97725$

$P(-\infty < Z < z)$ or $P(Z < z)$ — $\int_{-\infty}^{z} F(z)\, dz$

(3) $\boxed{\phi(z)}$ — Biomedical Table

[Probability]

① $P(x<a)$ — [less than]

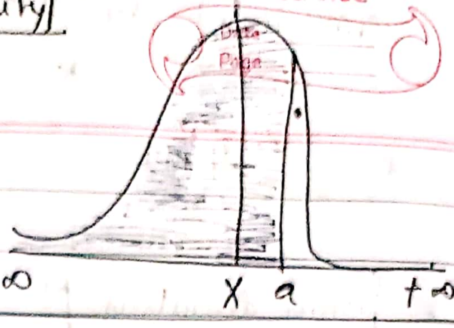
[convert in Z]

$$P\left(\frac{x-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right)$$

let this = Z — let this (K)

$= P(Z<K)$

$\phi(K)$



0 [1st subtract mean, than divide (S. deviation)]

② $\mu = 500$, $\sigma = 100$ (Prob of $x < 500$)

$$P(x<500) = P\left(\frac{x-500}{100} < \frac{500-500}{100}\right)$$

$P(Z<0)$

$\boxed{\phi(0) = 0.5}$

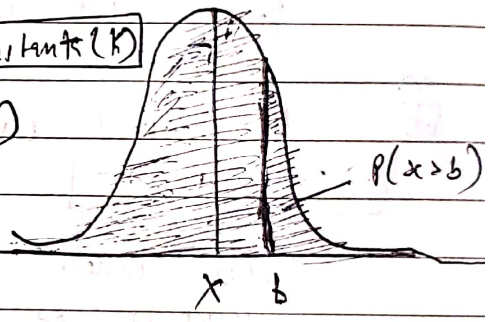
② $\boxed{P(x>b)}$ = (1st convert in less than form)

[greater than]

$1 - P(x<b)$ — [Constants (K)]

$1 - P(Z<b)$ or $1 - P(Z<K)$

$1 - \phi(b)$ or $(1 - \phi(K))$

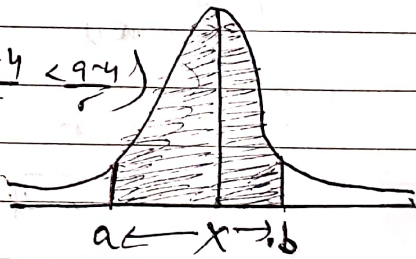


③ $P(a<x<b) = P(X<b) - P(x<a)$

$$P\left(\frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) - P\left(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right)$$

$P(Z<K_2) - P(Z<K_1)$

$\boxed{\phi(K_2) - \phi(K_1)}$ Ans

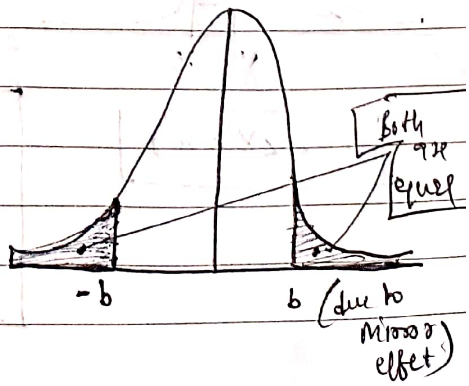


④ $P(x<-b) = P(x>b)$

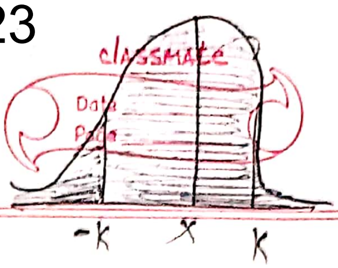
$1 - P(x<b)$

$1 - P(Z<K_1)$

$\boxed{1 - \phi(K)}$



1. $P(Z>K) = \boxed{1 - P(Z<K)}$
2. $P(Z<-K) = \boxed{P(Z>K)} = \boxed{1-P(Z<K)}$
3. $P(K_1<Z<K_2) = \boxed{P(Z<K_2) - P(Z<K_1)}$
4. $P(Z>-K) = \boxed{P(Z<K)}$



