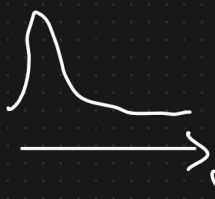
Statistics

Z-score → Normal Distribution, Log Normal Distr, Power law

Agenda

1. Continuation of Z-score {Super perfect}
2. Central Limit Theorem {Interview} ⇒ Solving problem [n≥30]
3. Probability ✓ {Done}
4. Permutation And Combination {Completed}. { 2, 3, 4, 5, 8, 10, 15, 20, 0, -5, -8, 24, 28, 42,
5. Covariance {Computed}
6. Pearson And Spearman Rank Correlation {Tom}



① Z-score =

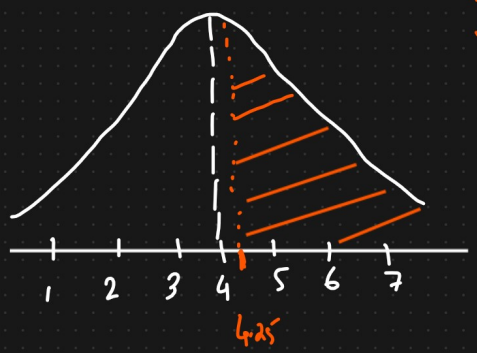
$X = \{1, 2, 3, 4, 5, 6, 7\}$ ⇒ $\mu = 4$ (common)

$\sigma = 1$

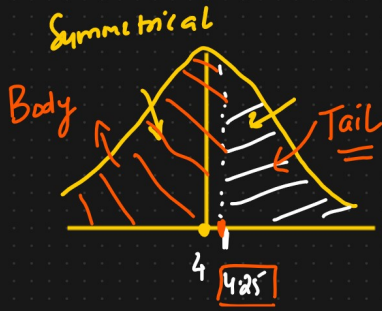
$X = \boxed{4.25}$

$$Z\text{-score} = \frac{x_i - \mu}{\sigma}$$

$$Z\text{-score} = \frac{4.25 - 4}{1} = \boxed{0.25}$$

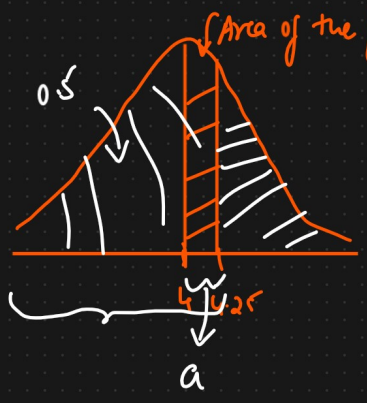


Question: What percentage of score falls above 4.25 ?

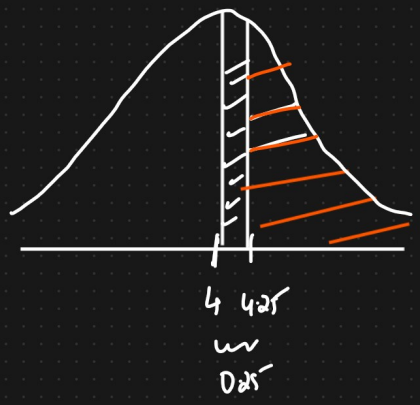


Area of the curve = 1 or 100%.

⇒ 1 − Area of the body = Area of the tail.



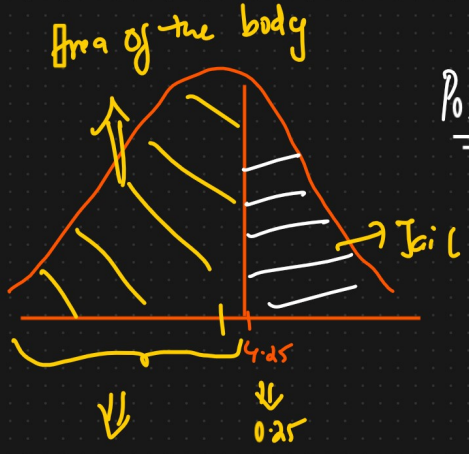
$= 1 - (0.5 + a)$: Area under the tail



$$Z\text{-Score} = \frac{4.25 - \mu}{\sigma} = \frac{4.25 - 4}{1} = 0.25$$

① Z table ⇒ Area under the curve.

