Stats Day 4.

1. Continuation of Z-Score
2. Central limit Theorem
3. Probility
4. Permutation and Combination
5. Covariance
6. Pearson and Spearman Ran Correlation

* Continuation of Z-Score.

1. Z - Score

$x = \{1, 2, 3, 4, 5, 6, 7\}$, let Consider the mean = 4 and Standard deviation =

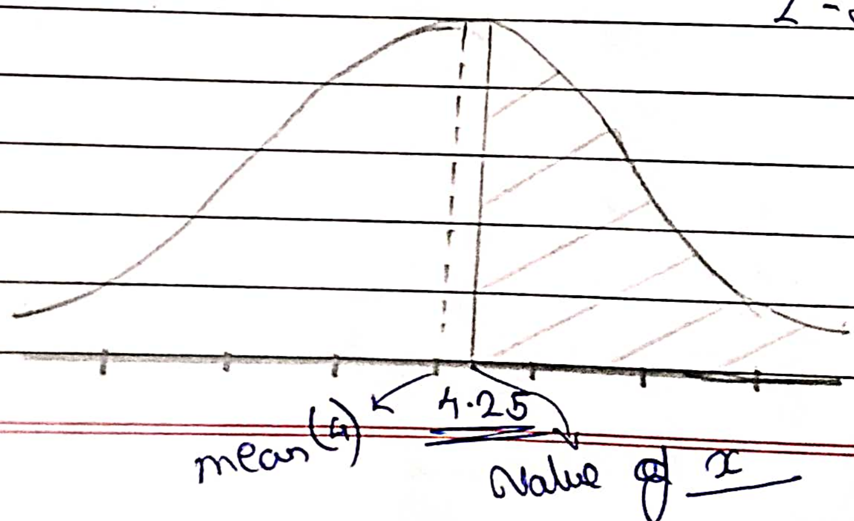
$\boxed{\bar{x} = 4.25}$

$$Z\text{-Score} = \frac{x_i - \mu}{\sigma/\sqrt{n}}$$

$\sigma/\sqrt{n} = 1$

$$= \frac{4.25 - 4}{1}$$

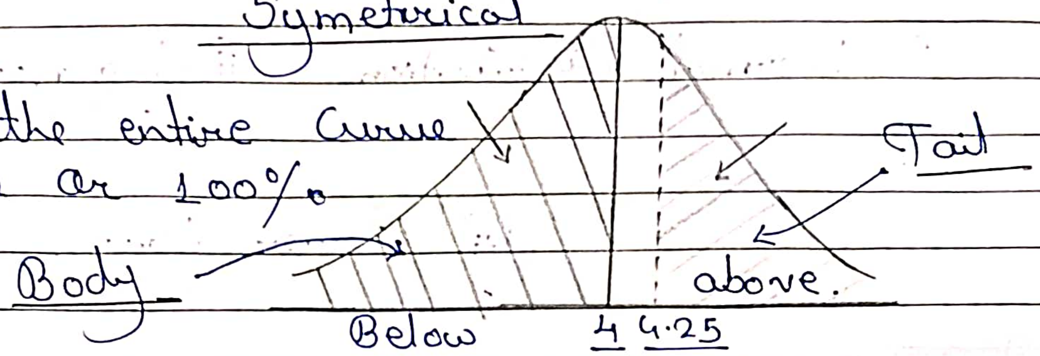
$$= 0.25$$



Q. What Percentage of Score falls above 4.25?

Symetrical

area of the entire Curve = 1 or 100%



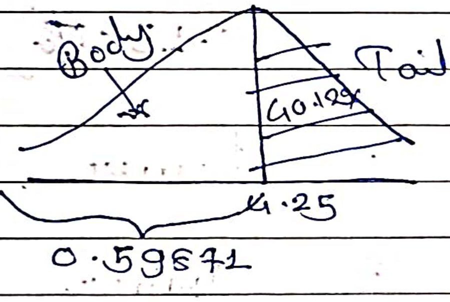
If we subtract entire area with Tail It will result, called **Body**

Formula = 1 - area of Body = Area under the Tail

As here our distance between 4 and 4.25 was 0.25, with the help of Z-score table we find the value of 0.25 is 0.59871.

