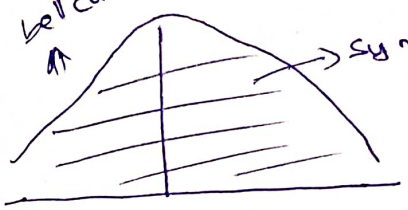
# Day - 3

## Statistics

1. Normal Distribution
2. Standard Normal distribution
3. Z - score
4. Standardization and Normalization

1. **Gaussian / Normal Distribution** :- is a probability distribution that is symentric about mean, showing the data near the mean are more frequent in occurence than data far from the mean

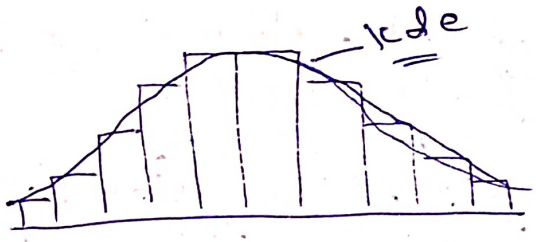
In graphical form, the normal distribution appears as a "bell curve"




distribution

Age, weight, height ⇒ distribution

Domain Expert




Example:-

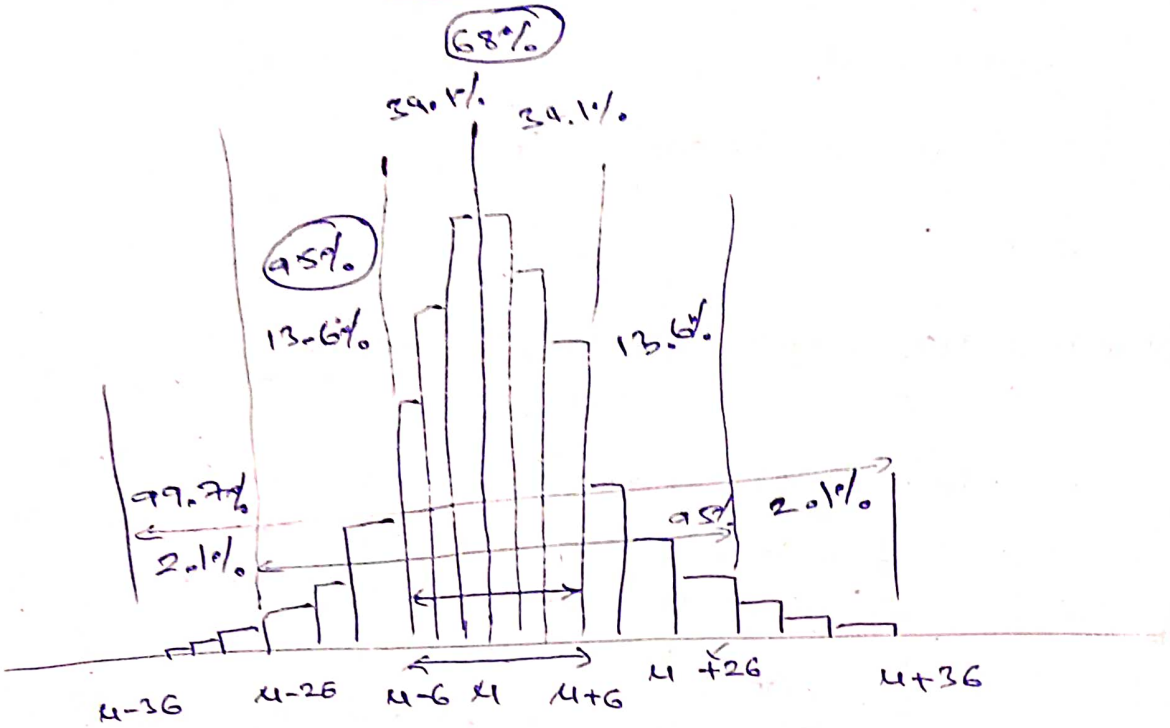
### IRIS DATASET

Petal length, Sepal width, petal width, Sepal length

⇓ Gaussian distribution

## * Emperial Rule of Normal distribution

Emprial rule :- **68 - 95 - 99.7%**




## Standard Normal Distribution

$X \approx$ Gaussian Distribution $(\mu, \sigma)$

$X = \{1, 2, 3, 4, 5\}$

$\mu = 3$

$\sigma = 1.41$

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

$y \approx SND\ (\mu = 0, \sigma = 1)$

$$z\text{-score} = \frac{x_i - \mu}{\frac{\sigma}{\sqrt{n}}} \quad n = 1$$

$\Rightarrow$ Standard Error

$\Rightarrow$ Inferential stats

$$z\text{-score} = \frac{x_i - \mu}{\sigma} \in \text{Simple}$$

$X = \{1, 2, 3, 4, 5\}$

$\mu = 3, \quad \sigma = 1.414$

$\Rightarrow \frac{1-3}{1.414} = -1.414$

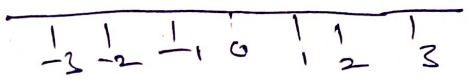
$\Rightarrow \frac{2-3}{1.414} = -0.707$

$\frac{3-3}{1.414} = 0$

$\frac{4-3}{1.414} = 0.707$

