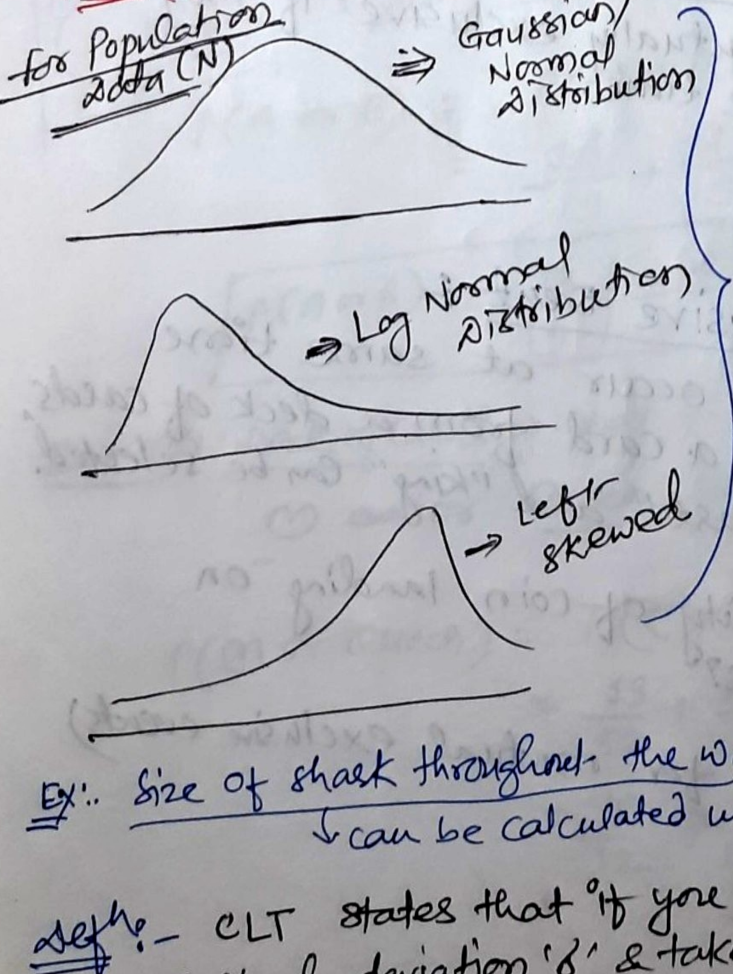
11th sept '22 | DAY #4 | FSDS ~~Bootcamp~~ | Statistics

- **Agenda:**
  1. Central Limit Theorem
  2. Probability
  3. Permutation & Combination
  4. Covariance, pearson correlation, spearman Rank correlation.
  5. Bernoulis Distribution.
  6. Binomial Distribution.
  7. Power law Distribution.

#1) Central Limit Theorem (CLT) :-



⇒ $\boxed{n \geq 30}$ $\boxed{n}$ ← Size of sample $\boxed{m}$ ← No. of samples ← larger the value Better the Result

$\rightarrow \bar{x}_1 = \bar{S}_1$

$S_1 \Rightarrow \{x_1 x_2 x_3 \cdots x_n\} \rightarrow \bar{x}_2 = \bar{S}_2$

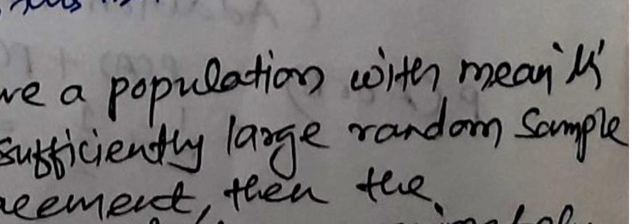
$S_2 \rightarrow \{x_3, x_4, x_5 \cdots x_n\} \rightarrow \bar{x}_3 = \bar{S}_3$

$S_3 \rightarrow \{x_4, x_5, \cdots x_n\}$

"Sampling with replacement"

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

⇒



Ex:. Size of shark throughout the world?
↓ can be calculated using this method

defn:- CLT states that if you have a population with mean 'μ' & stand. deviation 'σ' & take sufficiently large random sample from the population with replacement, then the distribution of the sample means will be approximately normaly distributed.

#2) Probability :-

- It is a measure of the likelihood of an event.

Eg: ① Tossing of fair coin

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

In sholay movie
⇓
unfair coin tossing
⇓
As $P(H) = 1$

② Rolling a dice :

$P(1) = 1/6$
$P(2) = 1/6$
⋮
$P(6) = 1/6$ ... etc.

* ① Mutual Exclusive Event :-

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

Ex:- ① Tossing a coin
② Rolling a dice.