So with Z-score table

1 - 0.59871 = 0.40129



So we can say that 40.129% score fall above 4.25

Statistics

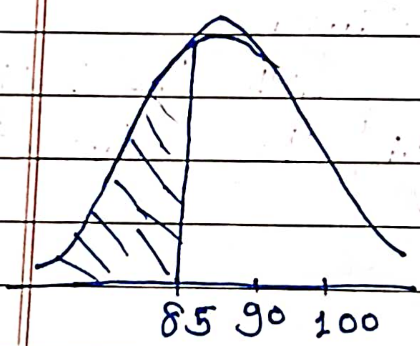
* Z-Score to know how many standard deviation away from the mean a specific variable is there.

(i) Normal Distribution (ii) log normal Distribution

(iii) Power law.

~~Agenda~~

Q. In India the average IQ is 100, with a standard deviation of 15, what is percentage of population would you expect to have an IQ lower than 85?



$m = 100$
$\sigma = 15$

$$Z\text{-Score} = \frac{85 - 100}{15}$$

$$= -1$$

(0.1587)

(i) Lower than 85 = 0.1587 — area of Body

(2) Higher than 85 = 84.13 % — $0.1587 - 1 = 0.8413$

(3) Between 85 to 100 = 0.3413 (34.13 %)

$0.5 - 0.1587 = 0.3413$

Page No.
Date

Find the y. of Score between 100 to 125 Standard deviation is 15?

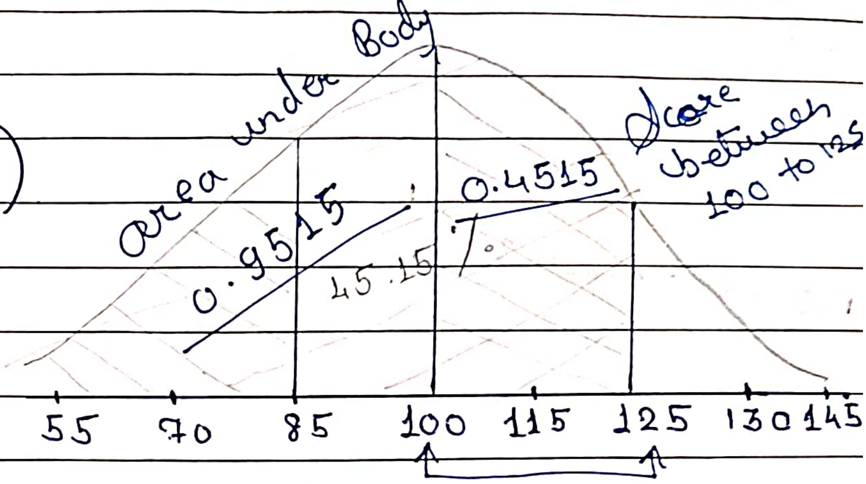
$m = 100$
$\sigma = 15$

$* = \dfrac{125 - 100}{15}$

$= 1.66 \quad (.9515)$

$\therefore 0.9515 - 0.5$

$\therefore 0.4515$



Find the y of Score between 85 to 115

Zscore $: \dfrac{85 - 115}{15} = \dfrac{-30}{15} = -2$

$= -2 \ (0.02275)$.