* **Properties of Normal distn** ⟷ **St. density Probability Fun F(Z)**

1. If X & Y are independent normal variable with mean (u) & S. Dev. (σ) than $(X+Y) = ?$

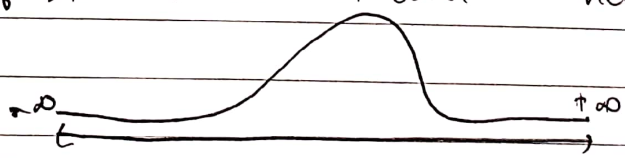
$X \sim N(u_1, \sigma_1)$, $Y \sim N(u_2, \sigma_2)$ $\Rightarrow$ $\boxed{(X+Y) \sim N\left(u_1+u_2, \sqrt{\sigma_1^2+\sigma_2^2}\right)}$

* **Some imp Property of St. normal variate**

(7) u, Median = mode)

* If we take $u=0$, $\sigma=1$, we have $\boxed{-\infty < Z < \infty}$

1. Z has Mean, Median, Mode all equal & 0.
2. σ of Z = 1
3. σ^2 or variance is Also 1
4. Also $\boxed{MD} = \boxed{0.8\,(SD)}$ & $\boxed{Q.D} = \boxed{0.675\,(SD)}$
5. The standard normal distn About (Z) = 0
6. Point of Inflexion of Standard normal distn are
   (1) $\boxed{u - \sigma = -1}$ (2) $\boxed{u+\sigma = 0+1 = 1}$
7. The Two Tails of standard normal curve never touch its Horizontal Axis.



8. (1) $P(Z>Z_P) = \boxed{P}$
   (2) $P(Z < Z_{1-P}) = P$ → due to Mirror effect
   (3) $P(Z<-Z_P) = \boxed{P}$
9. We Have $\boxed{Z_{\boxed{0.0005}} = 2.58}$ than $P(Z > 2.58)$ Ans is = $\boxed{0.0005}$

↳ To ये है यहां

$\boxed{\text{Further Ex}}$ $\quad Z_{0.001} = 2.33 \quad P(z > 2.33) = 0.001$

Ans.

Correlation and Regression

* Two variable are said to in corelation, if change in ① variable cause change in another.

Ex → Healthy food = Healthy person.

- Corelation means → Corelate B/w (X) & (Y) → 2 variable → Bivariate data.
- Grouped freq. distn

→ Bivariate data

- Marginal distn → Sirf ek hi variable ke data
  - Ex → (in class note book)
  - → only either data of (X) with its Frequency or (Y) with its Frequency
  - They are independent of other variable.
- Conditional Distn → Jab ek variable ke condition Ho or ~~dusre ke data~~ ~~dusre~~ ~~ke data~~ bhara Ho' (ex = in class note Book)
  - → (X) ki condition. ke (Y) ka distn or (vice. versa)

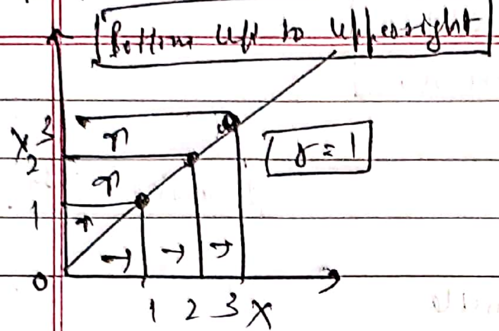
* Positive or Negative corelation
  - → Corelation is +ve when one variable ↑, other Also ↑.
  - → ——— -ve ——— ↑ ——— ↓
- Direct relation in +ve co-relation, Opposite or inverse relation -ve co-relation

* Method of Studying Corelation *

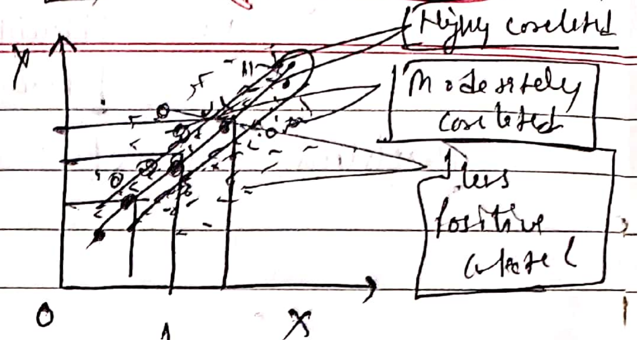
1. Scatter dig method
2. Karl Pearson's Method
3. Speaman Rank corelation
4. Coefficient of con-current deviations.

