24th March 2023

let's start @ 9:05

Correlations

⇒ loanID Education Income loan Amt.

graduate Non-graduate

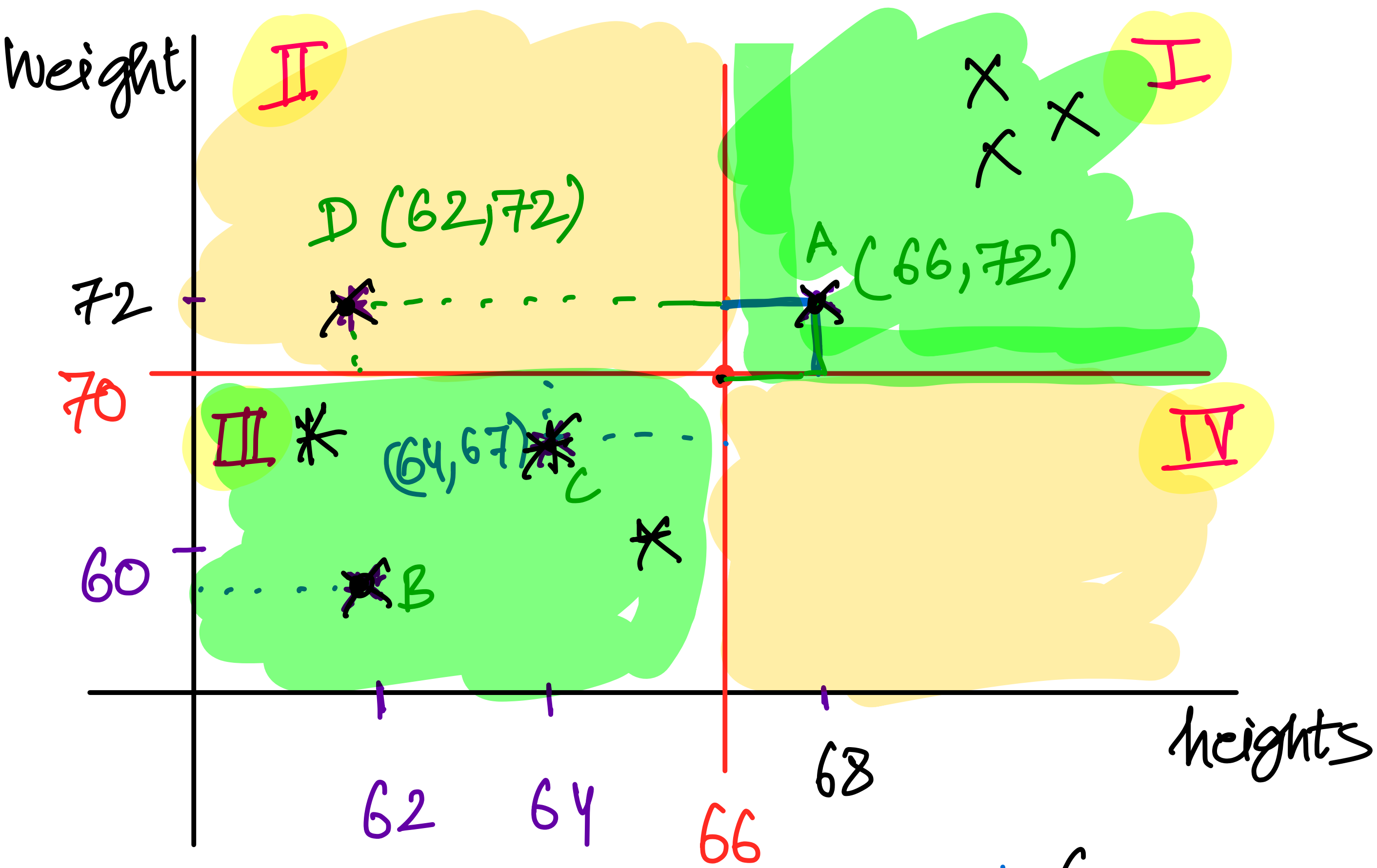
high collateral

[100 K, 10000 K]

100 K, 1 LK.

the Effect

Height (inches)	Weight (kg)
68	72
62	58
64	67
61	72 ✓
70	79
66	61
61	68
65	64
71	80
72	79
$\bar{h} = 66$	$\bar{w} = 70$