Area of Body $= 0.02275 - 0.5$

$= -0.47725$

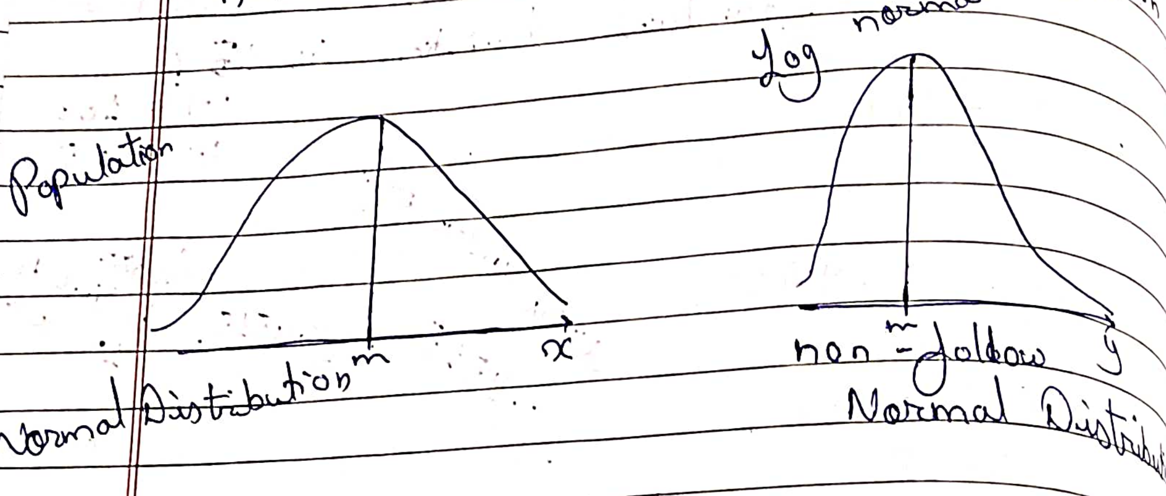
$= 47.725\,\%$

Important Topic

Central Limit Theorem.

Page No.
Date

$n > 30$ = Z Table
$n < 30$ = T Table



~~Age~~

Q. As from the population we Collected Sample where as $n > 30$ and made the average of Sample.

Sample →

eg: $\{x_1, x_2, x_3, \text{------}, x_n\} \rightarrow \bar{x}_1$

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

$\{x_1, x_2 \text{------} x_n\} \rightarrow \bar{x}_m$

As $\{\bar{x}_1, \bar{x}_2, \bar{x}_3, \text{--------}, \bar{x}_m\}$

As If we plot an all this Sample m... we can get an Normal distribution

(If we have or not have Normal distri...

If we take multiple sample mean with $n > 30$ we get the Normal following distribution }.

It is specifically apply for log normal distribution.

Interview Question

Q. Tell the size of the sharks of the entire world ?

→ By making an assumption and applying multiple of sample mean with $n > 30$ we can get a normal following distribution. (Consider)

Inferential Statistics

* Probability *

Meaning :- Probability is a measure of a likeli of an event.

fair
a) Probability (~~measure~~ of Head = 0.5

b) Probability of Tail = 0.5

- $\underbrace{P(H) = 0.5}$ $\underbrace{P(T) = 0.5}$

when there is an Two outcomes this called an Bernoulli's Distributions

eg :- $\underbrace{P(H)}_{p} = 0.5$ $\underbrace{P(T)}_{q} = 0.5$

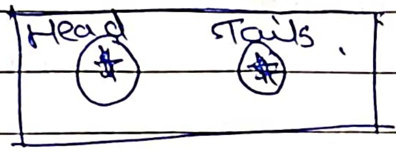
$\therefore q = 1 - p$.

