

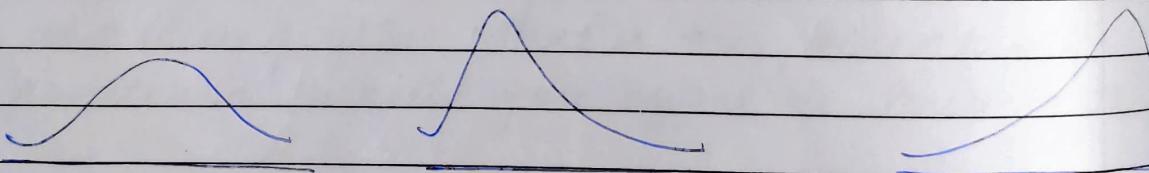
Day 4 Stats

- (1) central limit theorem
- (2) Probability
- (3) Permutation and combination
- (4) Covariance, Pearson correlation, Spearman Rank correlation
- (5) Bernoulli distribution
- (6) Binomial distribution
- (7) Poisson law & Parity distribution)

V = V - true

Central Limit Theorem

- For any kind of distribution
- Distribution can be



Gaussian /
Normal
dist.

log normal
dist

left + skewed

or some other distribution

Population data - N

sample data (n)

No.	Sample data	mean
s_1 (first) $\rightarrow \{n_1, n_2, n_3, \dots, n_n\}$		\bar{n}_1

$$\text{Population } \{x_1, x_2, \dots, x_3, \dots, x_n\} \rightarrow \bar{x}_2$$

$$\text{Sample } \{x_2, x_3, \dots, x_n\} \rightarrow \bar{x}_3$$

$$\{x_1, x_2, \dots, x_{n-1}\} \rightarrow \bar{x}_4$$

⋮

∴ for this we will take m no. of samples

$$S_m \{x_3, \dots, x_n\} \rightarrow \bar{x}_m$$

Here we have to take sample size (n) greater than 30

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

For any kind of distribution where population datum
 If we take m no. of samples where sample size is ≥ 30 and if we take the mean of all the samples mean and plot it as a histogram then it will be gaussian plot / Normally distributed

If $n < 30$, then central limit theorem will not be applicable

$n \geq 30$ is scientifically framed

Q) Size of shark throughout the world?

Soln: Take sample of size ≥ 30 .

Let us take 10 nos. of samples from different regions.
Then plot the mean of all the samples.
We will get Gaussian distribution and we can do many assumptions.

* More no. of m will give more accurate results.

Probability

Probability is a measure of the likelihood of an event.

Eg - Tossing of a coin (unbiased)

Head - $1/2$

Tail - $1/2$

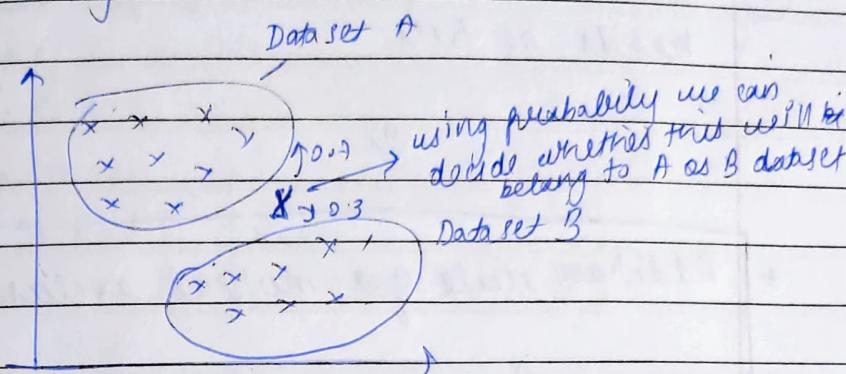
- Rolling of a dice

$$P(1) = 1/6 \quad (\text{Prob. of getting } 1)$$

$$P(6) = 1/6$$

→ In some companies aptitude questions like this will be asked (only for freshers).

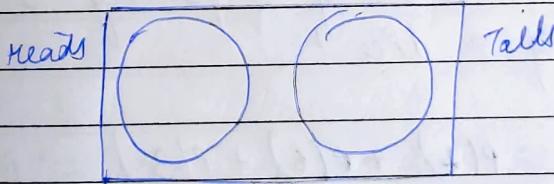
use of Probability



① mutually Exclusive

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

Ex - Tossing a coin



→ Dice rolling of a dice

② Non mutual except exclusive Events

Two events can occur at the same time

→ Picking ^{two} red card from a deck of card.