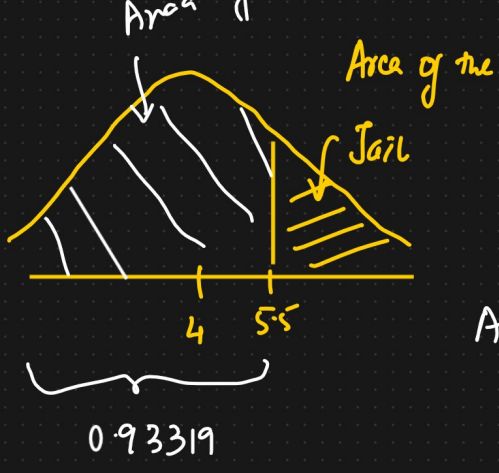
0.59871



Positive Z table

$1 - 0.59871 = 0.40129$

40.129%



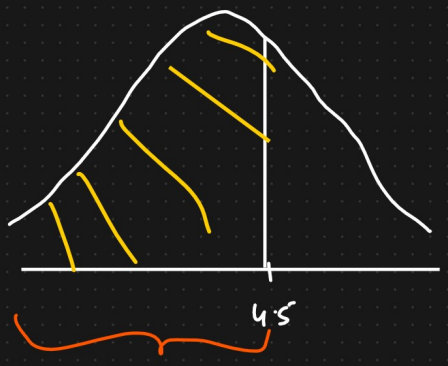
① Find the std using Z-score

$\mu = 4$

$\sigma = 1$

$$Z\text{-Score} = \frac{5.5 - 4}{1} = 1.5$$

Area of the Jail $= 1 - 0.93319 = 0.06681$



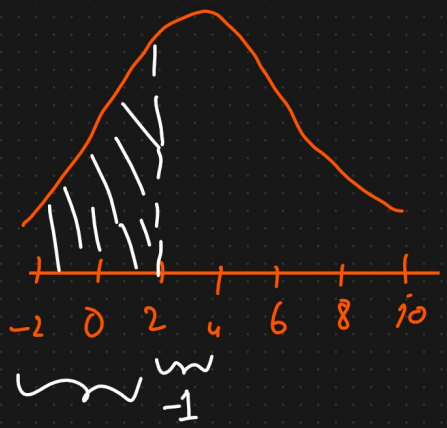
$\mu = 4$

$\sigma = 1$

Percentage of scores less than 4.5?

$$Z\text{-score} = \frac{4.5 - 4}{1} = 0.5$$

0.69146 $\Rightarrow$ Answer = 69.146%



$\mu = 4$

$\sigma = 2$

What is the % of scores less than 2

① $$Z\text{-score} = \frac{2-4}{2} = \frac{-2}{2} = -1$$

⇓

0.15866 = 15.86%

(K) In India the average IQ is 100, with a standard deviation of 15. What is the percentage of population would you expect to have and IQ lower than 85?