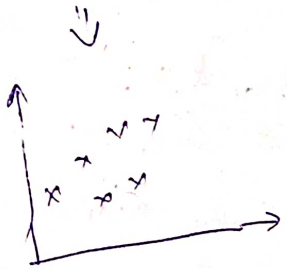
$\frac{5-3}{1.414} = 1.414$

$y = \{-1.414, -0.707, 0, 0.707, 1.414\}$




Why? [Standardization] $\Rightarrow \boxed{\mu = 0 \quad \sigma = 1}$

| Age (years) | weight (kg) | Height (cm) |
|---|---|---|
| 24 | 72 | 150 |
| 26 | 78 | 160 |
| 32 | 84 | 165 |
| 33 | 92 | 170 |
| 34 | 82 | 150 |
| 28 | 83 | 180 |
| 29 | 80 | 175 |

Feature scaling

$\frac{x-\mu}{\sigma}$

Normalization

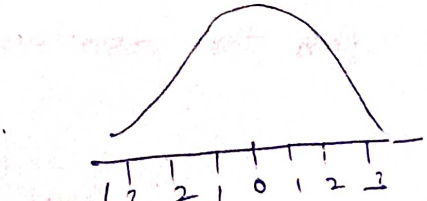
standardization {z-score}

[0-1]

[0-5]

$\mu = 0 \quad \sigma = 1$

[-3, 3]



## Normalization: (lower scale ⟷ high scale)

### ① Min Max Scaler (0–1)

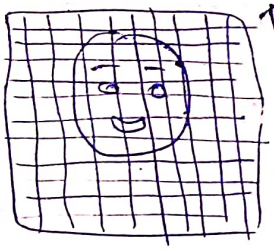
$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

| x | y | y1 – standar |
|---|---|---|
| 1 | 0 | -1.414 |
| 2 | 0.25 | -0.707 |
| 3 | 0.5 | 0 |
| 4 | 0.75 | 0.702 |
| 5 | 1 | 1.414 |

$\Rightarrow$ $\frac{1-1}{5-1} \Rightarrow 0$ $\frac{2-1}{5-1} \Rightarrow \frac{1}{4} = 0.25$ $\frac{3-1}{5-1} \Rightarrow 0.5$

$\frac{4-1}{5-1} \Rightarrow 0.75$, $\frac{5-1}{5-1} = 1$

* we mainly used in deep learning, Image process (CNN, ANN, RNN)



0-255 $\Rightarrow$ 0-1

Normalization,

### standardization

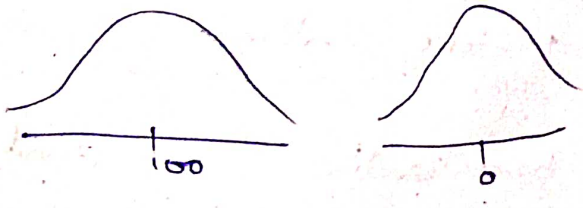
$$z\ score = \frac{x_i - \mu}{\sigma}$$

$x \rightarrow$ Normal distribution $(\mu, \sigma)$

$\Downarrow$ z-score

$y \rightarrow$ SND $(\mu = 0, \sigma = 1)$

### why do we this?




$\rightarrow$ bring the feature in the same scale

Normalization [0-1]

