Subject: Data Science

MON TUE WED THR FRI SAT SUN

1, st September, 2022

Statistics Note (Day-4)

by Krish Naik Sir

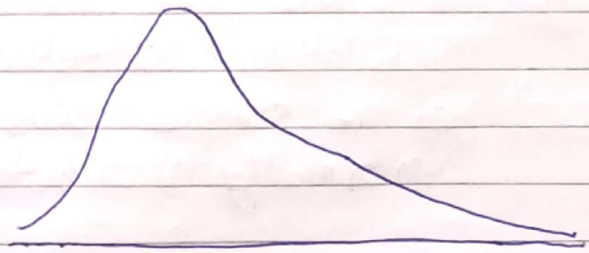
(A) Agenda :-

1. Central Limit Theorem
2. Probability
3. Permutation And Combination
4. Covariance
5. Pearson Correlation
6. Spearman Rank Correlation

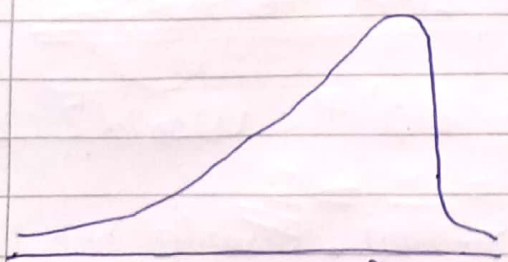
(1) Central Limit Theorem :-

Population Data (N)

[Normal Distribution]



(Log Normal Distribution)



(~~Right~~ ~~Skewed~~ Left Skewed)

* In the above Distribution, data if we consider sample data (n)

Let 1st Sample $\Rightarrow \{x_1, x_2, x_3, \ldots x_n\} \Rightarrow \bar{x}_1$

2nd Sample $\Rightarrow \{x_1, x_2, x_3 \ldots x_n\} \Rightarrow \bar{x}_2$

3rd Sample $\Rightarrow \{x_1, x_2, x_3 \ldots x_n\} \Rightarrow \bar{x}_3$

$\vdots$

$\bar{x}_m$

$\therefore$ Size of Sample $\Rightarrow$ n

$\therefore$ no. of samples $\Rightarrow$ m

Subject _____

MON TUE WED THR FRI SAT SUN

Let, $n \geq 30$, if we plot all the sample data then the central limit theorem, Then we will get a Gaussian / Normal Distribution.

Let us Consider a population data which may be normally distributed or not be then the central Limit Theorem say that if we take sample of size $n \geq 30$ and selected 'm' no of. samples of respective mean and if we plot the data then we will get normal Distribution of the sample mean.

Defination of Central Limit Theorem :—

The central Limit Theorem States that if you have a population with mean μ and Standard Deviation σ and take sufficient large random samples from the population with replacement, then the distribution of the sample means will be approximately normally Distributed.

By Increasing sample Size, we can smoothen the curve.

Importance of Central Limit Theorem :—

Size of Shark through out the world?

⇒ We will take 10 Different region, sample population of size $n \geq 30$ and make an assumption on the size of Shark.

Subject _____ MON TUE WED THR FRI SAT SUN

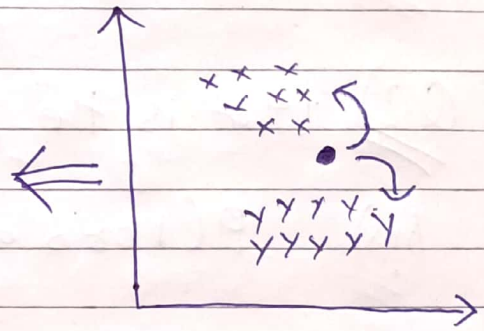
② **Probability :-** Probability is a measure of the likehood of an event

Eg:- Tossing a fair coin ⇒ $P(H) = 0.5$ $P(T) = 0.5$

Rolling a Dice, $P(1) = \frac{1}{6}$, $P(2) = \frac{1}{6}$, $P(3) = \frac{1}{6}$

③ To check whether '•' data belongs to '× ×× ×××' or 'y yy yyy'

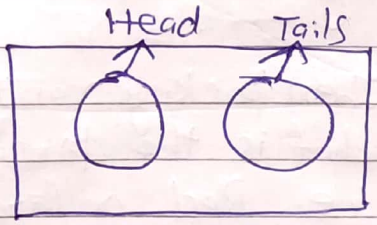
We will use possibility



(i) **Mutual Exclusive Event :-**

⇒ Two events are mutually exclusive if they cannot occur at the same time.