① Scatter diagram method

① Perfect corelation

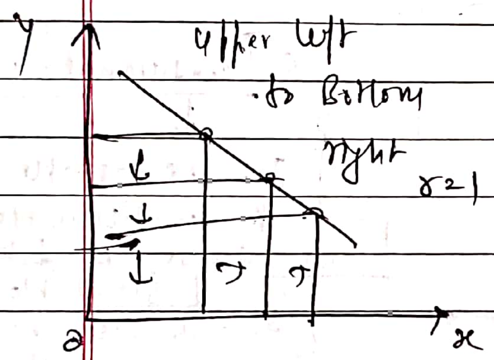


② Perfect +ve corelation



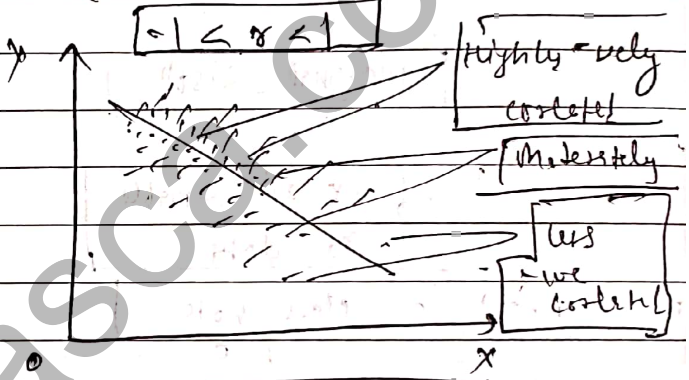
0 < r < 1

③ Perfect -ve corelation



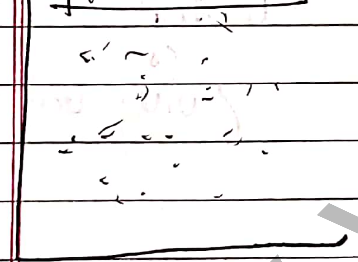
④ -ve corelation

-1 < r < 1



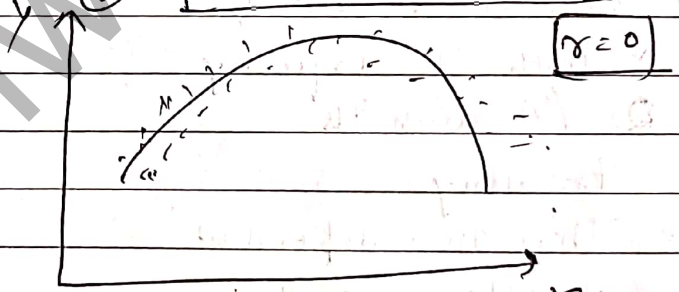
⑤ No relation

r = 0



⑥ curvilinear corelation

r = 0



* The corelation cofficient * (r) (denotation)

→ The quantitative measure of strength in the linear relationship b/w 2 variable is called (corelation cofficient)

* r = linear relationship *

→ Measure point cluster about the straight line.

→ ranges b/w $-1 < r < 1$ [same] $-1 < r < 1$

→ No relation than 0

→ More corelation differ from 0, the stronger the linear relationship b/w 2 variable.

Degree	+ve	-ve
No relation	0	0
very low	0 — .25	-.25 — 0
Low	.25 — .50	-.50 — -.25
Moderate	.50 — .75	-.75 — -.50
Fairly High	.75 — .90	-.90 — -.75
very High	.9 — 1	-1 — -.9
Perfect	1	-1

Covariance

→ Avg. of Product of variable x, y from their means $\bar{x}$, $\bar{y}$

= ① $Cov(X,Y) = \boxed{\dfrac{\Sigma (x_i - \bar{x}) \cdot \Sigma (y_i - \bar{y})}{n}}$

n = No. of Pairs of X & Y

② $cov(X,Y) = \boxed{\dfrac{\Sigma X \cdot Y}{n} - \bar{X} \cdot \bar{Y}}$

② **Karl Pearson's Method** → $r = $ ① $\boxed{\dfrac{Cov(X,Y)}{\sigma_x \cdot \sigma_y}}$

② $r = \boxed{\dfrac{\Sigma (X-\bar{X}) \cdot \Sigma (Y-\bar{Y})}{\sqrt{\Sigma (X-\bar{X})^2 \cdot \Sigma (Y-\bar{Y})^2}}}$

③ $r = \boxed{\dfrac{n \cdot \Sigma X \cdot Y - \Sigma X \cdot \Sigma Y}{\sqrt{n \Sigma X^2 - (\Sigma X)^2 \cdot n \Sigma Y^2 - (\Sigma Y)^2}}}$

* Characteristics of correlation coeff. (r)

① → Independent of change in origin or scale.