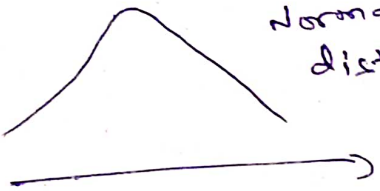
① min max scaler ⟺ standadization

⇓ ⇓

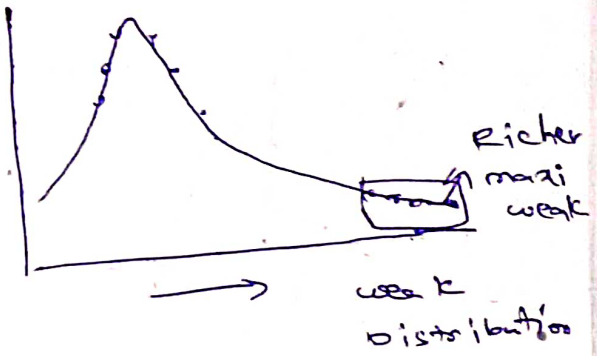
ml nd

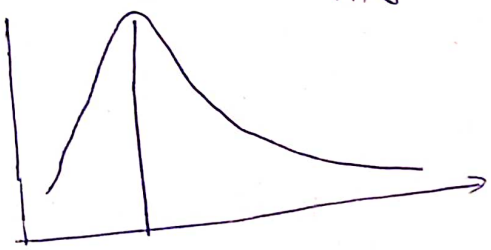

minmax scal

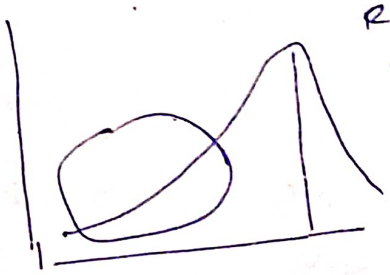
## ① log Normal Distribution :-




log normal distribution

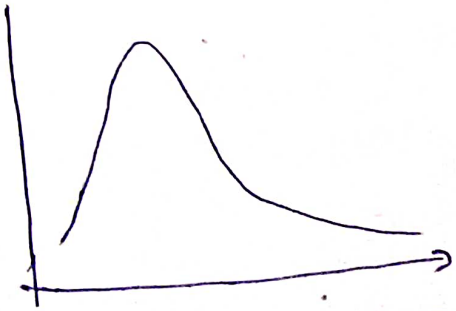


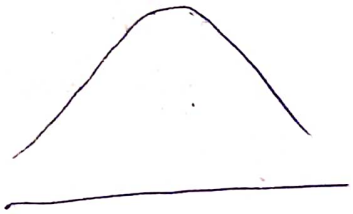

from ascending order given the relationship of mean median & mode?.

* log Normal distribution to Normal distribution




$y = \ln(x)$

$X \approx$ log Normal distribution $(\mu, \sigma)$



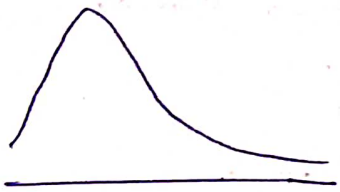
$\Rightarrow$ (Inverse)

antilog

$\Downarrow$

exp $(x)$

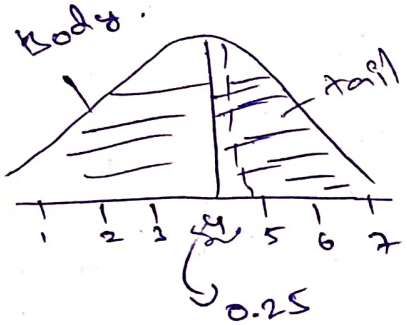
$\Downarrow$



$x \approx$ ..

$\mu$

* $x = \{1, 2, 3, 4, 5, 6, 7\}$ $\quad \mu = 4$, $\quad \sigma = 1$




area = 1

Question: what is the percentage of score that falls above 4.25%

$x$

fall below 3.75?

$= \quad =$

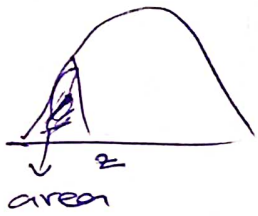
$$z\text{-score} = \frac{x - \mu}{\sigma} \Rightarrow \frac{4.25 - 4}{1} = 0.25$$

$\Downarrow$ 0.59

$1 - 0.59$

$\Rightarrow 0.41\%$

$= \quad =$

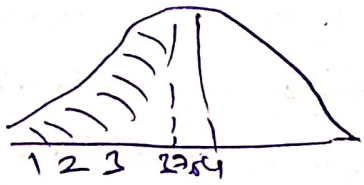



z-table cover the area

↳ +ve table

↳ -ve table

$$z = \frac{x_1 - \mu}{\sigma} = \frac{3.75 - 4}{1} = -0.25$$




40%

$\Rightarrow$

$\Rightarrow \frac{4.25 - 4}{1} \Rightarrow \frac{0.25}{} \Rightarrow$

$\Rightarrow \frac{3.75 - 4}{1} \Rightarrow$

* In India the average IQ is 100 with a standard deviation of 15. what is the percentage of popula-tion could you expect to have an IQ