Tosses of coin and always get outcome as (H) it is not an a fair coin toss

(i) mutual Exclusive event

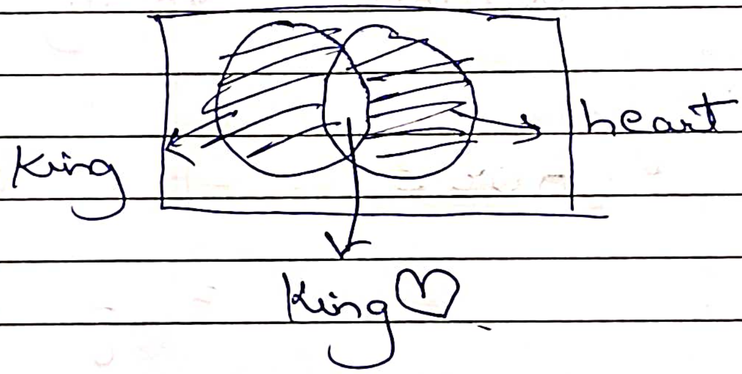
Two events are mutually exclusive if they cannot occur at the same time.

No overlaps happen in mutual exclusive event.



(ii) Non mutual exclusive event.

Picking randomly from a deck of Cards, two event "heart" & "King" are not mutual exclusive event may overlaps in non-mutual exclusive event.



Bag of marbles.   Red, Green, R & G.

Red Or Green

Page No.
Date

* Mutual exclusive event.

1. What is a probability of coin landing on heads or tails.

→ Rules. (Addition Rule)

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

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

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

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

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

Page No.
Date

Non mutual exclesive event

1. Picking up randomly a marble from bag, what is probability of Chossig marbles that is Red or green

Bag of marble = R 10, G = 6, R and G = 3

$P(R) = \frac{10}{19}$ $P(G) = \frac{6}{19}$ $P(R \text{ and } G) = \frac{3}{19}$

Addition Rule for non-mutual.

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$
$$= \frac{10}{19} + \frac{6}{19} - \frac{3}{19}$$

$$= \frac{13}{19}.$$

mutiplication R

Page No.
Date

① Independent event

② Non - Independent event

Independent event : Two events are independent if they do not affect one another.

eg : Tossing a coin.

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

On one not impact another event

eg : Rolling a "five" and then rolling a "three" in a dice.

$\underbrace{P(5) = \frac{1}{6}}_{\text{Independent}}$ and $\underbrace{P(3) = \frac{1}{6}}_{\text{Independent}}$ and $\underbrace{P(4) = \frac{1}{6}}_{\text{Independent}}$

Example

Probability of rolling a "5" and a "3" with a normal six - sided dice ?

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

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

$= \frac{1}{36}$

$A = 5$
$O = 6.$

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

$= \frac{5}{11} \times \frac{6}{1\cancel{0}} = \frac{3\cancel{0}}{11}$

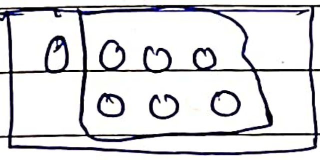
$P(A \text{ and } O) = \frac{3}{11}$

Dependent event : Two event are dependent when they affect one another.

eg : Bag of Colour marbles.

3 Orange
4 yellow } 7. marbles

$P(\text{Orange}) = \frac{3}{6}$ $P(\text{yellow}) = \cancel{6}\ \frac{4}{7}$



$\frac{3}{6}$ Because 1 marble is already remove.

nd event

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

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

$P(y) = \frac{3}{6}.$

$P(O \text{ and } y) = P(O) \times P(y)$

$= \frac{4}{7} \times \frac{3}{6}$

$= \frac{12}{42} = \frac{6}{21} = \frac{2}{7}.$

$\therefore \frac{2}{7}.$ (Probability).

2.02.12

Permutations and Combination

e:- We went to a Zoo where
is numbers of animals are
presented.
(Tiger, Lion, Zebra).

$\underline{6} \times \underline{5} \times \underline{4} = 120$ way or Possibility

{Tiger, Lion, monkey}
{Tiger, Lion, Zebra}
{Lion, money, Tiger}

Permutation mean all the possibility of writing an object.
(Order matters)
That will be Count

Formula :- $^{n}P_{r} = {}^{6}P_{3} = \frac{n!}{(n-r)!}$

n (Total number of object).
R (Total number of we are picking object.