② → If variable (X, Y) replaced by (U, V)

$\boxed{U = aX + b}$ $\boxed{V = cY + d}$

$\boxed{r(U,V) = r(X,Y)}$ — Sign of these

If sign of A & C are		result
A	C	
+	+	$+r(X,Y) = +r(U,V)$
−	+	$r(U,V) = -r(X,Y)$
+	−	$r(U,V) = -r(X,Y)$
−	−	$r(U,V) = +r(X,Y)$

$a \times c$
$|a|\,|c|$ → If a & c are * Arbitrary constants. *

③ Spearman's Rank Correlation coff. method → denote by (ρ)

used for Qualitative data

like → ~~weight~~, ~~weight~~, Ability, Honesty, Beauty etc

ρ Range from $(-1 \le \rho \le 1)$

→ $\rho = 1$ then same rank by (Judges) (Identical)

→ $\rho = -1$ Reverse ranks.

Formula = $\rho = 1 - \dfrac{6\sum D^2}{n(n^2-1)}$

* when ranks not equal

D − Deviation of rank (Difference of rank)

n = Total no. of pair

* when ranks are equal = $1 - \dfrac{6\left(\sum D^2 + \sum \frac{(t^3 - t)}{12}\right)}{n(n^2-1)}$

t = tie count

(4) Coff. of Concurrent deviation → Denoted by rc

* $\boxed{r_c \Rightarrow \pm\sqrt{\pm\frac{2c-m}{m}}}$ = Jo sign Andar hai vo bahar

Data analysis (2)

m → Total Pair − 1

c → (no. of Pair) which (Have same sign)

* Regression * Check how much one variable depending on another.

Regⁿ eqⁿ → A mathematical eqⁿ that allows us to predict value of one variable from known value of one or more variable is called regression eqⁿ.

- Dependent variable whose value is to be predicted from Independent variable

(1) X on Y
(2) Y on X

- Independent variable where value is used to find other dependent called independent

* (Simple Regression Analysis), (Simple linear Analysis)
→ Study only (2) variables: dependent, Independent

- Let say Y depend on X : $Y = a + bX$ → regression line of Y on X (Y dependent, X Independent)

- A regression line is a line which gives minimum sum of vertical distance from points:
$\Sigma(Y_i = a + bx_i) \Rightarrow$ Min. sum

* Method of least sq. regression
(1) Y on X

$\hookrightarrow \Sigma Y = n \cdot a + b\Sigma x$ (1)

$\Sigma xY = a\Sigma x + b(x \cdot x)$ (2)

Do sum in 1 eq, Multiply x in 2nd equation & sum.
& Multiply (a) by n.
$\boxed{\text{constant}}$

→ Then eqn we get are →

$$\left.\begin{aligned} (1)\quad \Sigma xy &= a\Sigma x + b\Sigma x^2 \\ (2)\quad \Sigma y &= n \cdot a + b\Sigma x \end{aligned}\right] \text{ These are } \boxed{\text{Inormal}} \text{ eqn of } \boxed{Y \text{ on } X}$$

$$b = \frac{Cov(x,y)}{var\ of\ x} \quad (1) \quad \Bigg| \quad \frac{n\Sigma X \cdot Y - \Sigma X \cdot \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2} \quad (2)$$

→ $\boxed{y = a + bx}$ → iske done ($\bar{x}$ or $\bar{X}$) Mean lie kste hai

$$\bar{Y} = a + b \cdot \bar{X}$$

$\hookrightarrow$ (1) $\frac{Cov(X,Y)}{Var\ of\ x}$ | (2) $\frac{n\Sigma X \cdot Y - \Sigma X \cdot \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2}$

* $\boxed{Y \text{ on } X}$ $\boxed{\text{regression coff.}}$

$Y = a_{yx} + b_{yx}x$

$\boxed{a_{yx}}$ (1) , $\boxed{b_{yx}}$ (2)

$$\boxed{a_{yx} = \bar{Y} - b_{yx} \cdot \bar{X}}$$

② $\boxed{X \text{ on } Y}$

$X = c + dy$: $x = a + by$

$\boxed{a_{xy}}$ $\boxed{b_{xy}}$ only this $\boxed{\text{regression coff.}}$

①

$\boxed{b_{xy}} = \frac{Cov(X,Y)}{\text{Variance of } Y}$ | $\frac{n \cdot \Sigma X \cdot Y - \Sigma X \cdot \Sigma Y}{n \cdot \Sigma Y^2 - (\Sigma Y)^2}$

$\boxed{a_{xy}} = \bar{X} - b_{xy} \cdot \bar{Y}$

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

$\boxed{r, \; b_{xy}, \; b_{yx}} \rightarrow$ $\boxed{\text{All three Have same sign.}}$ either all +ive, or all -ive.