* ② Non-Mutual Exclusive Event :-

- Two events can occur at same time.

Ex:- ① Picking Randomly a card from a deck of cards, two events "heart" and "king" can be selected.

Mutual Excl. Event Problems :

Que ① : What is probability of coin landing on Heads or tails?

⇓

(Addition Rule for mutual exclusive events)

$$P(A \text{ or } B) = P(A) + P(B)$$
$$= \frac{1}{2} + \frac{1}{2} = 1$$

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

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

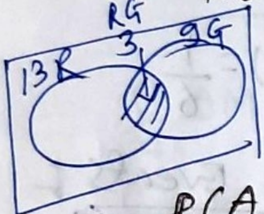
$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

Non-mutual Exclusive Event Problems:

Que.① Bag of marble = 10 Red, 6 Green, 3 (R&G)
- when picking Randomly from bag of marbles what is the probability of choosing a marble that is red or green?

⇓

(Non-mutual Excl. Event, Addition Rule)



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

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

$$\boxed{P(A \text{ or } B) = \frac{19}{19} = 1}$$

Que.② Deck of cards → what is probability of choosing ♡ or Queen cards

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

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

$$= \frac{16}{52} \Rightarrow \frac{4}{13}$$

* Multiplication Rule :-

① **Dependent Events** :- Two events are dependent, if they affect one another.

[Que: what is probability of drawing a black marble & then drawing a Red marble from a bag?]

Ex:- Bag of marble $\left\{ \begin{matrix} 0\ 0\ 0\ \cancel{0} \\ 0\ 0\ 0 \end{matrix} \right\}$

Red marble

Sol^n:- $P(B) = \frac{4}{7}$ $\xrightarrow[\text{removed}]{\text{as 1 Blacked}}$ $P(R) = \frac{3}{6}$ { conditional probability

1 Black marble

(Here 1 event is affecting other event)

$\therefore P(B \text{ and } R) = P(B) \times P(R)$

$= \frac{\cancel{4}}{7} \times \frac{3}{\cancel{6}} \Rightarrow \frac{2}{7}$

② **Independent Events** :-

problem statement :

Que ① what is the probability of rolling a "5" and then a "3" with a normal six sided dice?

Sol^n :- $P(1) = \frac{1}{6}$, $P(2) = \frac{1}{6}$, $P(3) = \frac{1}{6}$ .... $P(6) = \frac{1}{6}$

$\therefore$ **Multiplication Rule for Independent Events** :-

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

$$= P(5) \times P(3)$$

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

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

$\boxed{\begin{matrix} \text{and} \Rightarrow \times \\ \text{or} \Rightarrow + \end{matrix}}$

#3). Permutation :-

Ex: School children entered into Chocolate factory
⇒ {Dairy milk, Kit kat, Milky bar, Sneakers, 5 star}

they have been asked to write first 3 chocolates they see.

$$\Rightarrow \underline{5} \times \underline{4} \times \underline{3} = \underline{60 \text{ ways}}$$

possible.

↳ permutation

with permutation order matters.

{DM. KK MB} { } { }
{KK DM MB} { } { }
{ } { } { } ...etc

possible permutations -

Formula $\boxed{{}^{n}P_{r} = \frac{n!}{(n-r)!}}$, where n → Total no. of subjects
r → no. of selections

∴ using formula [(∵ n = 5) & (r = 3)]

$$\therefore {}^{5}P_{3} = \frac{5!}{(5-3)!} = \frac{5\times4\times3\times\cancel{2!}}{\cancel{2!}} = 60$$

#4). Combination :-

- Repeatation will not occur.

{DM, KK MB} —only unique combination, will be taken

× {MB KK DM} ← no repeatation.

Formula: $\boxed{{}^{n}C_{r} = \frac{n!}{r!\,(n-r)!}}$

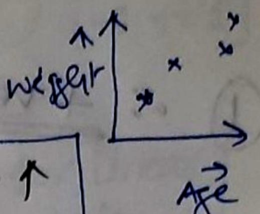
n → Total no. of subjects
r → no. of selections.

$${}^{5}C_{3} = \frac{5!}{3! \times (5-3)!} = \frac{5 \times 4 \times 3!}{3! \times 2!} = \frac{20}{2}$$

Combination → $\boxed{{}^{5}C_{3} = 10}$

Ex: - DREAM 11. → selection.

#5] 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 betn X & Y using mathematical Equations.