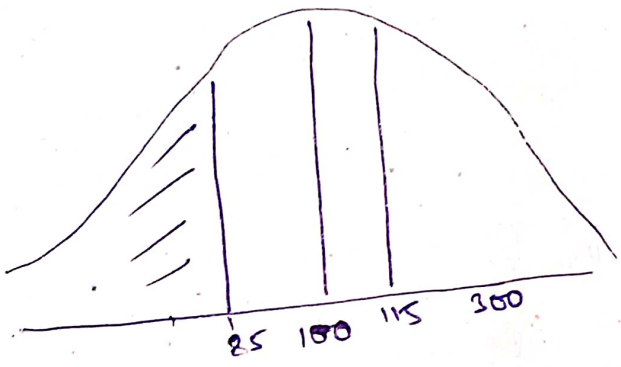
① lower then 85 = 0.1587

② Higher then 85 = 0.8413

③ Between 85 and 100 = 0.3413

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

0.5 − 0.1587



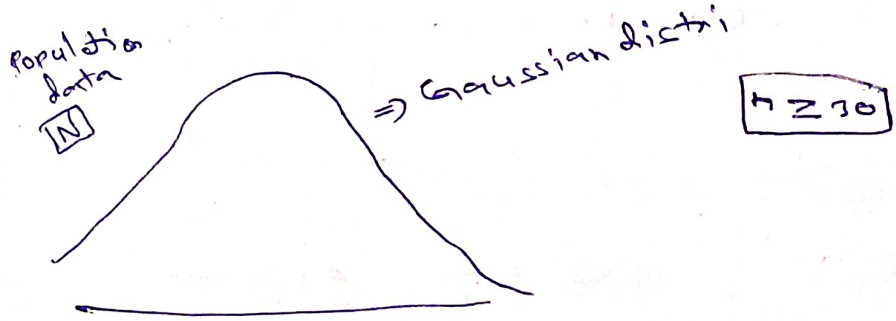
# Day - 4

## Stats

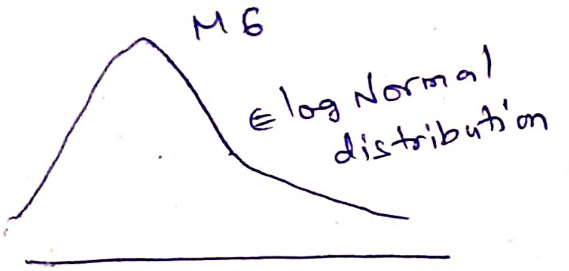
1. Central limit Theorem
2. Probability
3. Permutation and Combination
4. Covariance, Pearson correlation, Spearman Rank correlation
5. Bernoullis Distribution
6. Binomial Distribution
7. Power law

### Central limit Theorem :-

The CLT states that the distribution of sample means approximates a normal distribution as the sample size gets larger, regardless of population's distribution. Sample size equal to or greater than 30 are often considered sufficient for the CLT to hold




* Population data is Norm dis or logND or left skew we take $n \geq 30$ and n size of sample and n is no. of samples mean we can draw the Gaussian distribution.




$n \geq 30$ size of sample ($n$) → No. of samples ($m$)

↑ mean

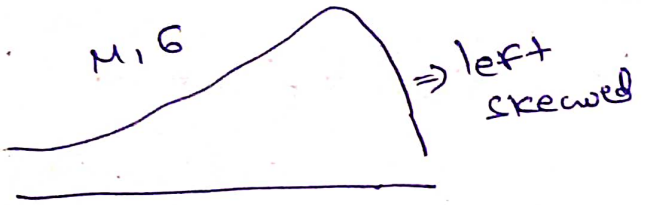
$\rightarrow S_1 \rightarrow \{x_1, x_2, \ldots x_n\} \rightarrow \bar{x}_1 = \bar{S}_1$

$\rightarrow S_2 \rightarrow \{x_3, x_4 \ldots x_n\} \rightarrow \bar{x}_2 = \bar{S}_2$