$\mu = 100$ $\sigma = 15$

① Lower than 85 = 0.1587

② Higher than 85 = 84.13

③ Between 85 to 100 = 34.13%

% of Score between 100 to 125

⇓

45.15%

$$Z\text{-score} = \frac{85 - 100}{15} = \frac{-15}{15} = -1$$

Area of Body = 0.1587

1 - 0.1587 = 84.13%

$0.5 - 0.1587 = 0.3413$

$\frac{125-100}{15} = \frac{25}{15} = \frac{5}{3}$

$= 0.9515 - 0.5$

$= 0.4515$

$\Downarrow$

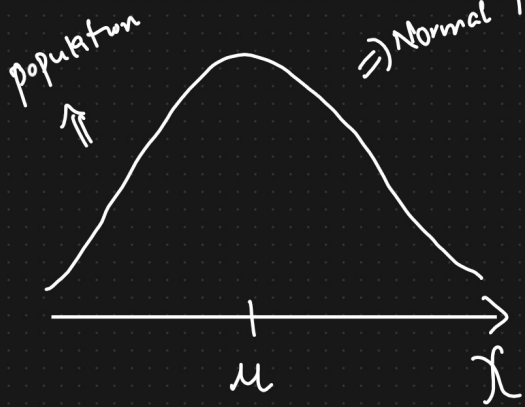
45.15%

$n \geq 30 \Rightarrow$ Z table ✓

$n < 30 \Rightarrow$ t table ✓

$n \leq 30$

② Central limit theorem



$\Rightarrow$ Normal Distribution

$n \geq 30$ — Sample mean

$\{x_1, x_2, \ldots x_n\} \rightarrow \bar{x}_1$

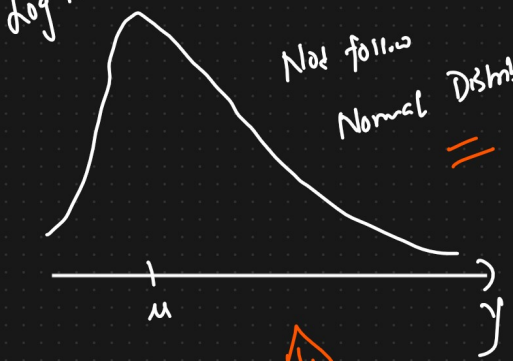
$\{x_1, x_2 \ldots x_n\} \rightarrow \bar{x}_2$