Formula :-

Covariance → $$\boxed{\text{cov}(X, Y) = \frac{\sum (x_i - \bar{x}) \times (y_i - \bar{y})}{n-1}}$$

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

$$\therefore \text{cov}(X, X)$$

$$\therefore \boxed{\text{cov}(X, X) = \text{Var}(X)}$$

Sol^n: $$\text{cov}(x, y) = \frac{(12-15)\times(40-51) + (13-15)(45-51) + (15-15)(48-51) + (17-15)(60-51) + (18-15)(62-51)}{5-1}$$

$$= \frac{(-3)(-11) + (-2)(-6) + (0)(-3) + (2)(9) + (3)(11)}{4}$$

$$\Rightarrow 24 = (33 + 12 + 0 + 18 + 33)/4 \Rightarrow 96/4 \Rightarrow 24$$

P.T.O.

① +ve covariance

X ↑	Y ↑
X ↓	Y ↓

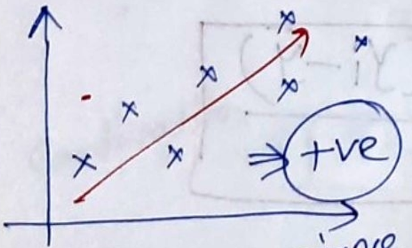
③ when Covariance = 0

⇓

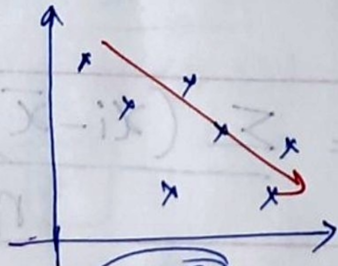
No relationship betn X & Y.

② -ve covariance

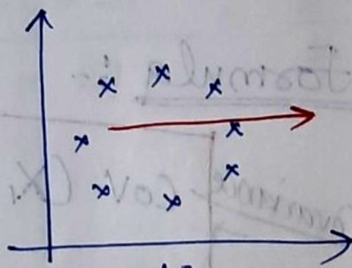
X ↑	Y ↓
X ↓	Y ↑



+ve covariance



-ve covariance



No Relation.

Ex. ② for -ve covariance

X	Y
10	4
8	6
7	8
6	10
$\bar{x}=7.75$	$\bar{y}=7$

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

$$= \frac{(10-7.75)(4-7)+(8-7.75)(6-7)+(7-7.75)(8-7)+(6-7.75)(10-7)}{4-1}$$

$$= \frac{(2.25)(-3)+(0.25)(-1)+(-0.75)(1)+(1.75)(3)}{3}$$

$$= -6.75-0.25-0.75+5.25/3$$

$$= -2.5/3$$

$$\boxed{\text{cov}(x,y) = -0.833} \rightarrow -\text{ve}$$

#6) Pearson Correlation Coefficient :- (ρ) [-1 to 1] (Row)

Good for linear data only

Formula :-

$$\boxed{\rho(x,y) = \frac{cov(x,y)}{\sigma x . \sigma y}}$$

where, $\rho \rightarrow$ pearson correlation coefficient
cov $\rightarrow$ covarience
$\sigma \rightarrow$ standard deviation.

→ More the value of 'ρ' towards '+1', More '+ve' correlated it is.

→ More the value of 'ρ' towards '−1' More '−ve' correlated it is.

for Non-linear data, we use

#7) Spearman's rank Correlation Coefficient :- [-1 to 1]

Formula :-

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

where R → Rank

Only Ascending order

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

Bernouli's Distribution :-

- ① outcomes are binary — Ex. - Tossing coin: $\underline{H}$ or $\underline{T}$.

$P = P(H) = 0.5$, $q = P(T) = 0.5$

$\boxed{P = 1 - q}$ $\boxed{q = 1 - P}$

— Bernouli Distribution describes events having exactly two outcomes i.e. it represents the success or failure of single Bernouli trial.

Binomial Distribution :-

Exp#1 ↓ Binary outcomes

Exp#2 ↓ B.O

Exp#3 ↓ B.O.

Exp#4 ↓ B.O.

Exp#5 ↓ B.O.

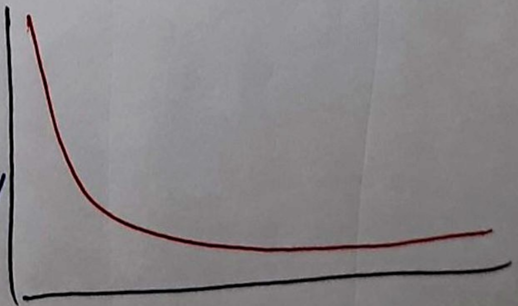
Ex. Tossing a coins.

→ The Binomial Distribution represents the number of successes and failures in 'n' independent Bernouli trials for some given value of 'n'.

Power law :-

(The 80-20 Rule)

- A relative change in the ~~other~~ one quantity results in proportional relative change in the other quantity i.e. one quantity varies as power of another.



Annova Test (F-Test)