$(68-66).(72-70) = (+2)(+2) = 4$

$(62-66)\cdot(58-70) = (-4)\cdot(-12) = 48$

$(64-66)\cdot(67-70) = (-2)(-3) = +6$

$(62-66)(72-70) = (-4)(+2) = -8$

$\vdots$

$(72-66)(79-70) = 60$

$\frac{1}{10}(4+48+6-8+\cdots+60)$

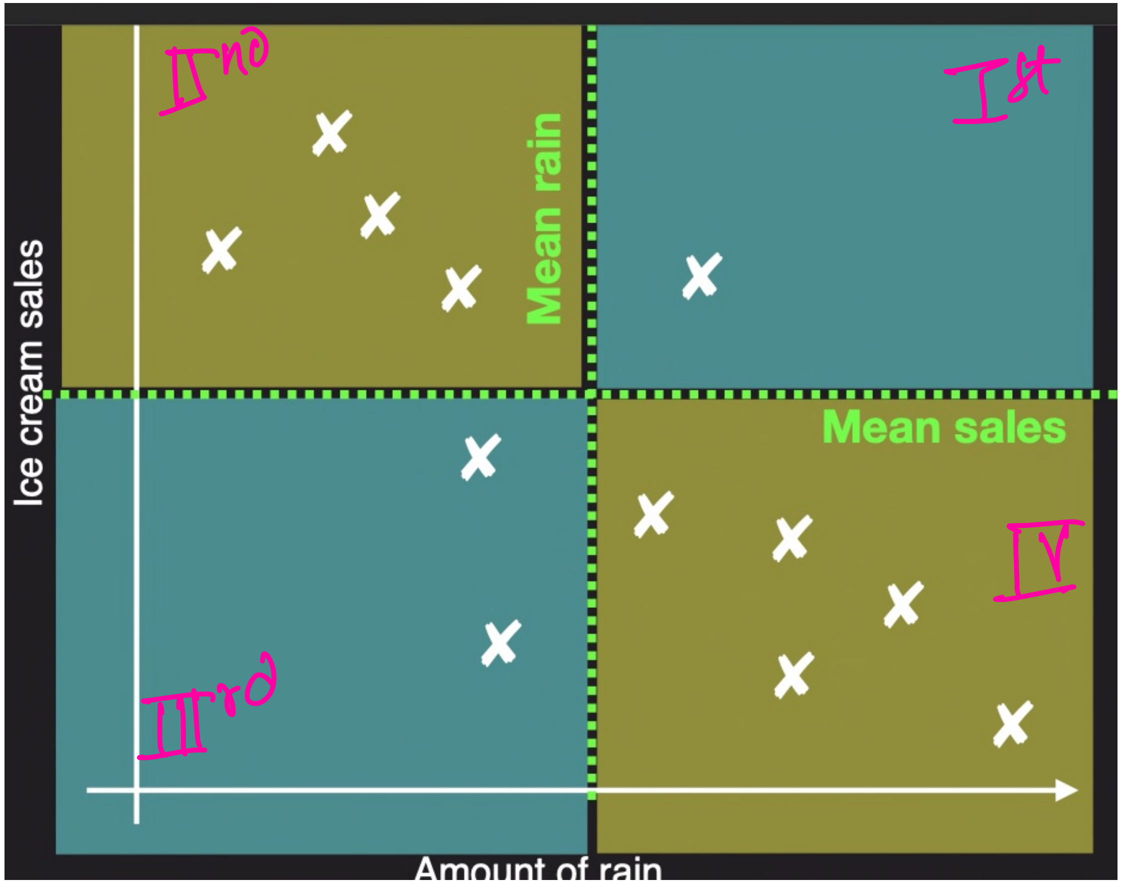
Avg. of these #'s

(COVARIANCE)