* $\boxed{\text{Some imp results.}}$ * of $\boxed{Y \text{ on } X}$

① $\boxed{Y \text{ on } X} \rightarrow$ $Y = a_{yx} + b_{yx} \cdot x$ — ①

$\boxed{a_{yx}} = \bar{Y} - b_{yx} \cdot \bar{X}$ — ②

Put ② in ①

$= Y = \bar{Y} - b_{yx} \cdot \bar{X} + b_{yx} \cdot x$

$\boxed{(Y - \bar{Y}) = b_{yx} (X - \bar{X})}$

1. Put these b_{yx} value

$\boxed{(Y - \bar{Y}) = \frac{r \sigma y}{\sigma x} (X - \bar{X})}$

* $\boxed{\text{Slope of } Y \text{ on } X \text{ is } = b_{yx}}$ *

② $r = \frac{Cov(X,Y)}{\sigma x \cdot \sigma y}$ (a) | $b_{yx} = \frac{Cov(X,Y)}{(\sigma x)^2}$

Divide (b) by (a)

$\frac{r}{b_{yx}} = \frac{\cancel{Cov(X,Y)}}{\sigma x \cdot \sigma y} \times \frac{(\sigma x)(\sigma x)}{\cancel{Cov(X,Y)}}$

$\frac{r}{b_{yx}} = \frac{\sigma x}{\sigma y}$ $\boxed{b_{yx} = \frac{r \sigma y}{\sigma x}}$ ②

(1) $\boxed{x = a + by}$ (1)

$a = \bar{X} - b \cdot \bar{Y}$ (2)

Put (2) in (1)

$x = \bar{X} - b \cdot \bar{Y} + by$

$\boxed{(X - \bar{X}) = b_{xy}(Y - \bar{Y})}$ (1')

(2) $r = \dfrac{Cov(X,Y)}{\sigma x \cdot \sigma y}$ (a)

$b_{xy} = \dfrac{Cov(X,Y)}{(\sigma y)^2}$ (b)

Divide (a) by (b)

$\dfrac{r}{b_{xy}} = \dfrac{\cancel{Cov(X,Y)}}{\sigma x \cdot \cancel{\sigma y}} \times \dfrac{\cancel{(\sigma y)}(\sigma y)}{\cancel{Cov(X,Y)}}$

$\dfrac{r}{b_{xy}} = \dfrac{\sigma y}{\sigma x} =$ $\boxed{b_{xy} = \dfrac{r \cdot \sigma x}{\sigma y}}$ (2')

Put b_{xy} value in the

$\boxed{(X - \bar{X}) = r \cdot \dfrac{\sigma x}{\sigma y}(Y - \bar{Y})}$ (3)

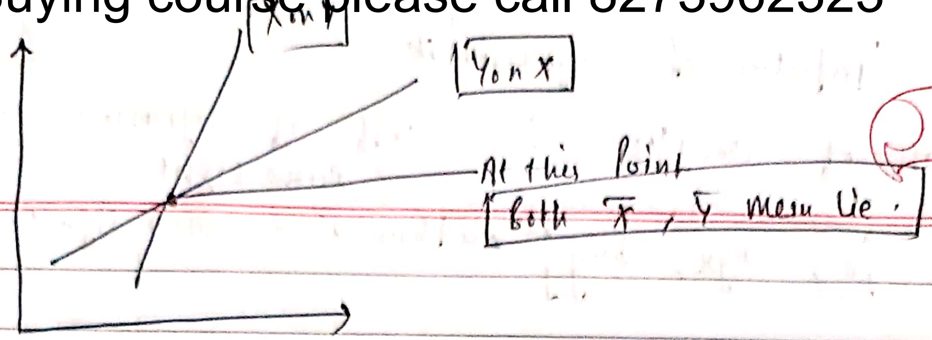
$\boxed{\text{Slope of } X \text{ on } Y}$ → $\boxed{\dfrac{1}{b_{xy}}}$

