Day 4 - STATS

1. Central Limit Theorem. ✓
2. Probability. ✓
3. Permutation And combination ✓
4. Covariance, Pearson Correlation, Spearman Rank Correlation. ✓
5. Bernoulli's Distribution
6. Binomial Distribution
7. Power Law (Pareto Distribution).

$\boxed{n<30}$

Size of sample

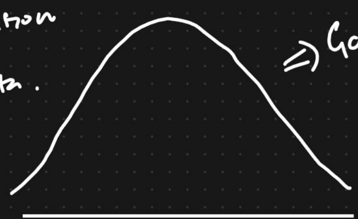
The larger the value the better

No. of samples

1. { Central Limit Theorem }

$\Rightarrow \boxed{n \geq 30}$ $\boxed{n}$ $\rightarrow \boxed{m}$

population data.



$\Rightarrow$ Gaussian / Normal Distribution

μ, σ

$\boxed{N}$

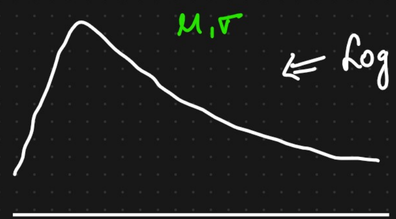
$\rightarrow S_1 \rightarrow \{ x_1, x_2, x_3 \, - - - \, x_n \} \rightarrow \bar{x}_1 = \bar{S}_1$

$\rightarrow S_2 \rightarrow \{ x_3, x_4 - x_1, - - - \, x_n \} \rightarrow \bar{x}_2 \quad \bar{S}_2$

$\rightarrow S_3 \rightarrow \{ x_4, x_1, - - - \, x_n \} \rightarrow \bar{x}_3 \quad \bar{S}_3$

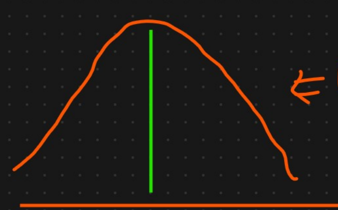
$\bar{x}_m \quad \bar{S}_m$

μ, σ



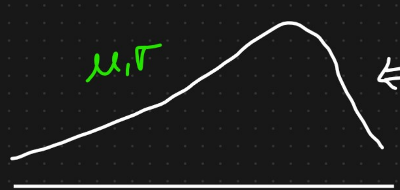
$\Leftarrow$ Log Normal Distribution =

Sampling with replacement



$\Leftarrow$ Gaussian Distribution

μ, σ



$\Leftarrow$ Left Skewed

10 different region $\boxed{n \geq 30}$

$\hookrightarrow$ Assumptions

Size of shark through out the world?

$\boxed{n < 30}$ $m\uparrow\uparrow\uparrow$

② **Probability** : Probability is a measure of the likelihood of an event

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

$\Downarrow$

SHOLAY $\longrightarrow$ COIN $P(H) = 1$

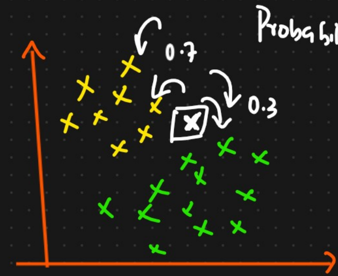
$\Downarrow$

unfair coin

Strong Basic

$\Uparrow$

Probabilities

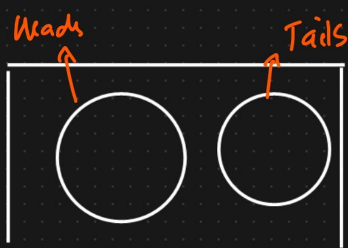


Rolling a Dice $P(1) = \frac{1}{6}$ $P(2) = \frac{1}{6}$ $P(3) = \frac{1}{6}$

① Mutual Exclusive Event

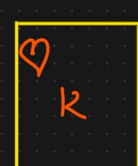
Two Events are mutually exclusive if they cannot occur at the same time

① Tossing a coin ② Rolling c dice



② Non Mutual Exclusive Events

Two events can occur at the same time.



Bag of Marbles

① Picking randomly a card from a deck of cards, two events "heart" and "king" can be selected.