the

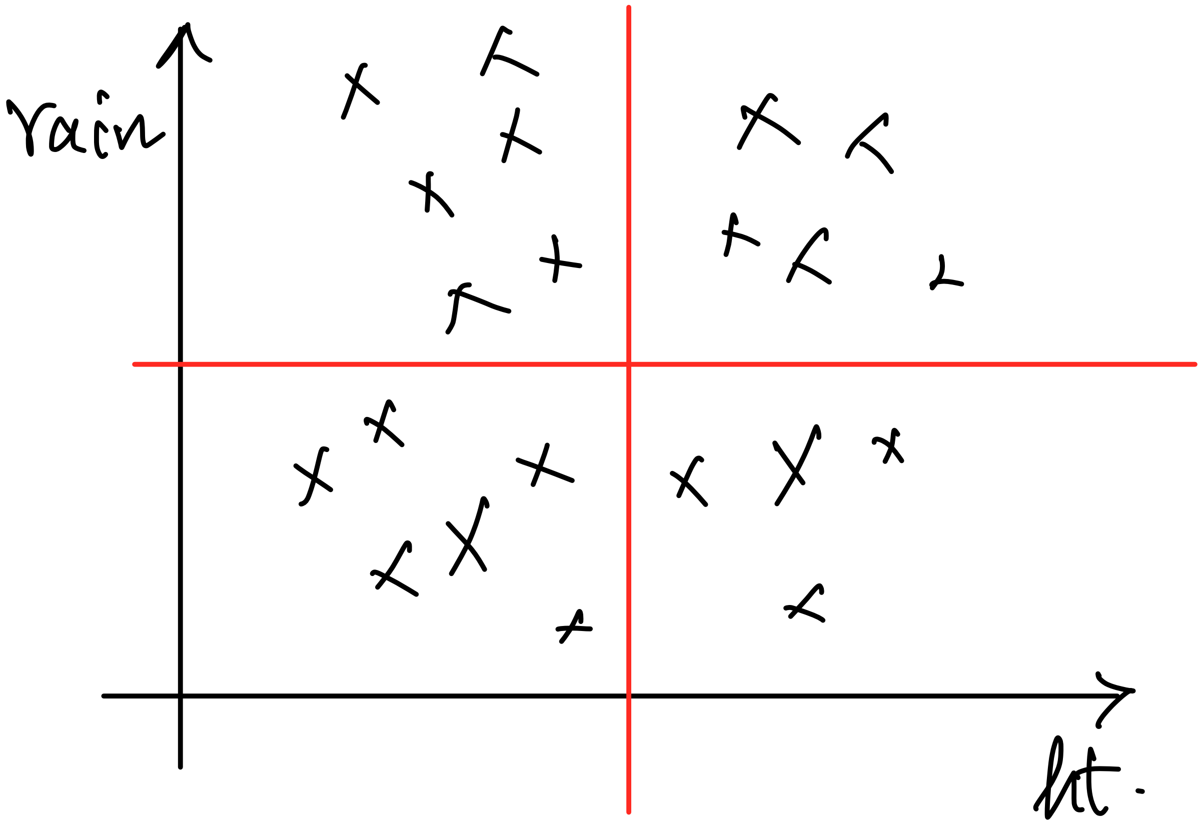
$$\cdot \; Cov(x,y) = \frac{1}{n}\sum_i (x_i - \bar{x})(y_i - \bar{y})$$

Icecream Vs Rain



−ve covariance

Height v/s Rain



Cov. ≃ 0

↓

very low

$$\boxed{\text{Covariance} \in (-\infty \text{ to } +\infty)}$$

$$\rightarrow \frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y})$$

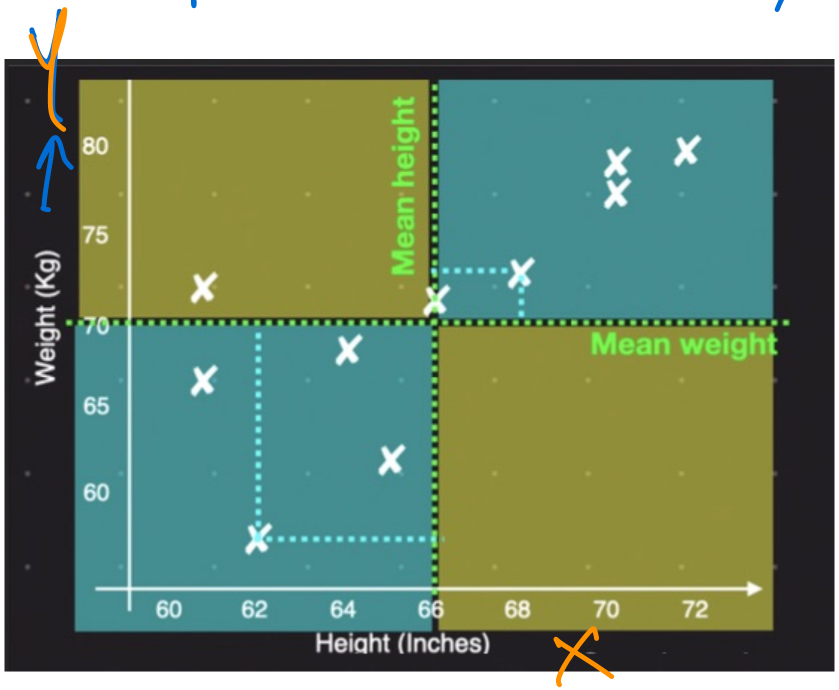
Cov.

-1000

-20000

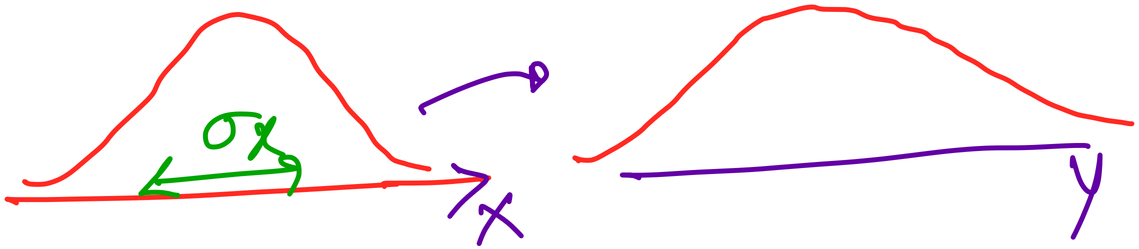
$+1.1$

$1'' = 2.5 \text{ cm}$, $1 \text{ Kg} \simeq 1.7$ pounds.