$\frac{6!}{(6-3)!} = \frac{6!}{3!} = \frac{6 \times 5 \times 4 \times \cancel{3!}}{\cancel{3!}}$
$= \underline{\underline{120}}$

① {Krish, youtube channel, Stats, iNeuron, Support, Power BI}.

$\underline{6} \times \underline{5} \times \underline{4} \times \underline{3} \times 2$ (R).

$n = 6$
$R = 5$

$^{n}P_{r} = {}^{6}P_{5} = \frac{6!}{(n-r)!}$

$= \frac{6!}{(6-5)!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times \cancel{1!}}{\cancel{1!}}$

$= \underline{720}$

② {Sun, Rain, moon, Stars, planet, climate, Galaxy}

$\underline{7} \times \underline{6} \times \underline{5} \times \underline{4}$ (R).

$n = 7$
$R = 4$

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

$= {}^{7}P_{4} = \frac{7!}{(7-4)!} = \frac{7 \times 6 \times 5 \times 4 \times \cancel{3!}}{\cancel{3!}}$

$= \underline{\underline{840}}$

*- Combination

(Order not matters)

Lion, Tiger, Zebra } 1
Zebra, Lion, Tiger }
Lion, Zebra, Tiger } 1

- In Combination Repeation not Occurs

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

$n = 6$
$R = 5$

$$= \frac{6!}{5!(6-5)!}$$

$$= \frac{6 \times \cancel{5 \times 4 \times 3 \times 2 \times 1}}{\cancel{5 \times 4 \times 3 \times 2 \times 1}}$$

$$= 6$$

(2) $6 \times 5 \times 4$ (R)

$^{n}C_{R} = \frac{n!}{r!(n-r)!} = {}^{6}C_{3} = \frac{6!}{3!(3!)} = \frac{6 \times 5 \times 4}{3 \times 2}$

$= \frac{120}{6}$

Co-Variance.

Page No.
Date

$X \uparrow$ age $\leftrightarrow$ $Y \uparrow$ weight

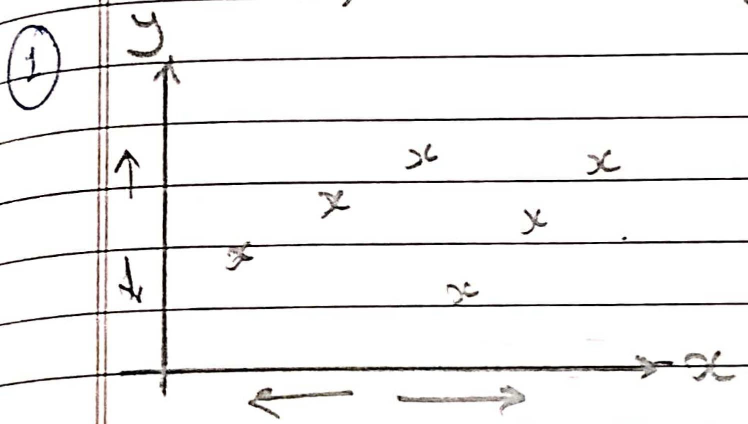
Q mathematically can we find relation between X and y

age (x_i)	weight (y_i)
12	40
15	45
18	50
19	55
20	60

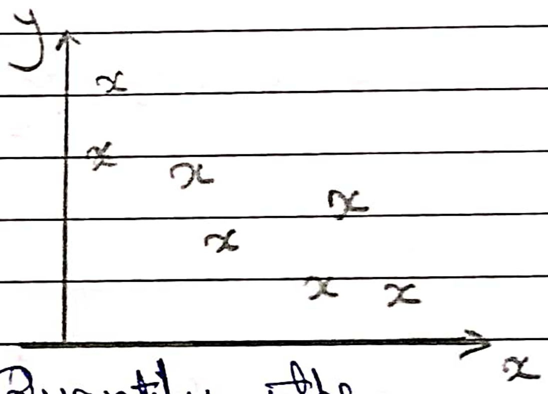
(1) $\begin{Bmatrix} x\uparrow & y\uparrow \\ x\downarrow & y\downarrow \end{Bmatrix}$

(2) $\begin{Bmatrix} x\downarrow & y\uparrow \\ x\uparrow & y\downarrow \end{Bmatrix}$

(1)



(2)



- Scarted graph
- mathematical formula to Quantify the ~~of~~ relationship between x & y ?

