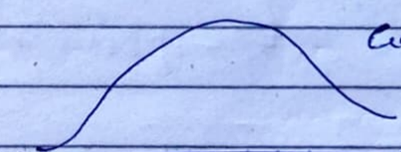
[Statistics - day 4]

Topics:-

1. central limit theorem
2. Probability
3. Permutation and Combination
4. Covariance, Pearson correlation, Spearman Rank correlation.

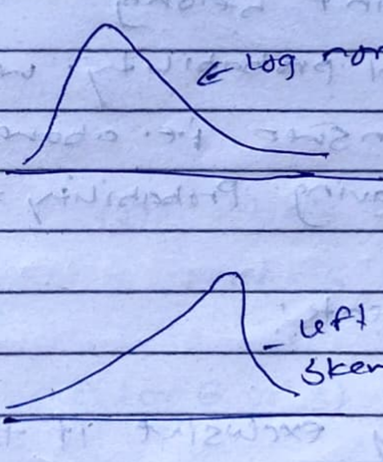
① central limit theorem:-



Gaussian

Size of sample $= n$

no. of samples $= m$

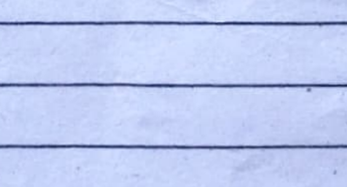


← log normal

$\rightarrow S_1 = \{x_1, x_2, x_3 \cdots\} = \bar{x}_1$

$\rightarrow S_2 = \{x_1\ x_2\ x_3 \cdots\} = \bar{x}_2$

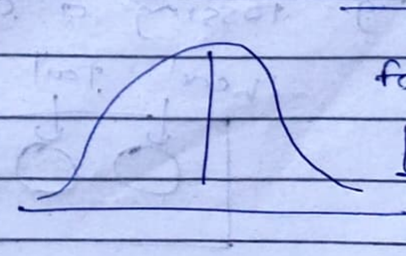
$\rightarrow S_3 = \{x_1, x_2, x_3, \cdots\} = \bar{x}_3$



– left skewness

If we draw the mean of all samples

it shows Gaussian distribution



for value $\boxed{n \geq 30}$

(2) Probability :-

measure of the likehood of an event.

e.g. → (1) Tossing a fair coin

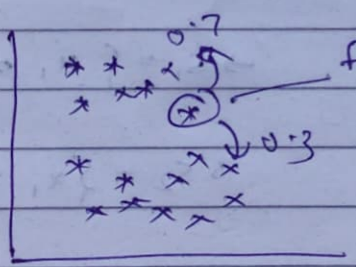
$P(H) = 0.5$ $\quad$ $P(T) = 0.5$

(2) Rolling a dice

$P(1) = \frac{1}{6}$ $\quad$ $P(2) = \frac{1}{6}$ .. etc.

<u>why it is used</u> !

suppose we have

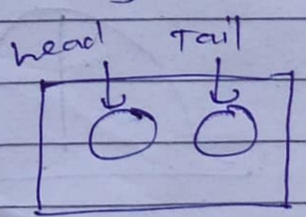


from which does this point belong by probability we can answer i.e. above having Probability $= 0.7$

(1) mutual exclusive events !

→ The events mutually exclusive if they cannot occur at the same time.

e.g. tossing a coin



if <u>head</u> comes then <u>tail</u> cannot occur

Page No.: Date. / /

② Rolling a dice

if 1 come in dice then other numbers cannot occur.

non-mutual exclusive

① Picking randomly a card from a dice deck of cards the events "heart" and "King" can be selected.

Problems :-

① what is the Probability of coin landing on heads or tails.

⇓

Addition rule for mutual exclusive events.

$$P(A \text{ or } B) = P(A) + P(B)$$
$$= 0.5 + 0.5$$
$$= 1$$

② what is probability of getting 1 or 6 or 3 while rolling a dice

$$P(1 \text{ or } 6 \text{ or } 3) = P(1) + P(6) + P(3)$$
$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

Non mutual exclusive Event:

events are non mutual exclusive if they occure at same time

Bag of Marbles = 10 red, 6 Green
3 (R & G)

when picking randomly from a bag of marbles what is the probability of choosing a marble that is red and green]

Addition rule:-

(1) $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

$= \frac{13}{19} + \frac{9}{19} - \frac{3}{19}$

$= \frac{19}{19}$

$= 1$

(2)

$P(\text{heart or queen}) = P(\text{heart}) + P(\text{queen}) - P(\text{heart and queen})$

$= \frac{13}{52} + \frac{4}{52} - \frac{1}{52}$

$= \frac{16}{52} = \frac{8}{26} = \frac{4}{13}$

Page No.:
Date. / /

Multiplication Rule:-

① Dependent event:- if they affect one another. then it is dependent event.

Bag of marble = $\left\{ \begin{matrix} \underbrace{0000}_{} \\ 000 \end{matrix} \right.$ → 4 white → 3 yellow

if 1 white marble selected then

$\boxed{P(W) = \frac{4}{7}}$ and $\boxed{total = 7-1 = 6}$

then,

$P(W) = \frac{3}{6}$

example:- what is the probability of rolling a "5" and then a "3" with a normal 6 sided dice?