$\Rightarrow \sigma_x = 2.5$ $\Big\}$ $\sigma_x^2 = (2.5)^2$.

$\mu_x = 65$ inches.



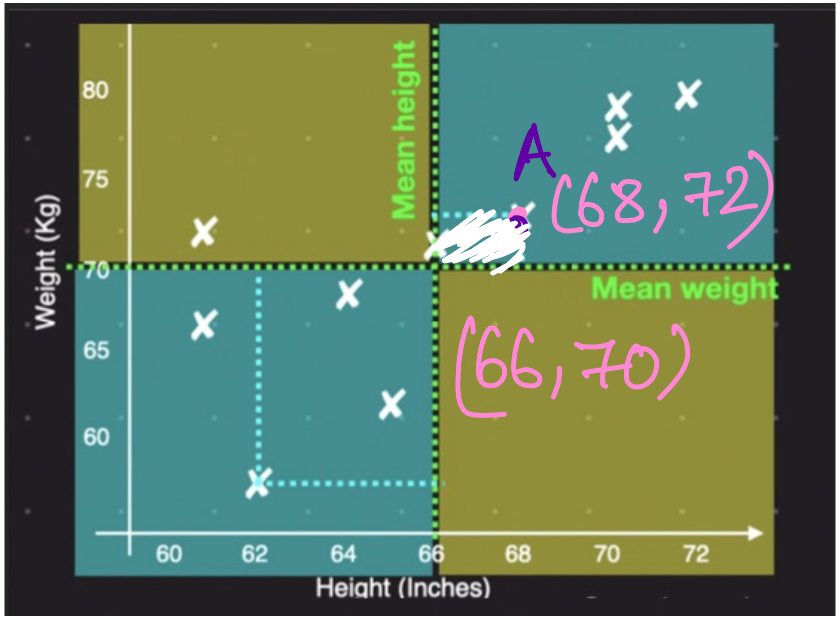
in cm

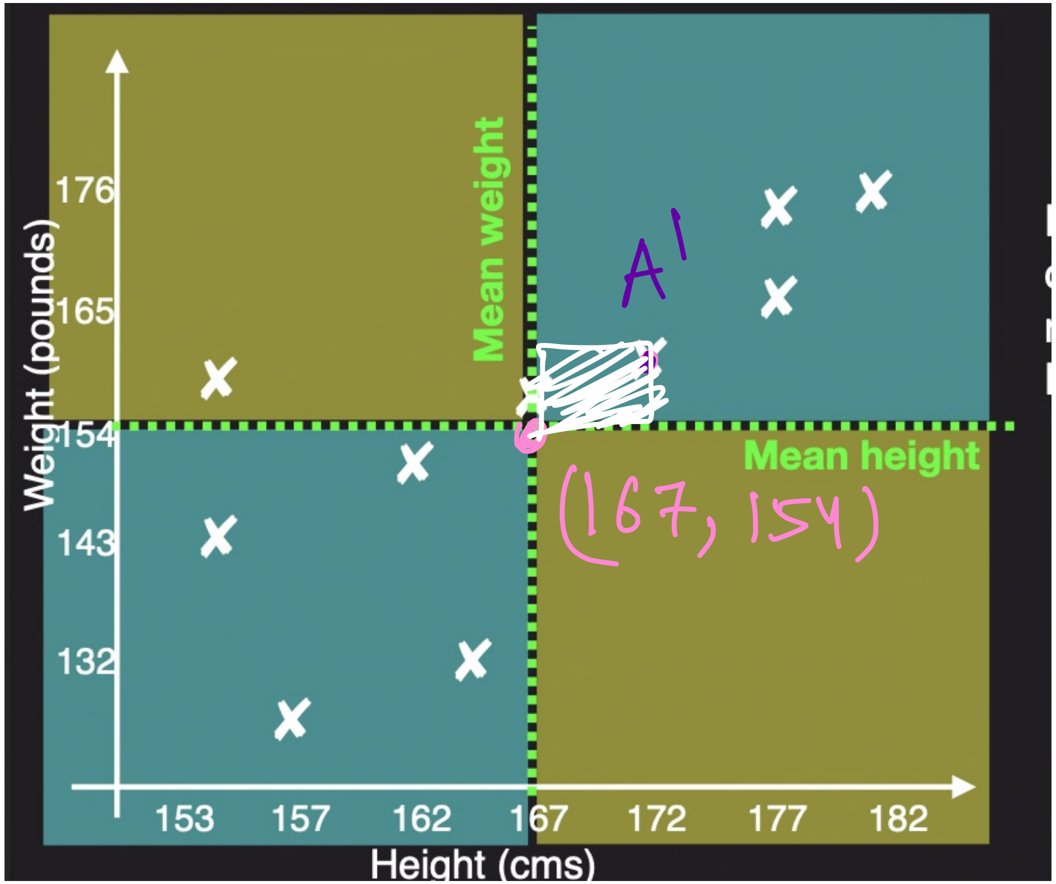
$\Rightarrow$ $X \rightarrow$ inches

$Y \rightarrow$ cm.