Mutual Exclusive Ques :-

- (Q) what is the probability of coin landing on heads or tails

Ans

* Addition rule for mutual exclusive events

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

$$\begin{aligned} P(\text{Heads or Tails}) &= P(\text{Heads}) + P(\text{Tails}) \\ &= \frac{1}{2} + \frac{1}{2} \\ &= 1 \end{aligned}$$

- (Q) what is the probability of getting 1 or 6 or 3 while rolling a dice

$\frac{1}{6} + \frac{1}{6} + \frac{1}{6}$

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

Non mutual Exclusive Event

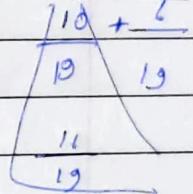
from a bag of marbles

10 - Red

6 - Green

3 - Red & Green

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



* Addition rule for non mutual exclusive events

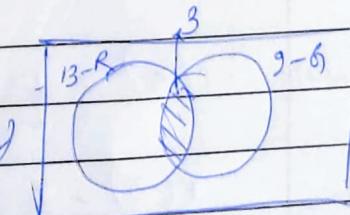
$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Q3

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

$$= \frac{10}{19} + \frac{6}{19} - \frac{3}{19}$$

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



(Q) From a deck of cards what is the probability of choosing Heart or Queen

Soln: Heart - 13
Queen - 4

$$\begin{aligned} P(H \cup Q) &= P(H) + P(Q) - P(H \cap Q) \\ &= \frac{13}{52} + \frac{4}{52} - \frac{1}{52} \\ &= \frac{16}{52} \end{aligned}$$

* Multiplication Rule

* Dependent Events :- Two events are dependent if they affect one another

(Q) What is the probability of rolling a "5" and then a "3" with a 6-sided die

Soln: $P(5) = \frac{1}{6}$

$P(3) = \frac{1}{6}$

(Independent Event)

$$P = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

multiplication rule for independent events

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

$$= \frac{1}{2} * \frac{1}{3} = \frac{1}{6}$$

From a bag of 4 orange and 3 yellow marbles.

Probability of drawing a "orange" and then drawing a "yellow" marble from the bag?

$$P(\text{orange}) = \frac{4c_1}{6c_1} = \frac{4}{6}$$

$$P(\text{yellow})$$

$$P(\text{yellow}) = \frac{3c_1}{6c_1} = \frac{3}{6}$$

$$\frac{4}{6} * \frac{3}{5} = \frac{2}{5} \quad (\text{Ans})$$

$$P(A) * P(B|A)$$

↑

conditional Probability

Conditional Probability will be used in ML for Naive Bayes.

Permutation

- with permutation, order matters
- Possible arrangements

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

n = total no. of objects

r = no. of selections

Ex.

chocolate = { Dairy milk, kit kat, sneakers,
5 stars, milkybar }

If we have to do selection of 3 chocolates
from the set then no. of possibilities will

$$5 \times 4 \times 3 = 60$$

{ DM, KK, SS }
{ DM, SS, KK }

$$5P_3 = \frac{5!}{(5-3)!} = \frac{5!}{2!} = \frac{5 \times 4 \times 3 \times 2!}{2!} = 60$$

Combination

- unique combination
- no repeated values

$\{ DM, KK, SS \}$
 $\{ DY, SS, KK \} X$

$$n_{C_3} = \frac{n!}{r!(n-r)!}$$

$$n_{C_3} = \frac{5!}{3!2!} = \frac{5 \times 4 \times 3!}{3! \times 2 \times 1} = 10$$

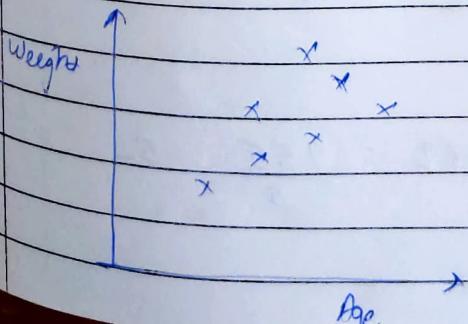
ij

Covariance

feature selection is done using this

Age	Weight
12	40
13	45
15	48
17	60
18	62

conclusion Age ↑ weight ↑
 Age ↓ weight ↓



Formula:

$$\text{Cov}(X, Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Variance of n