→ Independent

$P(1) = \frac{1}{6}$ $P(2) = \frac{1}{6}$ * $P(3) = \frac{1}{6}$, $P(5) = \frac{1}{6}$

$$P(A \text{ and } B) = P(A) * P(B)$$
$$= \frac{1}{6} * \frac{1}{6}$$

$$P(A \text{ and } B) = \frac{1}{36}$$

Page No.:
Date. / /

Permutation :-

in how many ways 3 chocklets

= { Dairy milk, kit kat, milky bar, Sneakers, 5 star }

$= 5 \times 4 \times 3$

$= \boxed{60 \text{ way}}$ ← permutation

formula is :-

$$^{n}P_{r} = \frac{n!}{(n-r)!} = \frac{5!}{2!} = \frac{5 \times 4 \times 3!}{3!}$$

$$= \frac{5 \times 4 \times 3 \times \cancel{2!}}{\cancel{2!}}$$

$$= 60$$

Combination :-

Repetation is not allowed:

formula: -

$$^{n}C_{r} = \frac{n!}{r!(n-r)!} = \frac{5!}{3!(2!)}$$

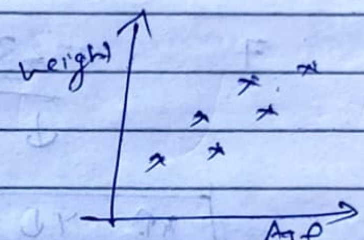
$$= \frac{5 \times 4 \times \cancel{3!}}{\cancel{3!} \times 2!} = \frac{20}{2}$$

$^{n}C_{r}$ $= 10$ ways

(*) Covarience $\xrightarrow{\text{used in}}$ feature Selection

X	Y
Age	weight
12	40
13	45
15	48
17	60
18	62
$\bar{x} = 15$	$\bar{y} = 51$

$\xrightarrow{\text{by looking at data}}$ $\boxed{\begin{array}{l} \text{Age}\uparrow \ \text{weight}\uparrow \\ \text{Age}\downarrow \ \text{weight}\downarrow \end{array}}$ ← quantify the relationship x & y using mathmatical question



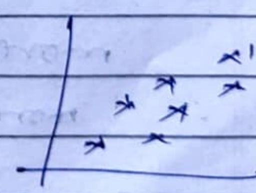
$$\text{cov}(x, y) = \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{n-1}$$

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

$$s^2 = \frac{\sum (x_i - \bar{x}) * (x_i - \bar{x})}{n-1}$$

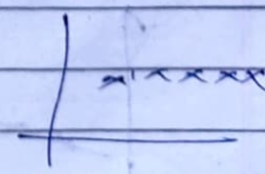
$$\boxed{\text{cov}(x, x) = \text{var}(x)}$$

+ Covariance $\boxed{\begin{array}{ll} x\uparrow & y\uparrow \\ x\downarrow & y\downarrow \end{array}}$



− Covariance $\boxed{\begin{array}{ll} x\uparrow & y\downarrow \\ x\downarrow & y\uparrow \end{array}}$

0 − covariance $\boxed{\text{no relation}}$



x	y
10	4
8	6
7	8
6	10
7.75	7

$$\text{cov}(x,y) = \underline{-ve}$$

$$= \left[(10-7.75)(4-7) + (8-7.75)(6-7) + (7-7.75)(8-7) + (6-7.75)(10-7)\right]/3$$

$$= \underline{-3.25}$$

$\downarrow$

$x\uparrow$	$y\downarrow$
$x\downarrow$	$y\uparrow$

Pearson Correlation coefficient (-1 to 1)

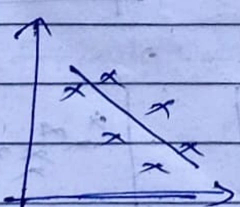
why to use this?

$\downarrow$

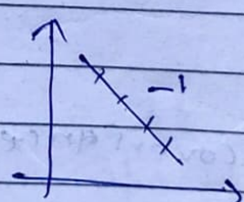
because covariance doesn't provide Scale

$$\rho(x,y) = \frac{\text{cov}(x,y)}{\sigma_x \cdot \sigma_y}$$

more the value towards +1
more +ve correlated is

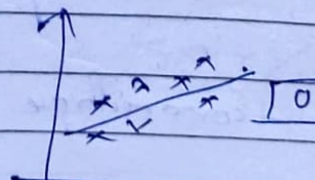


-1 to 0



$x\downarrow$ $y\uparrow$

$x\uparrow$ $y\downarrow$



[0 to 1]

Page No.:
Date. / /

(*) Spearman Rank Correlation:

$$r_s = \frac{Cov(R(x), R(y))}{\sigma(R(x)) * \sigma(R(y))}$$

x	y
10	4
2	6
7	8
6	10

R(x)	R(y)
4	1
3	2
2	3
1	4

→ Ascending

→ Spearman Rank correlation

why to use?

They are highly correlated

x y z O/p

correlation(y) = 0.2

there we can drop this

and x and z are 98% correlated

there we can drop (2)

A exp Degree city Sallary

96%