$\Big\}$ $Y = (2.54)X$.

$\boxed{\sigma_y = (2.54)X}$





Old.

(66,70)

A: (68,72)

New

(167,154)

A': (172, 159)

1 inch = (2.54) cm.

1 Kg = 2.2 pounds.

(68−66)·(72−70)

= +4

NewArea.

(172−167) * (159−154)

= 5×5

= +25

Old Cov. = +'m'.

New Cov = +'M'

M >> m

Pediatritians. Baby wt. ht.

$\tilde{w}$ $\tilde{h}$

0 1

$$\boxed{\tilde{w} = \left(\frac{w - \overline{w}}{\sigma_w}\right)}$$

$$\boxed{\tilde{h} = \left(\frac{h - \overline{h}}{\sigma_h}\right)}$$

$\Rightarrow X \rightarrow$ Random Var.

$Y = (2.54) X$.

$\sigma_y = (2.54)\, \sigma_x$

$\sigma_x = (2.54)$

New Var $\rightarrow A = \dfrac{X}{(2.54)}$

$\sigma_A = \dfrac{\sigma_x}{(2.54)}$.

$$\Rightarrow \sigma_A = \frac{2.54}{2.54} \Rightarrow \boxed{\sigma_A = 1}$$

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

$$\frac{(x_i - \bar{x})}{\sigma_x} \qquad \frac{(y_i - \bar{y})}{\sigma_y}.$$

$$\Rightarrow \boxed{\frac{1}{n} \sum \frac{(x_i - \bar{x})}{\sigma_x} \cdot \frac{(y_i - \bar{y})}{\sigma_y} = \rho_{xy}}$$

Correlation

PEARSON

If $\text{Cov}(x,y) \rightarrow (-\infty \text{ to } +\infty)$

$\rho_{xy} \rightarrow (-1 \text{ to } +1)$

$(\text{Cov.})_{x,y} \rightarrow -100$

$(\text{Cov})_{m,n} \rightarrow -300$

$\rightarrow$ not able to conclude.