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

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

$$\sigma(x, n)$$

★ interview dues

$$\therefore \text{Cov}(X, X) = \text{Variance of } X$$

- Q) From the previous dataset of Age, Age and weight quantify the relationship X & Y using mathematical question

$$\bar{x} = \frac{12 + 13 + 15 + 17 + 18}{5} = 15$$

$$\bar{y} = \frac{40 + 45 + 48 + 60 + 62}{5} = \frac{255}{5} = 51$$

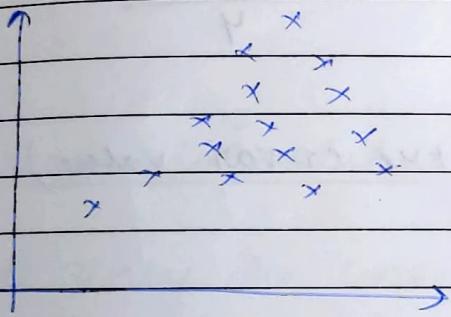
+ve covariance

$X \uparrow$	$Y \uparrow$
$X \downarrow$	$Y \downarrow$

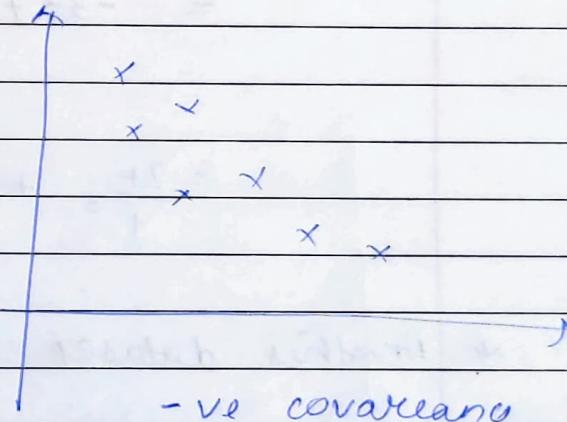
-ve covariance

$X \uparrow$	$Y \downarrow$
$X \downarrow$	$Y \uparrow$

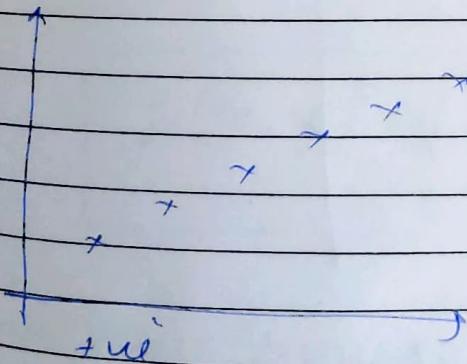
0 covariance \rightarrow No relation b/w $X \& Y$



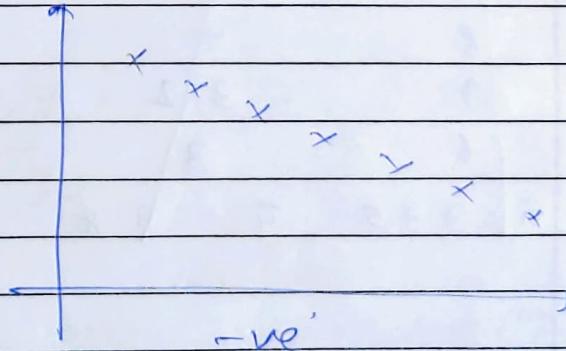
+ve covariance



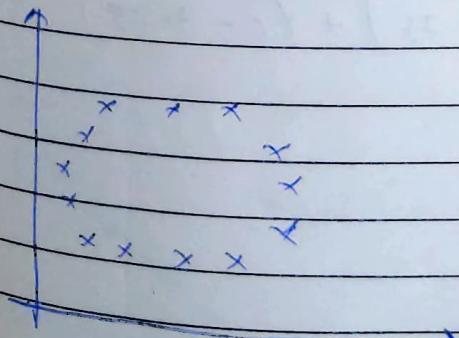
-ve covariance



0 covariance



-ve



Zeros covariance

$$\text{cov}(x, y) = (12 - 15)(10 - 5) + (13 - 15)(11 - 5)$$

$$\begin{aligned} \text{cov}(x, y) &= (12 - 15)(40 - 51) + (13 - 15)(45 - 51) + (12 - 15)(60 - 51) \\ &\quad + (18 - 15)(61 - 51) \end{aligned}$$