$\Rightarrow \{x_1 \ldots x_n\} \rightarrow \bar{x}_3$

$\{x_n \ldots\} \rightarrow \bar{x}_4$

$\Downarrow$

$\{x_1 \ldots x_n\} \rightarrow \bar{x}_m$



Not follow Normal Distribution $=$

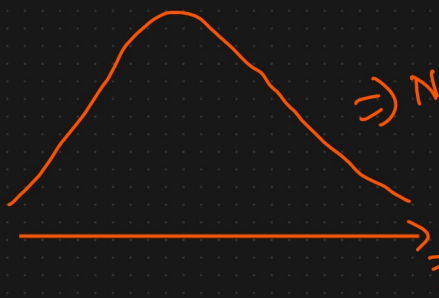
$n \geq 30$

$\Uparrow$

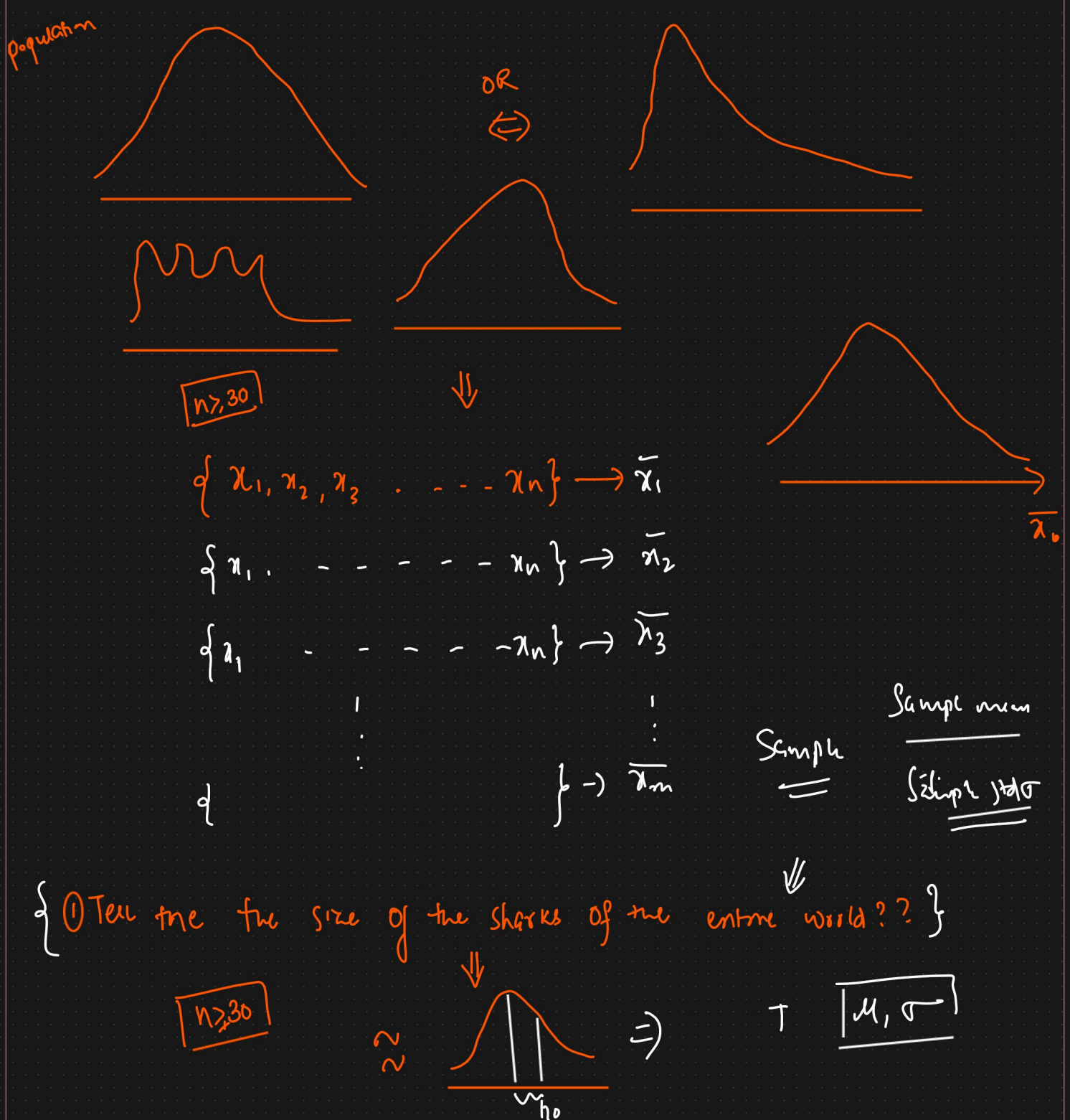
$n \geq 30$

$\{\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots \bar{x}_m\}$

$\Downarrow$



$\Rightarrow$ Normal or Gaussian Distribution



⑭ Probability : Probability is a measure of the likelihood of an event.

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

q = 1 - P

{ Bernoulli's Distribution } ⇒ Two outcomes

SHOLAY ?? ⇒ Bollywood

Tossed (4) ⇒ Biased coin

⇓

Not a fair coin

Rolling a Dice?

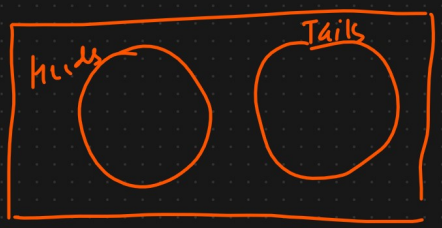
$\{1, 2, 3, 4, 5, 6\}$

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

① Mutual Exclusive Event

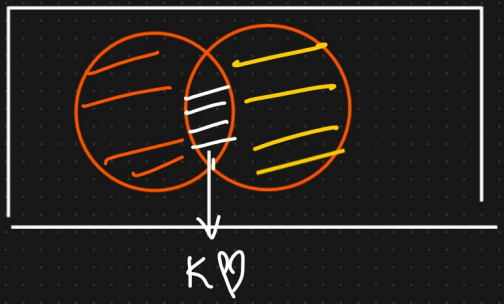
{ Two Events are mutually exclusive if they cannot occur at the same time }

Eg: Tossing a coin ② Rolling a dice



② Non Mutual Exclusive Event

Picking randomly from a deck of cards, two events "heart" & "King" are not mutual exclusive.



K OR ♡ ⇒ K♡

Bag of marbles | Red, Green, Red & Green

Red or GREEN

Addition Rule

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?

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

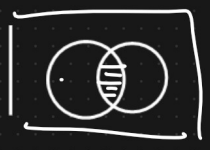
$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = 1/2 //.$$

R/4

Bag of marbles : 10 Red, 6 Green, 3 (Red & Green)

④ When picking randomly from a bag of marbles, what is the probability of choosing a marble that is red or green?

{ Non mutual Event }



$P(R) = \frac{13}{19}$ $P(G) = \frac{9}{19}$ $P(R \text{ and } G) = \frac{3}{19}$.

Addition Rule for Non mutual exclusive Event

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

$$= \frac{13}{19} + \frac{9}{19} - \frac{3}{19} = \frac{22-3}{19} = \frac{19}{19} = 1.$$

(#) Multiplication Rule

① Independent Event : Two events are independent if they do not affect one another

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

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

Eg: Rolling a "five" and then rolling a "3" in a dice.

$Pr(5) = \frac{1}{6}$ and $P(3) = \frac{1}{6}$ $P(4) = \frac{1}{6}$

$Pr(6) = \frac{1}{6}$ $Pr(6) = \frac{1}{6}$ $Pr(6) = \frac{1}{6}$.