xy $\quad$ mn $\rightarrow$ strong -ve relationship

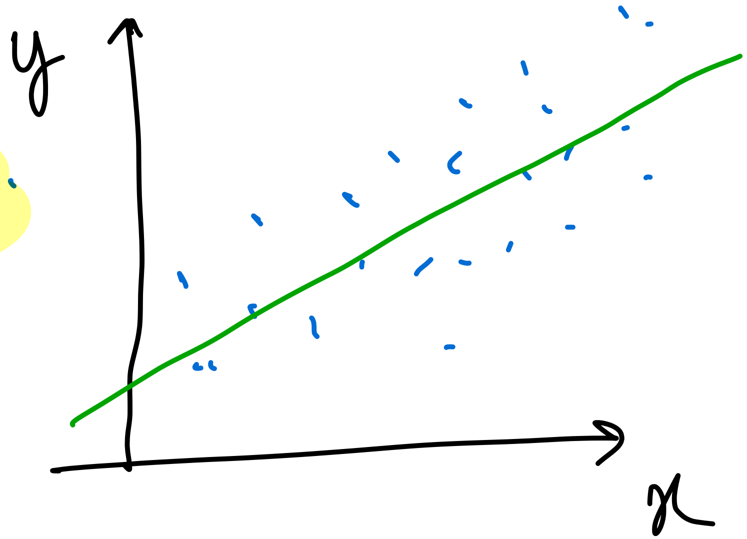
$\rho_{xy} = -0.80$

$\rho_{mn} = -0.95$

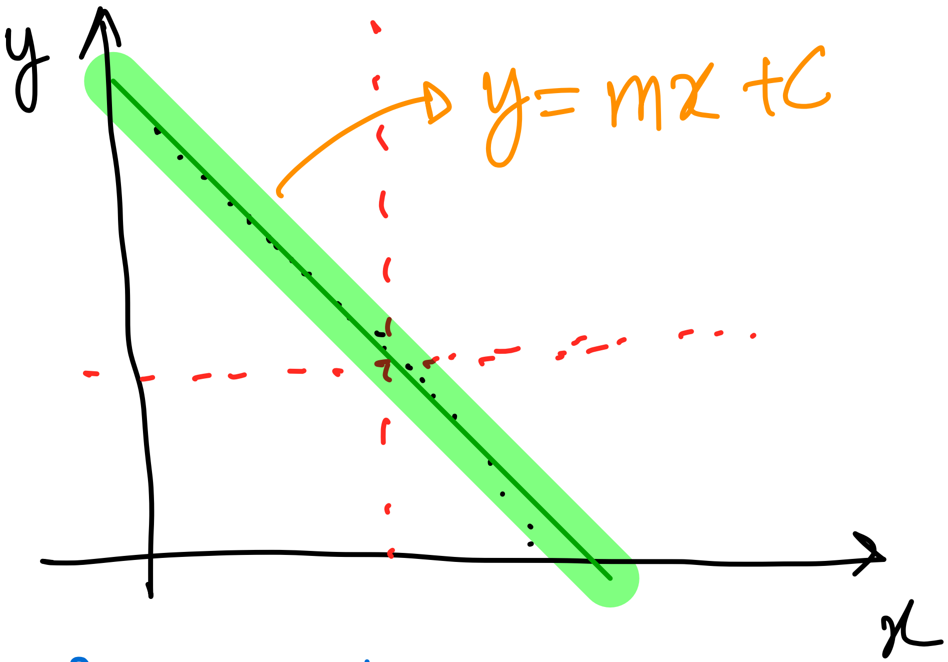
$\rightarrow$ m,n have strong -ve relationship wrt. x,y

Correlation $\longrightarrow$ (-1 to $+1$)

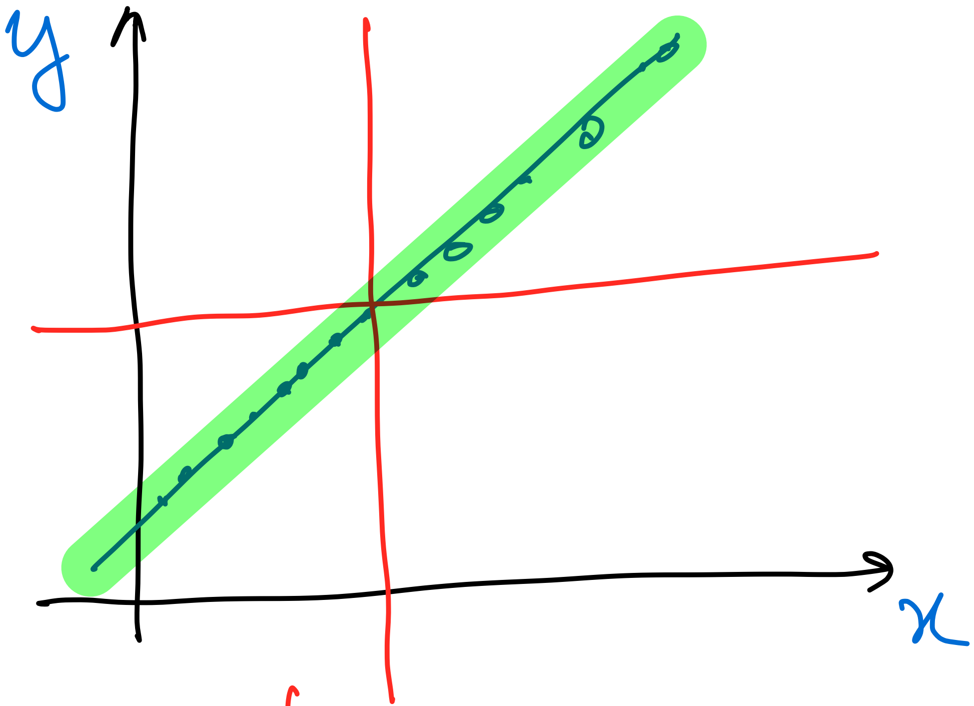
strength of linear relationship.



Corr. = -1



Perfect linear relationship.