Eg:- (a) Tossing a coin ⇒



(b) Rolling a Dice.

(ii) **NoN - Mutual Exclusive Event :-**

⇒ Two Event can occur at the same time.

Eg:- (a) Winning and Losing game.

(b) Picking Randomly a card from a deck of cards, Two events "Heart" and "King" - Can be selected.

Subject _____ MON TUE WED THR FRI SAT SUN

Problem Statement :- Mutual Exclusive Event:-

Q.1) What is the probability of coin landing on heads or Tails ?

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

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

Q.2) What is the probability of geting 1 or 6 or 3.

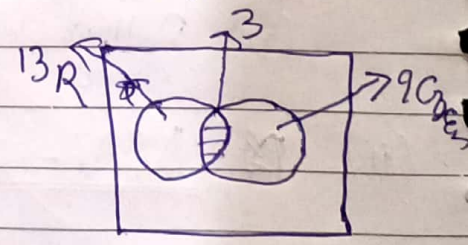
Ans) $P(1 \text{ or } 6 \text{ or } 3) = P(1) + P(6) + P(3)$

$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \Rightarrow \frac{3}{6} \Rightarrow \frac{1}{2}$

Problem Statement :- Non- Mutual Exclusive Event:-

Q.1) Bag of Marbles : 10 Red , 6 Green , 3 (R & Gr)

When picking randomly from a bag of marbles. What is the probability of choosing a marble i.e. Red or Green.

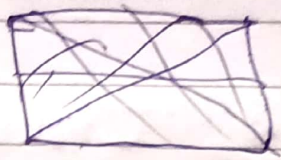


Ans) $P(R \text{ or } G_r) = P(R) + P(G_r) - P(R \text{ and } G_r)$

$= \frac{13}{19} + \frac{9}{19} - \frac{3}{19} = \frac{19}{19} = 1$

Subject _____

MON TUE WED THR FRI SAT SUN



Deck of Cards :-
What is the probability of choosing Heart and Queen?

$$P(H \text{ or } Q) = P(H) + P(Q) - P(H \text{ and } Q)$$

$$= \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{13}{52} + \frac{3}{52} = \frac{16}{52}$$

Multiplication Rule :-

Dependents Events :- Two events are dependent - if they effect one another

Eg :- Bag of Marbles { White - 4, Yellow - 3 }

$P(W) = \frac{4}{7}$ → $P(y) = \frac{3}{6}$

↑ 1 white marble Takeout

∴ Here white marbel effecting yellow marble's probability of occuring.

Q.2) What is the probability of rolling a "5" and then a "3" with a normal 6 size sided dice?

Ans $P(A \text{ and } B) = P(A) * P(B)$

$= \frac{1}{6} * \frac{1}{6} = \frac{1}{36}$

Subject _____ MON TUE WED THR FRI SAT SUN

Independent Event :—

Eg:- Tossing a coin :-

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

Q.2) There are bag of marbles (4 orange, 3 yellow) What is the probability of drawing a "orange" and then drawing a "yellow" marble from the bag [Dependent Event]

Ans Total No. of marbles = 4+3 = 7

probability of orange, $P(O) = \frac{4}{3}$

After taking out 1 marble orange

then, Probability of taking out yellow marble $P(Y/O) = \frac{3}{6} = \frac{1}{2}$

⇓

[Conditional Probability]

$\therefore P(O \text{ and } Y) = P(O) * P(Y/O)$

$= \frac{4}{7} * \frac{1}{2}$

$= \frac{4}{14}$

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

your writing partner

Subject _____ MON TUE WED THR FRI SAT SUN

(3) PERMUTATION :-

⇒ All the possible arrangement {Daily Milk, Kit Kat, Milky Bar, Sneakers, 5 Stars}

$5 * 4 * 3$

= 60 ways chocolate can be choosen

(*) With Permutation order matter

$$^{n}P_{r} = \frac{n!}{(n-r)!}$$

n = total no. of object = 5
r = no. of selection = 3

$$= \frac{n!}{(n-r)!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{(5-3)!} \quad \frac{5!}{(5-3)!}$$

$$= \frac{5 \times 4 \times 3 \times 2!}{2!} = 60$$

(4) Combination :-

* Repeation will not occur.
* Unique combination possible

Formula :-

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

Eg
Dream 11
Σ(Permutation)

$$= \frac{5 \times 4 \times 3 \times 2!}{3 \times 2 \times 2!} = 10$$

Choosing 3 Deserts from 10 of menu Dishes
Combination

∴ The Repeated one is removed

∴ Permutation value > combination value.