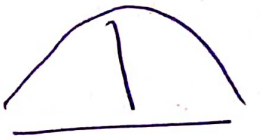
$\rightarrow S_3 \rightarrow \{x_4, x_5 \ldots x_n\} \rightarrow \bar{x}_3 \quad \bar{S}_3$

$\vdots \quad \vdots$

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




Sampling with replacement




size of shark through out the world ⇒ Assumption $n \geq 30$

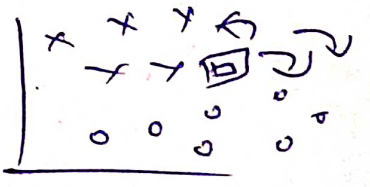
* **Probability** :- Probability is a measure of the likelihood of an event

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

⇓

Sholay → coin ⇓ unfair coin $P(H) = 1$

using probability ⇓

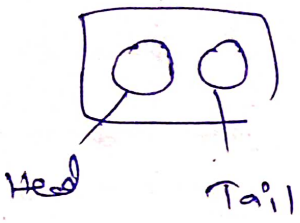


Rolling a Dice :- $P(1) = \frac{1}{6}, \quad P(2) = \frac{1}{6} \quad P(3) = \frac{1}{6}$

① **mutual exclusive event** :-

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

① Tossing coin ② Rolling dice




## ② Non mutal Exclusive Events

Two events can occur at the same time

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

Bag of marbels

## mutual exclusive events :

① what is the probability of coin landing on head or tails

Addition Rule for mutual exclusive events

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

$\frac{1\times2+2\times1}{2+2} = \frac{4}{4} = 1$

② what is the probabilities of getting 1 or 6 or 3 while rolling a dice ?

$$P(1 \text{ or } 6 \text{ or } 3) = P(1) + P(6) + P(3)$$
$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1/2$$

$\frac{3}{6}$

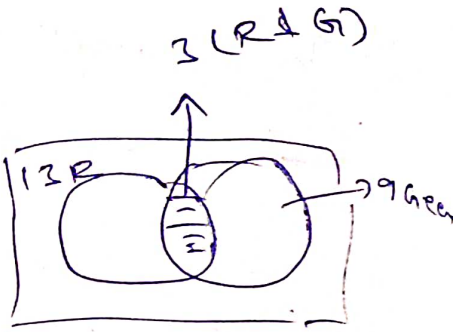
## Non Mutual exclusive events :-

Bag of marbles :- 10 red, 6 Green, 1 (R or G)

* 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 events :-

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

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

Dice of cards → what is the probability of choosing ♡ or Queen

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

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

* Multiplication Rule :-

* Dependent Events :- Two events are dependent if they effected one another

Bag of marble { white, yello }

$\Rightarrow P(w) = \frac{4}{7}$ (↑ white marble)

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

## Independent events :

what is the probability of rolling a "5" and then a "3" with a normal 6 six sides dice?

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

multiplication Rule for Independent Events

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

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

$P(A \text{ or } B) =$ { mutual Events / Non mutal Events

$P(K \text{ or } \heartsuit)$ $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$ → Non mutual

$P(A \text{ or } B) = P(A) + P(B)$ [mutual exclusive]

## Dependent and Independent Events

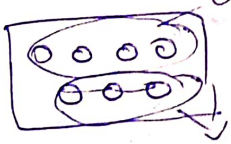
Tossing a coin

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

Event A Event B

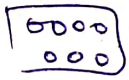
$P(A \text{ and } B) = P(A) * P(B)$

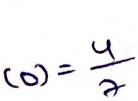
②




→ Dependent Event

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



$P(O) = \frac{4}{7}$ →  → $P(Y/O)$ → conditional probability

↳ $P\left[\frac{3}{6}\right] = \frac{1}{2}$

↓ orange marble

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

$= \frac{4}{7} * \frac{3}{6} = \frac{4}{14} \frac{2}{1} = \frac{2}{7}$

## * Permutation :-

The term permutation refers to a mathematical calculation of the number of ways a particular set can be arranged

* ~~Unique~~ combination

$^nP_r$ => permutation

n = Total no of objects

r = no. of objects taken at a time

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

* Possible arrangements

Example:- How many ways you can arrange the letters in the word MATH?

n=4 · r=1

$$^nP_r = \frac{n!}{(n-r)!} = \frac{4!}{(4-1)!} \Rightarrow \frac{24}{4} \Rightarrow 4$$