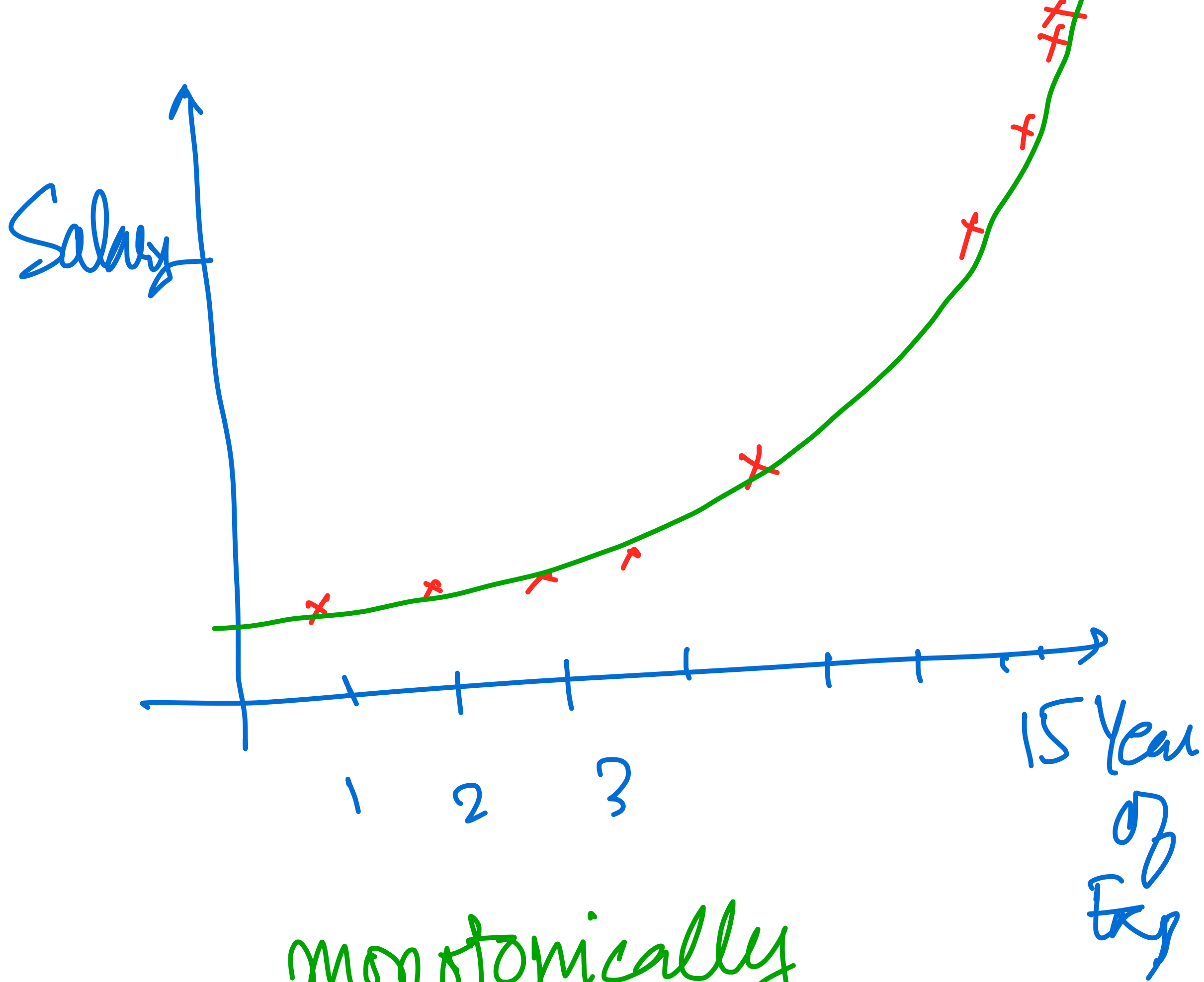


Corr = $+1$

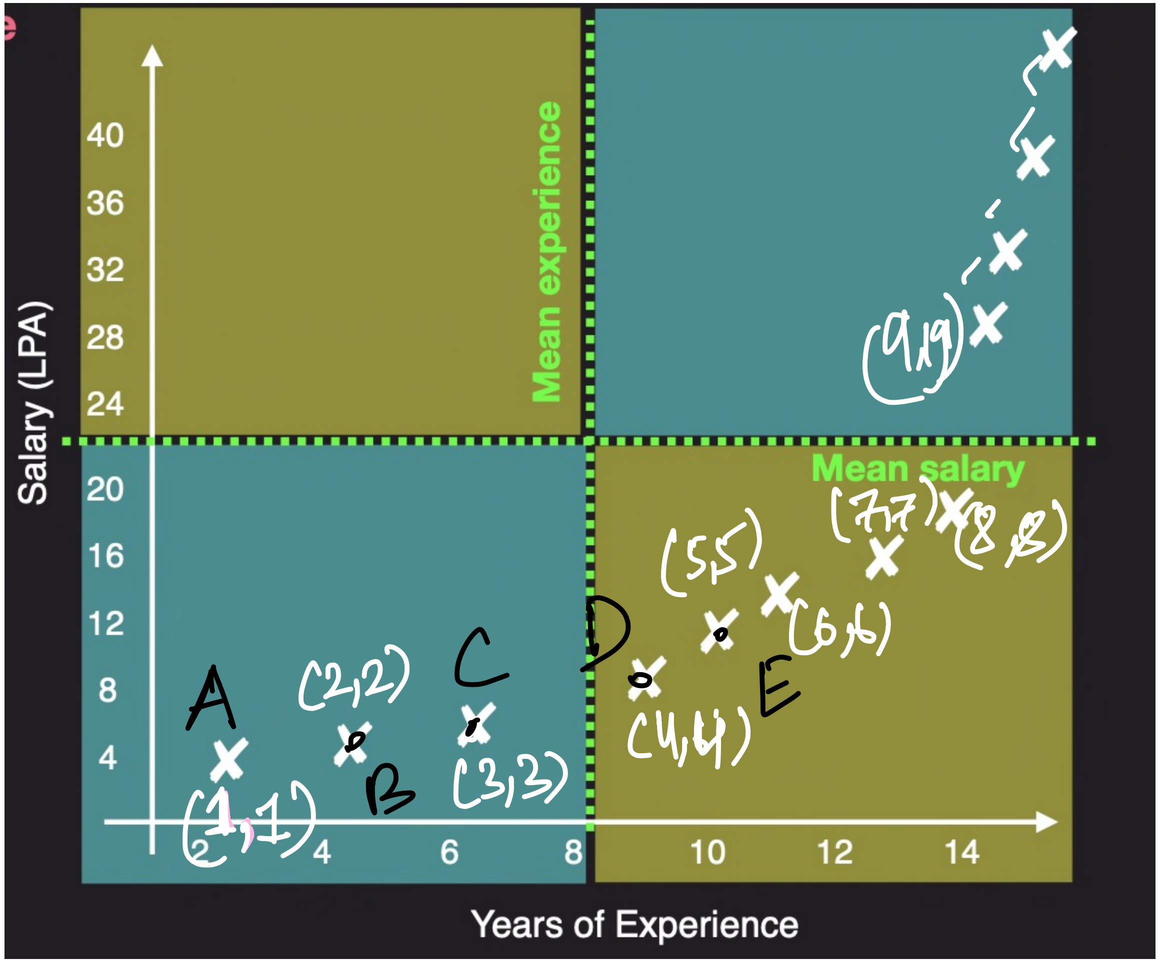
Salary v/s Years of Exp.