Mutual Exclusive Event

① What is the probability of coin landing on heads Or tails

⇓

Addition Rule for mutual exclusive events

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

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

② What is the probability of getting 1 or 6 or 3 while rolling a dice?

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

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

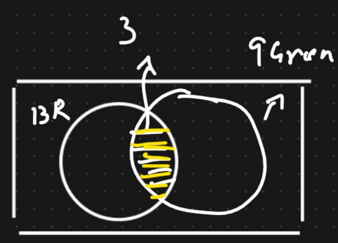
Non Mutual Exclusive Event

Bag of Marbles : 10 Red, 6 Green, 3 (R&G) (R/G)

① When picking randomly from a bag of marbles what is the probability of choosing a marble that is red or green?

⇓

Non mutual Exclusive



Addition Rule for Non Mutual Exclusive Event

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

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

Deck of cards → What is the probability of choosing $\boxed{\heartsuit \text{ or Queen}}$

$$P(\heartsuit \text{ or Queen}) = P(\heartsuit) + P(\text{Queen}) - P(\heartsuit \text{ and Queen})$$

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

(*) Multiplication Rule

(*) Dependents Events : Two events are dependent if they affect one another

Bag of marble $\left\{ \begin{matrix} o\ o\ o\ \cancel{o} \\ o\ o\ o \end{matrix} \right\}$

$\Rightarrow P(W) = \frac{4}{7} \longrightarrow P(Y) = \frac{3}{6}$

white
1 marble

(*) What is the probability of rolling a "5" and then a "3" with a normal 6 sided dice?

{Independent Dependent}

Ans) $P(1) = \frac{1}{6}$ $P(2) = \frac{1}{6}$ $P(3) = \frac{1}{6}$ $P(4) = \frac{1}{6}$

Multiplication Rule for Independent Events

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

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

$P(A \text{ or } B) \Rightarrow$
- Mutual Exclusive
- Non Mutual Exclusive

common

$P(K \text{ or } \heartsuit)$

$P(A \text{ or } B) = P(A) + P(B) - \boxed{P(A \text{ and } B)} \rightarrow$ Non Mutual Exclusive.

$P(A \text{ or } B) = P(A) + P(B)$ [Mutual Exclusive]

Sample $\{H, T\}$

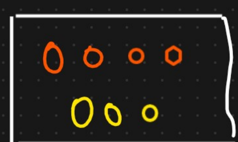
Dependent and Independent Events

Tossing a coin

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

Event A, Event B

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

② 

$\Rightarrow$ Dependent Events

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

Ans) 

$P(O) = \boxed{\frac{4}{7}}$

Orange Marble

$\boxed{P(Y/O)} \rightarrow$ Conditional probability

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

Naive Bayes

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

$= \frac{4}{7} * \frac{3}{6} = \frac{12}{42} = \frac{2}{7}$

Permutation

School of Children

{Dairy Milk, Kit Kat, Milky Bar, Sneakers, 5 star}

$\underline{5} \times \underline{4} \times \underline{3}$

$= \boxed{60 \text{ ways}} \Rightarrow$ Permutation

With permutation, order matters

{ DM KK MB } { } { }

{ KK DM MB } ✓ { } { }

{ } { } { }

Possible Arrangements

$n=5$

n = Total No. of objects

r = # of selection

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

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

Combination

Repetition will not occur

{ DM KK MB }

× { MB KK DM } ←

Unique Combination

$$^{n}C_{r} = \frac{n!}{r!(n-r)!} = \frac{5!}{3!(2)!} = \frac{5 \times \overset{2}{\cancel{4}} \times \cancel{3 \times 2} \times 1}{\cancel{3!} \times \cancel{2}}$$