② Dependent Events :{Two events are dependent if they affect one another}

Bag of color marbles

⇒ 7

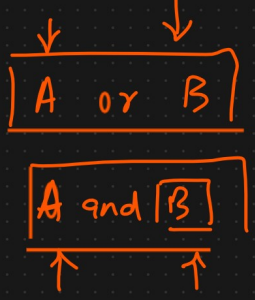
Orange ↓

→ P(Orange) = 4/7 and P(Yellow) = 3/6

Orange

{Independent Event}

* What is the probability of rolling a "5" and then a "3" with a normal six-sided dice?

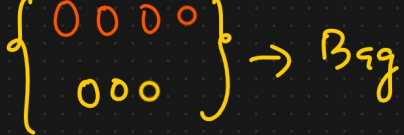


Ans) Independent

Multiplication Rule for Independent Events

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

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



→ Bag

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

Ans) Dependent Event

Dependent

⇓

Multiplication Rule of conditional Events

P(O) = 4/7 | P(Y) = 3/6

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

$= \frac{4}{7} * \left\{\frac{3}{6}\right\} = \frac{12}{42} = \frac{6}{21} = \frac{2}{7}$ //.

$P(Y/O)$ = Conditional Event

$m >>> 1$

$\left\{\frac{80}{100} \times (n+1)\right\} = \{\text{Index Position}\}$

(*) Packet of Biscuit

$\frac{3}{11}$.

$\{5 \quad A \quad 6 \quad O\} \Rightarrow P(A \text{ and } O)$

$P_r(A) * P(B/A)$

$\frac{3\cancel{0}}{11\cancel{0}}$ // $\frac{5}{11} * \frac{6}{10}$

→ Zoo

(*) Permutations.

{Tiger, Lion, Monkey, Zebra, Walrus, Snake}

$\rightarrow 6 \times 5 * 4 = 120$ ways or possibilities

→ {Tiger, Lion, Monkey}

→ {Tiger, Monkey, Lion}

→ {Lion, Monkey, Zebra}

→

→

With permutation, Order matters

that will be counted

n = Total no. of objects

r = # of packets we are picking

$${}^{n}P_r = {}^{6}P_3 = \frac{n!}{(n-r)!}$$

$n = 6 \quad r = 3$

$$= \frac{6!}{(6-3)!} = \frac{6!}{3!} = \frac{6\times5\times4\times\cancel{3!}}{\cancel{3!}} = 120$$

$${}^{6}C_3 = \frac{6!}{3!(3!)} = \frac{6\times5\times4\times\cancel{3!}}{3\times2\times1\times\cancel{3!}}$$

$= \frac{120}{6} = 20$ Combinations.

① $\underline{6} * \underline{5} * \underline{4} * \underline{3} * \underline{2}$ { KRISH, Youtube Channel, Stats, iNeuron, iNeuron Support, Power BI }

$n = 6$
$r = 5$

$${}^{n}P_r = \frac{6!}{(6-5)!} = \frac{6\times5\times4\times3\times2\times1}{1!} = 720$$

② Combination — Repetion will not occur

Lion Tigu Zebra
Zebra Lion Tiger
Lion Zebra Tigu
} → 1

$n = 6$
$r = 5$

$${}^{n}C_r = \frac{n!}{r!(n-r)!} = \frac{6!}{5!(6-5)!} = \frac{6\times\cancel{5\times4\times3\times2\times1}}{\cancel{5\times4\times3\times2\times1}} = 6$$

$n = 6 \quad r = 3$

$$= \frac{6!}{3!(6-3)!} = \frac{6!}{3!\times3!}$$

$n = 4$
$r = 3$

Dream 11 => Betting App { KRISH, Bharath, Sudhanshu, iNeuron }

$${}^{4}C_3 = \frac{4!}{3!(4-3)!} = \frac{4\times\cancel{3\times2\times1}}{\cancel{3\times2\times1}}$$

- KRISH Bharath Sudhanshu
- Bharath Sudhanshu iNeuron
- KRISH Sudhanshu iNeuron

KRISH Bharath iNeuron

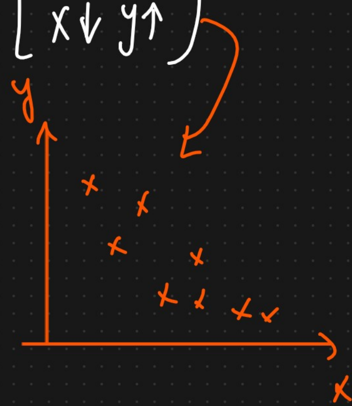
iNeuron Bharath KRISH

Covariance :