Subject

MON TUE WED THR FRI SAT SUN

(5) COVARIANCE [Feature Selection] :-

Age(x)	Weight(y)
12	40
13	45
15	48
17	60
18	62

Age ↑ Weight ↑
Age ↓ Weight ↓

(6) Quantity the relationship x & y using Mathematical Question!

$$\therefore \boxed{Cov(x,y) = \frac{\Sigma (x_i - \bar{x}) \times (y_i - \bar{y})}{n-1}}$$

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

$$= \frac{\Sigma (x_i - \bar{x}) \times \Sigma (x_i - \bar{x})}{n-1} = Cov(x,x)$$

$$\boxed{Cov(x,x) = Var(x)} \quad \therefore \text{(Interview Question)}$$

Subject _____

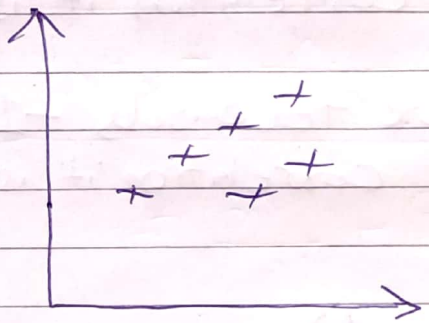
MON TUE WED THR FRI SAT SUN

Now, $\bar{x} = 15$, $\bar{y} = 51$

① +ve Covariance $x\uparrow\ y\uparrow$, $x\downarrow\ y\downarrow$

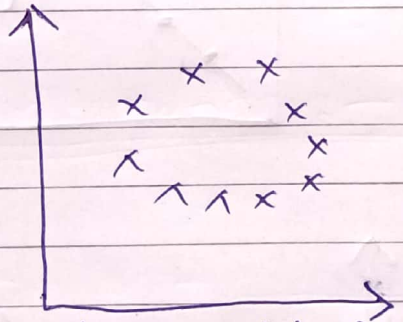
② −ve Covariance $x\uparrow\ y\downarrow$, $x\downarrow\ y\uparrow$

③ Covariance 0 [No relation with x & y]

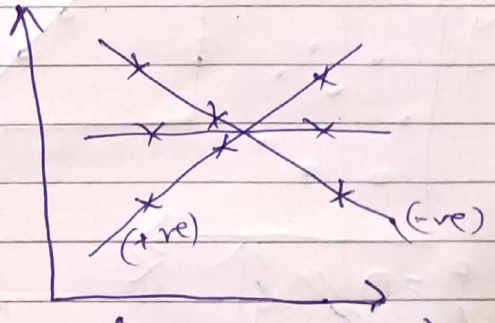


(+ve Covariance)

(−ve Covariance)



(No Relation)



[No Relation]

X	y
10	4
8	6
7	8
6	10

$\bar{x} = 7.75$ $\bar{y} = 7$

$$Cov(x,y) = \frac{(10-7.75)(4-7)+(8-7.75)(6-7)+(7-7.75)(8-7)+(6-7.75)(10-7)}{3}$$