5-1

$$\text{cov}(x, y) = \frac{(-3)(11) + (-2)(-6) + (2)(9) + (3)(10)}{4}$$

$$= \frac{-33 + 12 + 8 + 30}{4} = \frac{60 - 33}{4}$$

$$= \frac{27}{4} = + \underline{\underline{6.75}} \quad (\text{true co-var. value})$$

* Another dataset

x	y	y
10	5	4
8	4	6
7	3.2	8
6	3	10
$\bar{x} = 7.75$	$\bar{y} = 3.8$	$\bar{y} = 7$

$$(10 - 7.75) + (8 - 7.75) + (7 - 7.75) + (6 - 7.75)$$

$\approx 10 - 7.5$

$$(10-7.75)(4-7) + (8-7.75)(6-7) + (7-7.75)(8-7) + \\ (6-7.75)(10-7)$$

3

-3.25

value will be negative. can be observed from fig
egno above

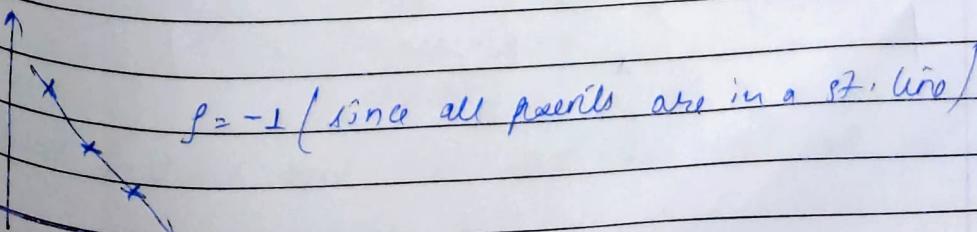
Pearson correlation coefficient (-1 to 1)

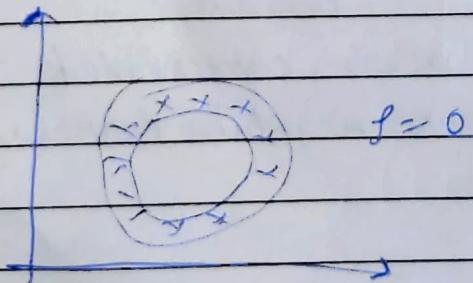
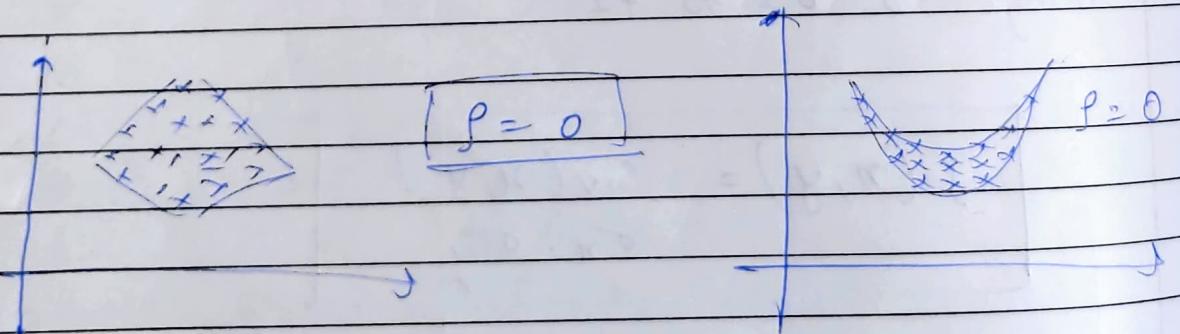
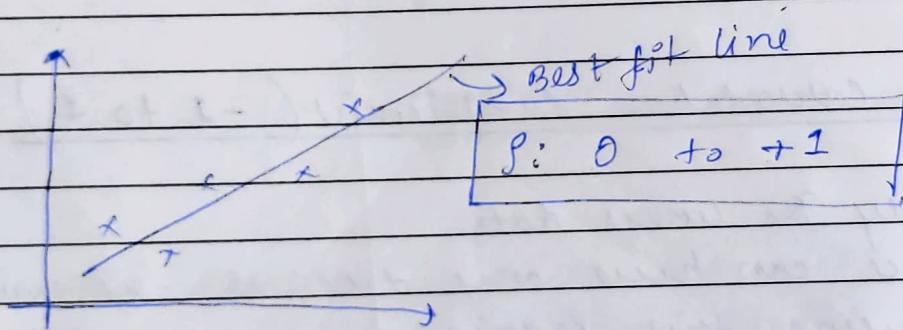
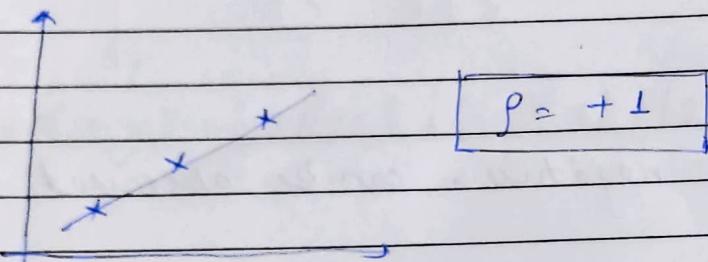
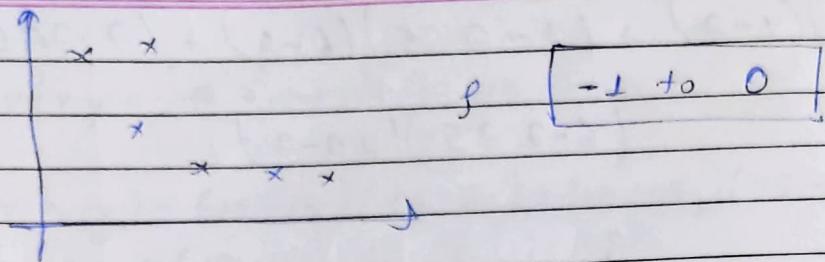
- holds only for linear data
- covariance can have any +ve or -ve values
but by using this Pearson we can define the range to -1 to +1