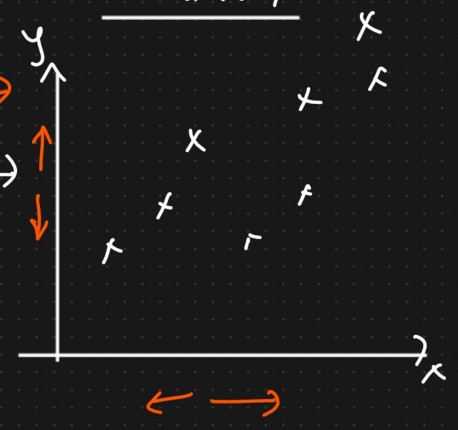
$$\left[\begin{Bmatrix} X\uparrow & Y\uparrow \\ X\downarrow & Y\downarrow \end{Bmatrix}\right]$$

$$\begin{Bmatrix} X\uparrow & Y\downarrow \\ X\downarrow & Y\uparrow \end{Bmatrix}$$

X	Y
Age	Weight
12	40
15	45
18	50
19	55
20	60



Mathematically



Mathematical formula to quantify the relationship between X & Y ??

Covariance

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

Sample Variance

$$Cov(X,X) = Var(X) = \frac{\sum (x_i - \bar{x}) * (x_i - \bar{x})}{N-1} = \frac{\sum (x_i - \bar{x})^2}{N-1}$$

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

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

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

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

1

X	Y
1	3
2	4
3	5

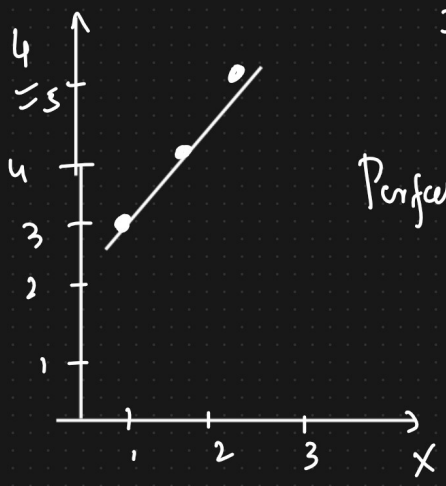
Sample mean

$\bar{x} = 2 \quad \bar{y} = 4$

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

$Cov(x,y) = 1$

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



Perfect +ve relation

X	Y
1	3
2	4
3	6

$\bar{x} = 2 \quad \bar{y} = 4.334$

$$Cov(x,y) = \frac{[(1-2)(3-4.334)+(2-2)(4-4.334)+(3-2)(6-4.334)]}{2}$$

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

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

$= 1.497$

$Cov(X,Y) = 1$

$Cov(X,Y) = 1.497$ ↑

+ve correlation

-1 to 1

Disadvantage

①

Z ← X ↔ Y

{ +1000 = −1000

+500 = −500

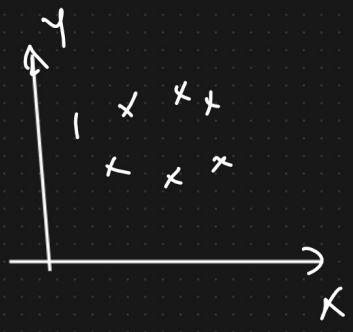
Cov → +2000 +1000 +3000 ↑↑

↓↓ 0

−1000 −

Define the strength (-1 to 1)

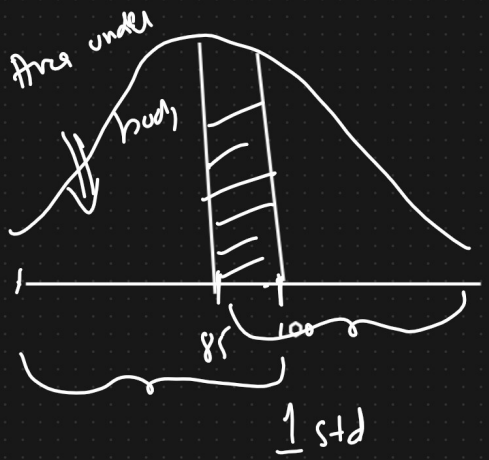
{Pearson Correlation And Spearman Rank Correlation}