To understand this we have an Co-variance.

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

$$\text{Cov}(x, x) = \text{Var}(x) = \frac{\sum (x_i - \bar{x}) \times (x_i - \bar{x})}{N-1}$$

$$= \frac{\sum (x_i - \bar{x})^2}{N-1}$$ ← (Sample Variance)

So the Cov(x, x) is Var (x)

Just, understand the formula.

- Formula of Variance

$$Var(x) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$ So,

$$\boxed{Var(x) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(x_i - \bar{x})}{n-1} \\ \Downarrow \\ Cov(x, x).}$$

Such as ;

$$Cov(x, y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

formula of Co-Variance to understand the relationship between x and y.

* Example.

x	y
1	3
2	4
3	5
6	12

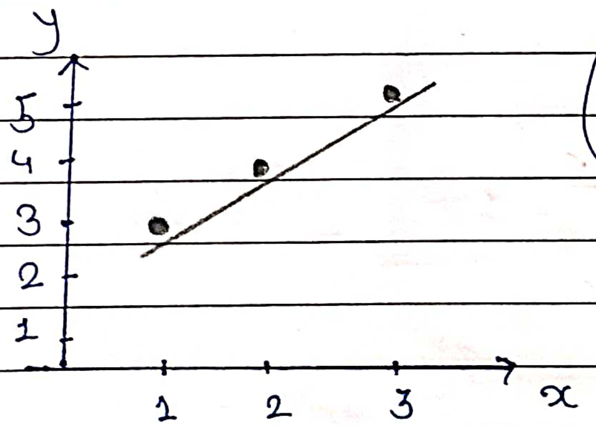
~~Cov (x, y) = [(1-2) (3-4)~~

$\bar{x}$: $6 \div 3$ $= 2$ $\bar{x}$: $12 \div 3$ $= 4$

$$\text{Cov}(x, y) = \frac{[(1-2)(3-4) + (2-2)(4-4) + (3-2)(5-4)]}{2}$$

$$= \frac{1 + 0 + 1}{2} = \frac{\cancel{2}}{\cancel{2}} = 1$$ (Positive value)

$\therefore$ Cov $(x, y) = 1$



(Perfect Positive relation x and y)

$x \uparrow$ $y \uparrow$

*. Example.

x	y
1	3
2	4
3	6
$\bar{x} = 2$	$\bar{x} = 4.33$

$Cov(x, y)$

$$\sum_{i=1}^{n} = \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

$$= \frac{[(1-2)(3-3) + (2-2)(4-4.334) + (3-2)(6}{2}$$

$$= \frac{(-1)(-1.334) + 0 + (1) * (1.66)}{2}$$

$$= \frac{1.334 + 1.66}{2} = \frac{2.994}{2}$$

$= 1.497$ $(Cov\ x, y) = 1.497$ (Positive Co-related)

Disadvantage of Co-variance

No specific limite define.
No specific strength define.

As $\underbrace{Z — x — Y}$

+1000. +500

-1000 -500

no limite and strength define in Co-variance

As,	$x \uparrow$	$y \uparrow$	(Increases)
	$x \uparrow$	$y \downarrow$	(Increases and decreases)
	$x \downarrow$	$y \uparrow$	(decreases and Increases)
	$x \downarrow$	$y \downarrow$	(decreases and decreases).

Covariance is method to Quantify the relationship between two data set.