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

positive

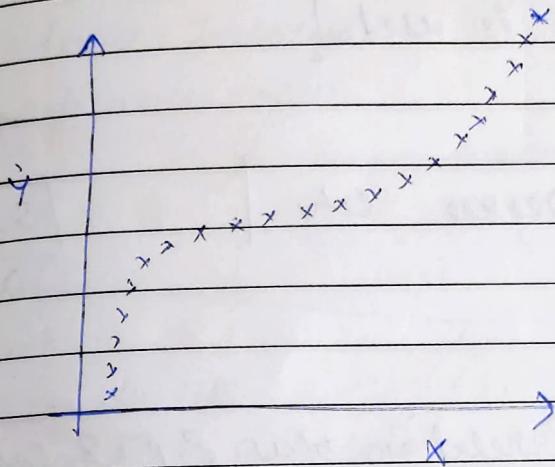
more value towards +1 \Rightarrow more correlated it is
" " " " -1 \Rightarrow more negative correlated it is





Spearman Rank Correlation

For non linear data



$$\rho_S = \text{Corr}$$

$$\rho_S = \frac{\text{Cov}(R(X), R(Y))}{\sigma_{R(X)} \sigma_{R(Y)}}$$

X	Y	R(X)	R(Y)
10	4	4	1
8	6	3	2
7	8	2	3
6	16	1	4

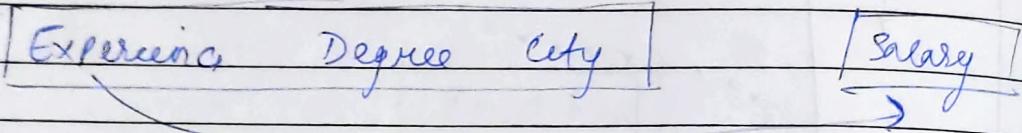
Rank is found
in ascending
order

After finding out the rank we will use only $R(X)$ & $R(Y)$

→ If two values are same then same rank will be given

Q) Why correlation is used?

i/p o/p



We will see correlation b/w Exp & Salary

\downarrow
This can be
true or -ve

Degree & Salary
city & Salary

And let's say we have one more input parameter A which gives same correlation as Experience, then we can drop A parameter.

X ~~A~~ Exp. Degree city Salary
Drop this

np.cov(df['total_bill'], df['tip'])

O/P:- covar([[79.25 8.32], [8.32 3.3]])

covariance of
total bill

covariance of
tip