x↑ y↑

x↓ y↑

x↓ y↓



95 to 100

Ages Weights

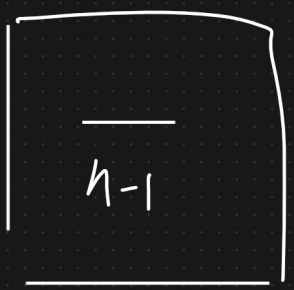
Age = 21

Age = 25

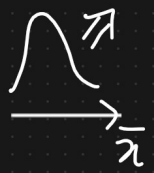
Age = 24

CLT

$n \geq 30$



$n \geq 30$



{$n < 30$}

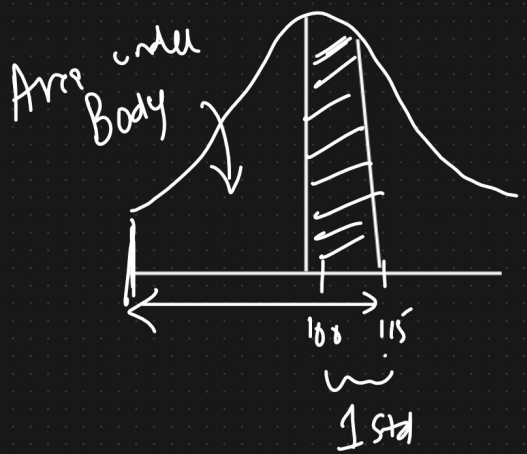
CLT

≈

$$\left\{ \begin{array}{l} Pr(1 \text{ or } 3) \\ = \\ Pr(1 \text{ and } 3) \\ = \end{array} \right\} \Rightarrow \text{Mutual or not Mutual}$$

$\Rightarrow$ Independent or Not Independent

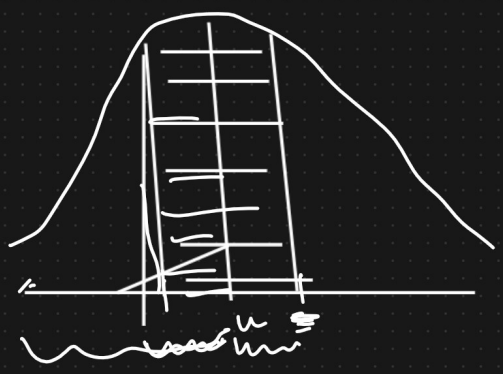
Z-Score Table $\leftarrow$ Area

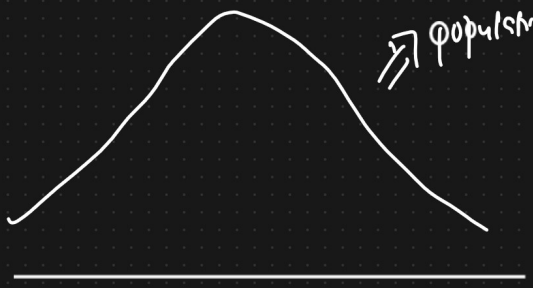


$\Downarrow$

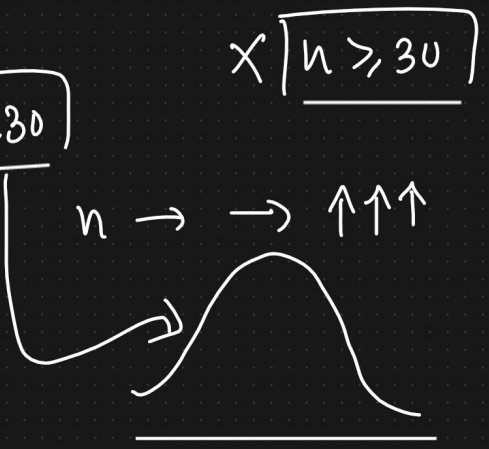
Zscore = How many Std it is away from mean

