to find the area of 1.75%.

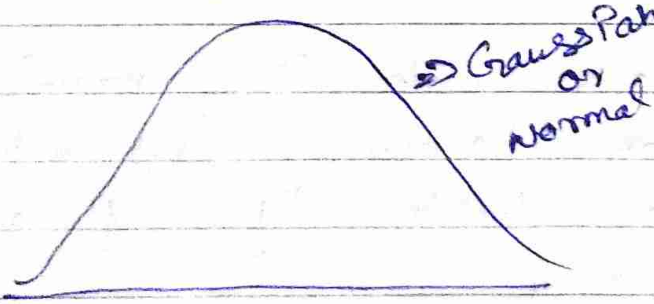
area of 1.75% is = 85%

Day 4

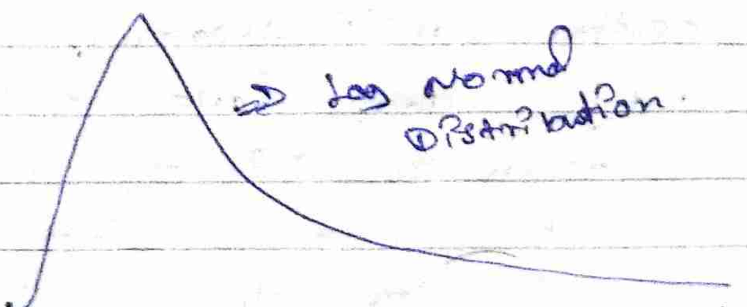
Topics

1.) Central Limit Theorem
2.) Probability
3.) Permutations & combinations
4.) Co-variance, pearson correlation, spearman Rank correlation
5.) Bernoulli's Distribution
6.) Binomial Distribution
7.) Power law {pareto Distribution}

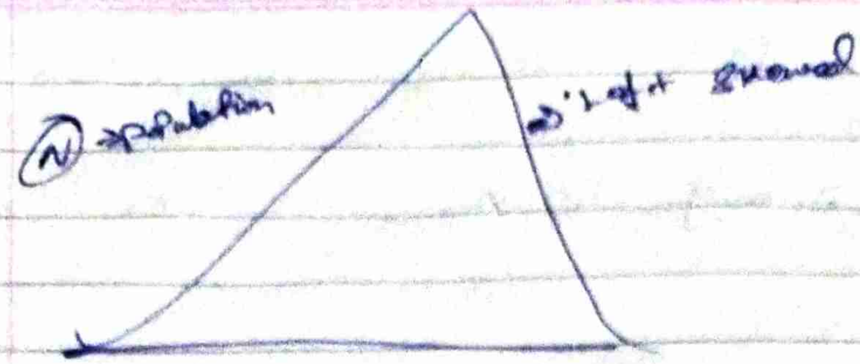
1 Central Limit Theorem



N → Population



N → population



* In any distribution when n is ≥ 30 for population it's fall in Gaussian distribution.

n → size of sample
m → No. of samples

$S_1 \rightarrow \{x_1, x_2, x_3 \cdots x_n\} \rightarrow \bar{x}_1$
$S_2 \rightarrow \{x_3, x_4, x_5 \cdots x_n\} \rightarrow \bar{x}_2$
$S_3 \rightarrow \{x_5, x_6 \cdots x_n\} \rightarrow \bar{x}_3$
$\vdots$
$S_n \quad \{x_n, \cdots x_{n+1}\} \rightarrow \bar{x}_n$

* If our data is normal distribution or not normal distribution that can be decide by if $n \geq 30$

def
* The central limit theorem says that if you have a population with mean μ & sd σ and take sufficiently large random samples ($n \geq 30$) from the population with replacement, then the distribution of the sample means will be approximately normally distributed.

* If $n < 30$ is not normally distributed

2. Probability

* Probability is a measure of the likelihood of an event

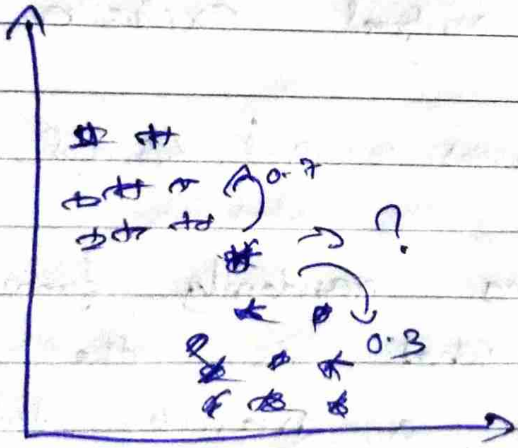
Eg: Tossing a coin

$$P(H) = \frac{1}{2} = 0.5$$

$$P(T) = \frac{1}{2} = 0.5$$

Eg: Rolling a Die

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



1] Mutual exclusive event

* Two events are mutually exclusive if they can't occur on the same time.