$= 10$

DREAM 11 ⇐

Covariance

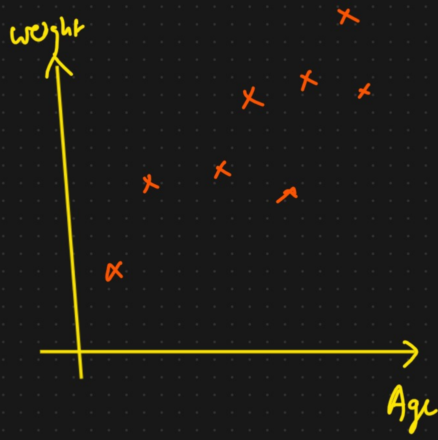
{Feature Selection}

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

Age ↑	Weight ↑
Age ↓	Weight ↓

⇑

Quantify the relationship x & y using mathematical question



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

⇓

$\boxed{24} \Rightarrow \underline{+ve}$

$Cov(x,x)$ ⇐

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

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

$$\boxed{Cov(x,x) = Var(x)}$$ ⇐

+ve Covariance

X↑	y↑
X↓	y↓

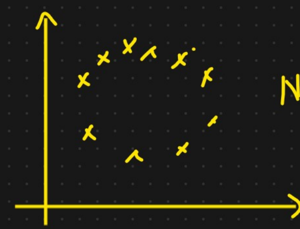
-ve Covariance

X↑	y↓
X↓	y↑

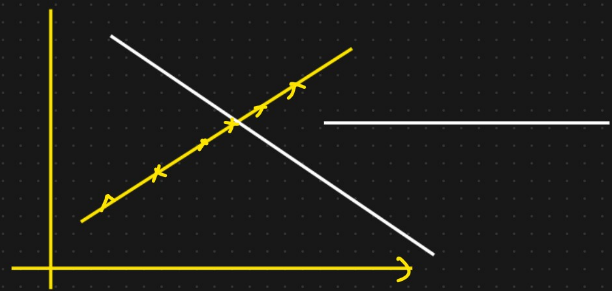
Covariance = 0 $\boxed{\text{No relation with x \& y}}$

⇒ +ve.

$\Rightarrow$ -ve Covariance



No relation.



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

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

$= [(10-7.75)(4-7) + (8-7.75)*(6-7) + (7-7.75)*(8-7) + (6-7.75)*(10-7)$

$= \underline{\underline{-3.25}}$

$\boxed{\begin{matrix} x\uparrow & y\downarrow \\ x\downarrow & y\uparrow \end{matrix}}$

② Pearson Correlation Coefficient (-1 to 1)

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

More the value towards +1

More +ve correlated it is

-1

negative correlated

Scale

Covariance = +ve

-ve

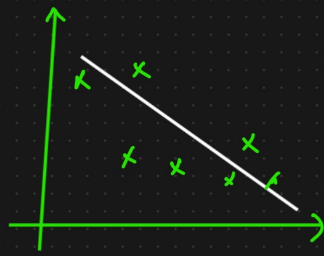
+ve

-ve

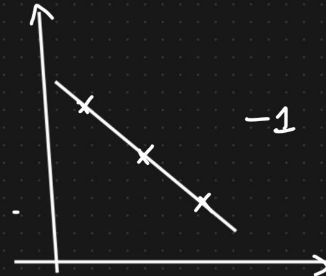
X Y Z

X Y Z $\Rightarrow \boxed{0.7}$ ✓

$\boxed{0.5}$



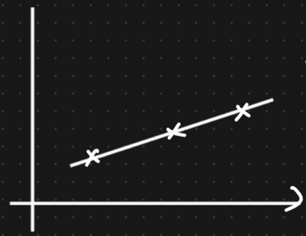
-1 to 0



-1

$x\downarrow\ y\uparrow$

$x\uparrow\ y\downarrow$



$+1$

0 to 1

(4) Spearman Rank Correlation

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

Ascending order

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

Ascending

Spearman Rank Correlation

Why this correlation will be used?

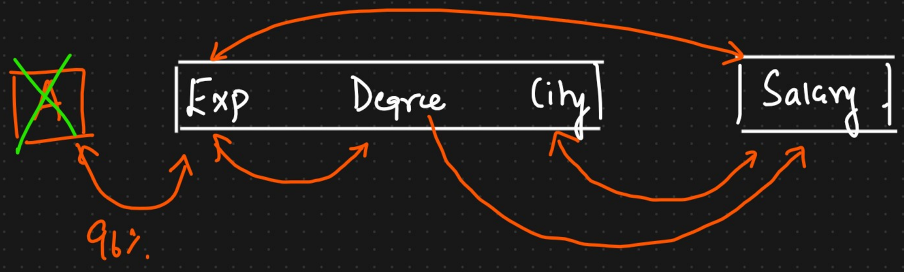
0.95

95%

+ve

-ve } Good

0.2 0.01



+ve Correlated

+ve Correlated

+ve Correlated