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

$= 24$ ways

## Combination:-

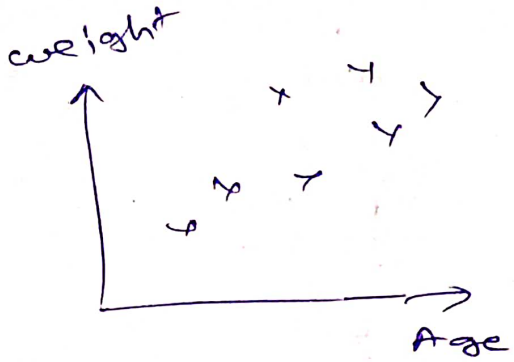
* unique elements
* Repeat will not occure

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

$= 10$ /.

## Covariance:

In statistics and mathematics, covariance is a measure of the relationship between two random variables




| x | y |
|---|---|
| age | weight |
| 12 | 40 |
| 13 | 45 |
| 14 | 48 |
| 15 | 60 |
| 17 | 62 |
| 18 | 66 |
| $\bar{x} = 15$ | $\bar{y} =$ |

Age↑ height↑
Age↓ height↓

Quantity the relationship x & y using mathematical equations

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

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

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

$Cov(x,x)$ ⇐ equal

$$Cov(x,x) = Var(x)$$

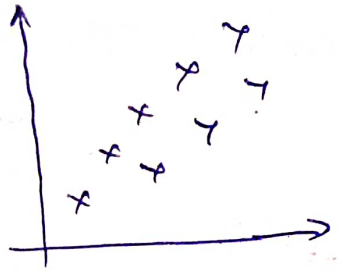
+ve covariance

| | |
|---|---|
| x↑ | y↑ |
| x↓ | y↓ |

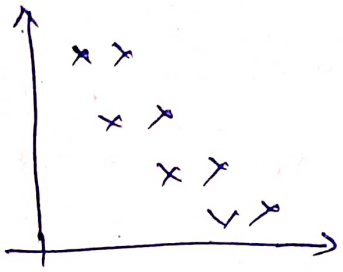
covariance = 0 [No relationship x & y]

-ve covariance

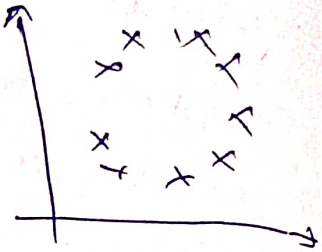
| | |
|---|---|
| x↑ | y↓ |
| x↓ | y↑ |



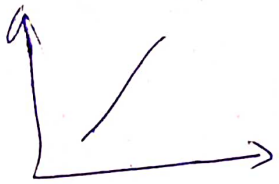
+ve

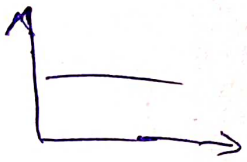


-ve



No relation





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

$cov(x,y) = -ve$

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

$\Rightarrow -3.25$

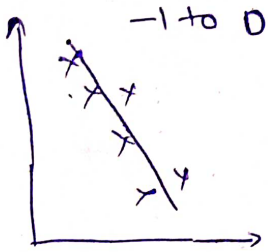
| | |
|---|---|
| x↑ | y↓ |
| x↓ | y↑ |

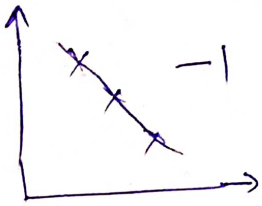
* Person corelation coefficient (-1 to 1)

is the most common way of measuring a linear correlation. it is number between -1 to 1 that measure the strength and direction of the relationship between two variables

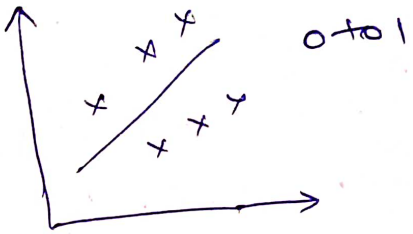
$$\rho(x,y) = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y}$$

scale
+: covar
-: covariate




-1 negative


Spearman's rank correlation coefficient

measures the strength and direction of associa
tion between two ranked variables

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

Example

Ascending order

| X | Y | R(X) | R(Y) | Spearmans rank correlation |
|---|---|---|---|---|
| 10 | 4 | 4 | 1 | |
| 8 | 6 | 3 | 2 | |
| 7 | 9 | 2 | 3 | |
| 6 | 10 | 1 | 4 | |