Eg: Tossing

* H, T can't occur on same time.

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

2.) Non - Mutual Exclusive Event

Two events can occur at the same time.

Eg: Picking randomly a card from a deck of cards, two events heart & King can be selected.

Examples of Mutual event

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

$$P(A \text{ or } B \text{ or } C) = \overset{1}{P(A)} + \overset{6}{P(B)} + \overset{3}{P(3)}$$

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

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

Example Non Mutual exclusive event

Ex: Bag of mables : 10 Red, 6 Green, 3 Red & Green

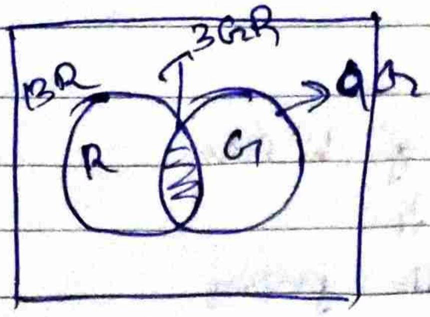
q: When picking randomly from a bag of marbles what is the probability of choosing a marble that is red or green?

Addition Rule for Non mutual

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

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

= [illegible]



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

$$= \frac{22-3}{19} = \frac{19}{19} = 1$$

5.) Multiplication Rule

i.) Dependent Events
ii.) Independent "

i.) Dependent event

* Two event are dependent if they affect one another.

Ex: Bag of marble = { □ □ □ ☒ ○ ○ ○ }

Event 1
$P(S) = \frac{4}{7}$

* Here 1 Square was removed so to marble is 6

Event : 2
so $P(C) = 3/6$ → This affected because of previous event
(Circle

III permutations

Eg: school of children
↓
chocolate factory
{DM, KK, MB, SNKS, 5Star}

* go and ask the student write the first 3 choclate name what ever you are going to see

$$\underset{\text{child 1}}{\underline{\text{5 chances (pick 5star)}}} \times \underset{\text{child 2}}{\underline{\text{4 chance (pick DM)}}} \times \underset{\text{child 3}}{\underline{\text{3 chances (pick KK)}}}$$

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

permutations $= 60$

* All the possible arrangement is called as permutations.

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

n = Total No. of choices
r = # of selections

$$\therefore nP_r = \frac{5!}{(5-3)!} = \frac{5!}{2!} = \frac{1\times2\times3\times4\times5}{2}$$

$$= \frac{24\times5}{2} = \frac{120}{2} = 60$$

$$\boxed{nP_r = 60}$$

Combinations

* Repeatation will not occur

{DM KK MB} → 1st occur select KK and then next occur we will have only unique combinations

formula

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

$$= \frac{5!}{3!(5-3)!}$$

$$= \frac{5!}{3!(2)!}$$

$$= \frac{\cancel{1}+\cancel{2}+\cancel{3}+4+5}{\cancel{1}+\cancel{2}+\cancel{3}+2}$$

$$= \frac{\cancel{20}\,10}{\cancel{2}}$$

$$\boxed{nC_r = 10}$$

4) Co-Variance

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

Age ↑ weight ↑

Age ↓ weight ↓

Spearman Rank Correlation

* Pearson Correlation can only consider linear data (increase straight line)

* Spearman Rank Correlation it can consider the non-linear data also

formula

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

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

∴ R(x) = Rank (x)

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

$\cancel{\sigma(R(x))} =$

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

$$\sigma(R(x)) = \sqrt{\frac{(R(x) - \overline{R(x)})}{n-1}}$$

Why this correlation will be used?

X ⊗ ? o/p

→ 0.2

0.95

95%

Expe ~~Degree~~ City Salary → +ve correlated

+ve correlated

unable to correlated

based on city

+ve correlated

95%

mostly correlated

so I can drop any one feature