$$= \frac{-13}{3} = -4.33 //$$

Subject _____

MON TUE WED THR FRI SAT SUN

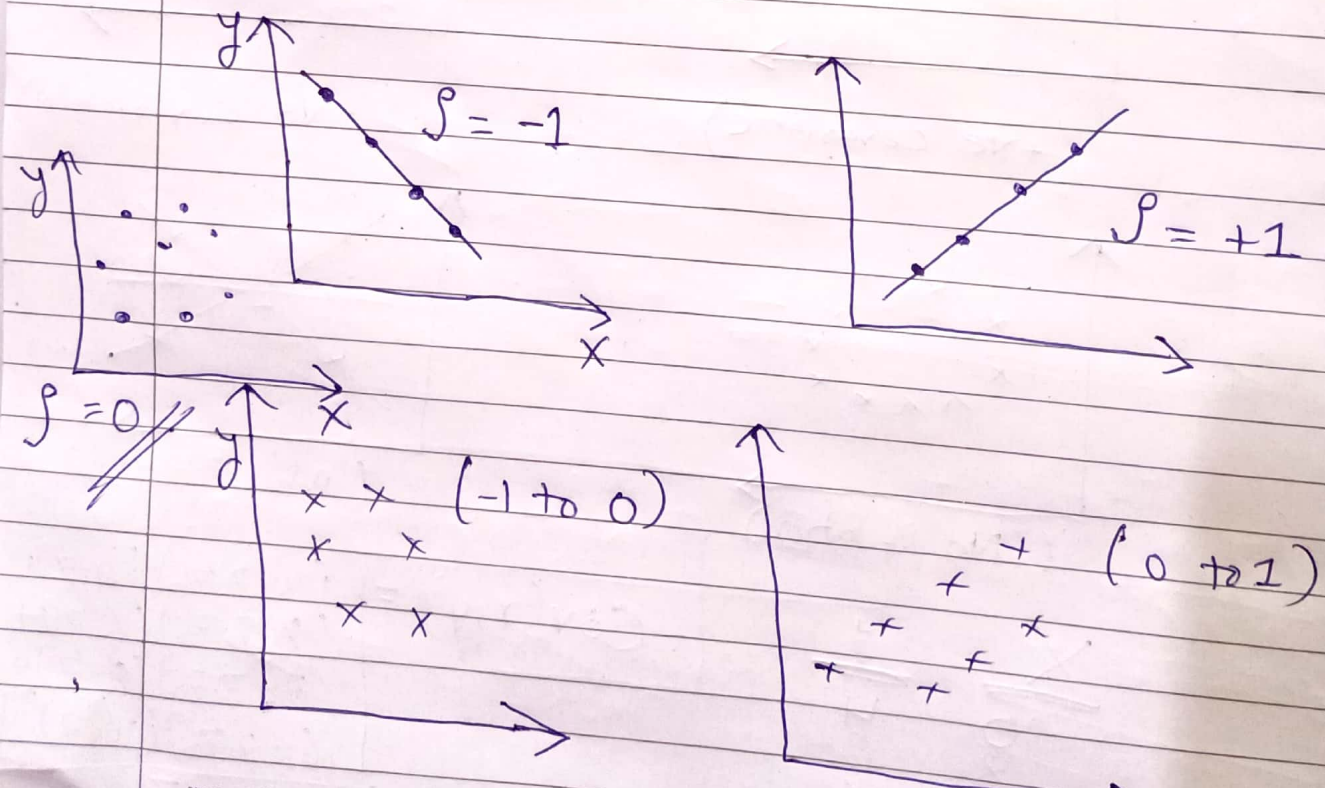
6. Pearson Correlation Coefficient [-1 & 1] :-

$$\rho(x,y) = \frac{Cov(x,y)}{\sigma x \, \sigma y}$$

* With the help of this we are trying to restrict the value between -1 and 1.

* More the value toward +1
  ⇒ More +ve correlation it is

* And, More the value towards -1.
  ⇒ More -ve correlation it is

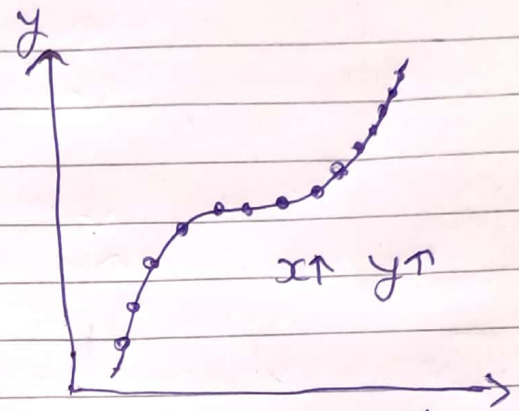


Subject _____

MON TUE WED THR FRI SAT SUN

(71) Spearman Rank Correlations :—

For Non-Linear Data we use Spearman Rank Correlations.



Spearman Correlation = 1
Pearson Correlation = 0.88

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

@ Rank = Assinging value by number in Ascending order

x	y	R(x)	R(y)
10	4	4	1
8	6	3	2
7	8	2	3
6	10	1	4