* $b_{xy} \times b_{yx} = r \cdot \dfrac{\sigma x}{\cancel{\sigma y}} \times r \cdot \dfrac{\cancel{\sigma y}}{\sigma x}$

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

$\boxed{r = \sqrt{b_{xy} \cdot b_{yx}}}$ $\boxed{\text{Hence Proved}}$

(All Have equal sign)



* Properties of Least Sq - Regression method. *

1. The constant (b) called $\boxed{b_{yx}}$ & regression coff. of $\boxed{Y \text{ on } X}$
2. It measure the change in (Y) According to X.
3. Thus b_{yx} represent the slope of $\boxed{Y \text{ on } X}$
4. The eqn of line of regression of Y on X can be written as

$$(Y-\bar{Y}) = \boxed{b_{yx}}(X-\bar{X})$$

↳ here ① $\boxed{r \cdot \frac{\sigma y}{\sigma x}}$ or ② $\frac{Cov(X,Y)}{(\sigma x)^2}$ or ③ $\frac{n \cdot \Sigma X \cdot Y - \Sigma X \cdot \Sigma Y}{n \cdot \Sigma X^2 - (\Sigma X)^2}$

5. On the other Hand if wish to estimate a value of (X) for a given value of (Y) we have to obtain regression line of $\boxed{X \text{ on } Y}$

$$(X-\bar{X}) = \boxed{b_{xy}}(Y-\bar{Y})$$

↳ ① $\boxed{r \cdot \frac{\sigma x}{\sigma y}}$, ② $\frac{Cov(X,Y)}{\sigma y^2}$ | ③ $\frac{n \cdot \Sigma X \cdot Y - \Sigma X \cdot \Sigma Y}{n \cdot \Sigma Y^2 - (\Sigma Y)^2}$

6. The constant (d) represent regression coff. of $\boxed{X \text{ on } Y}$ & denote by b_{xy}. It measure the change in X According to Y.
7. $\frac{1}{b_{xy}}$ is the slope of regression line $\boxed{X \text{ on } Y}$

classmate

Date
Page

* Properties of regression *

① The coff. of coreletion & coff. of regression Have same sighs

$\boxed{r, \; b_{yx}, \; b_{xy}}$ → All three Have same sigh.

② The coff. of coreletion is the G.M. of the 2 regression cofficients $= r = \sqrt{b_{xy} \times b_{yx}}$

ⓡ = $\boxed{-1 \le r \le 1}$ ①

② $\boxed{0 \le b_{xy} \cdot b_{yx} \le 1}$ or $\boxed{b_{yx} \cdot b_{xy} \le 1}$

③ If $b_{yx} > 1$ than b_{xy} will less than 1 or vice-versa.

③ If one the regression coff. is greater than unity (one)(1) than other must be less than 1.

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

$\boxed{-1 \le r \le 1}$ ① , $\boxed{0 \le r^2 \le 1}$ ②

③ $\boxed{0 \le b_{xy} \cdot b_{yx} \le 1}$

④ If $b_{xy} \cdot b_{yx} \le 1$, then if $b_{xy} \ge 1$ than $b_{yx} \le 1$ or (Vice-versa)

④ The 2 lines of regression intersect at point $(\bar{x}, \bar{y})$ where x & y variable under consideration.

* ⑤ * The regression coff. are independent of change of origin but not for scale.

Impr. * Property *

(Ex) Reqn = $u = p(x) + q$ $v = r(y) + s$

If we need $\boxed{b_{vu}} = \frac{\text{coff of } v}{\text{coff of } u} \times b_{yx}$ | $b_{uv} = \frac{\text{coff of } u}{v} \times b_{xy}$

→ Correlation Tells → Changes, variation caused by one in another (like) change in Profit According to Sales

→ Spurious Correlation → where there is no casual relation b/w 2 variables due to interference of third, this is called spurious correlation or non-sense correlation.

Coff. of Determination → denote by r^2

$$\boxed{r^2} = \frac{\text{Explained variance}}{\text{Total variance}}$$

for exe: $r = .7$

$r^2 = (.7)^2$

$r^2 = \boxed{0.49}$ or $\frac{49}{100}$

Means 49% is Explained.

So Non explained is $100 - 49 = 51\%$ (unexplained)

Explained

Also called Coff. of non determination

Total (−) Explained

or $= \dfrac{\text{unexplained}}{\text{Total variance}}$