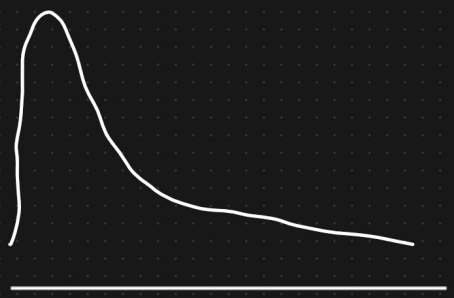
Left Z Score Table





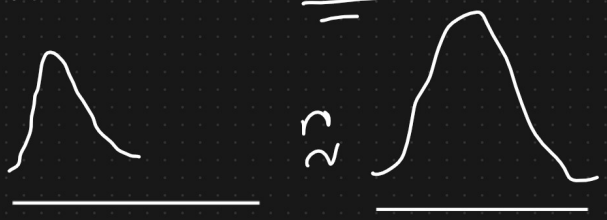
$n \geq 30$





$n \geq 30$ not on M

$n\uparrow\uparrow\uparrow$ $n \geq 30$



Age — Standardization → Z

Score down

-3 to $+3$

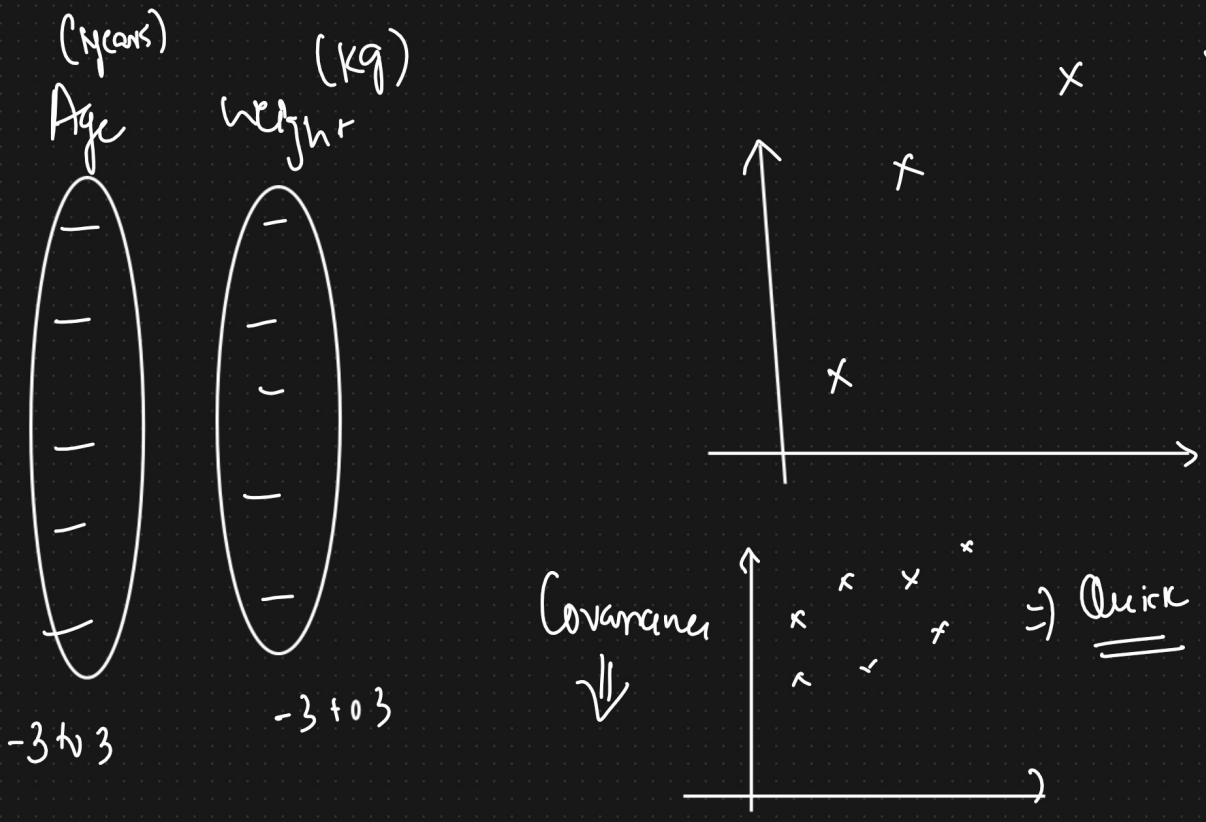
$x \leftrightarrow y$

$x\uparrow$ $y\uparrow$ } +ve relation

$x\downarrow$ $y\downarrow$ }

$x\uparrow$ $y\downarrow$ } -ve relation

$x\downarrow$ $y\uparrow$ }



Packet of Biscuit (Dependent)

4 A
5 B

$\Rightarrow Pr(A \text{ and } B)$

$= Pr(A) * Pr(B/A)$

$= \frac{4}{9} * \frac{5}{8} = \frac{20}{72}$ //.