linear → y
→ N

dec. at any time { → Y
→ N



monotonically increasing



Corr. should be +1

In real corr < 1

0.8, 0.7

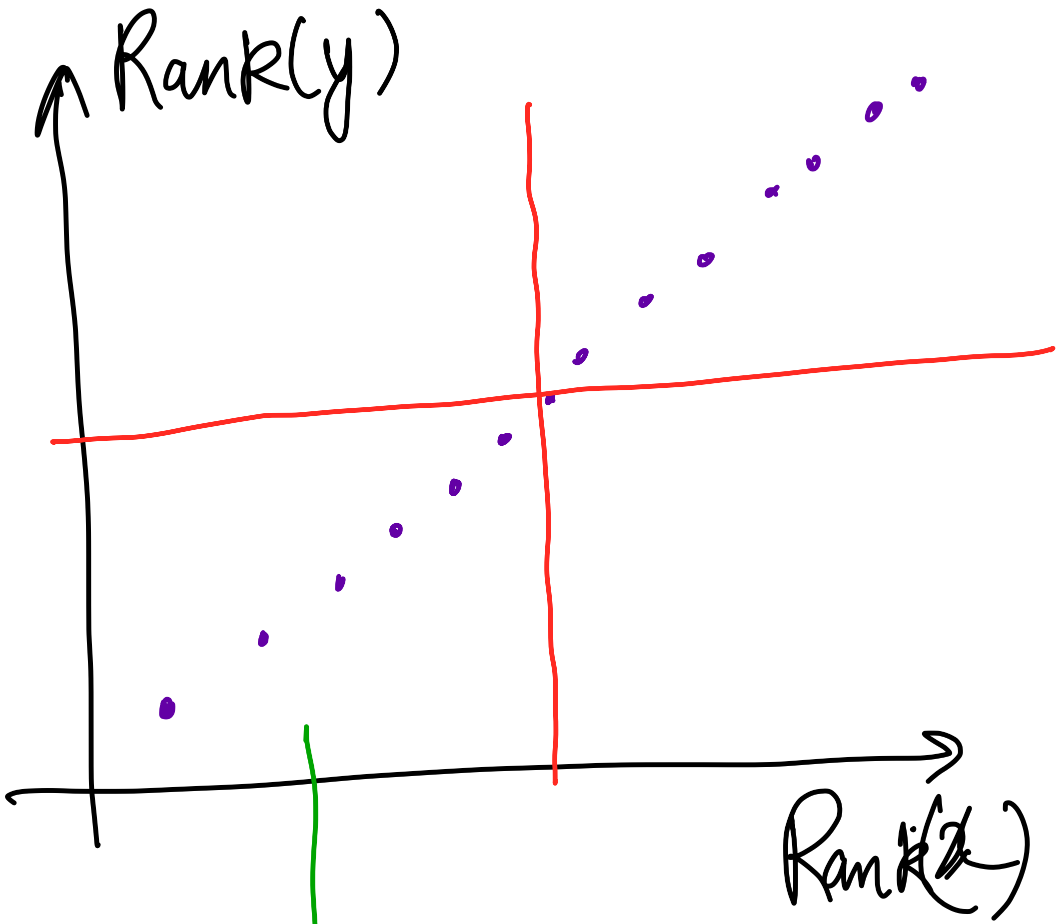
Spearman

Spearman Correlation

+1

Ranked Correlation

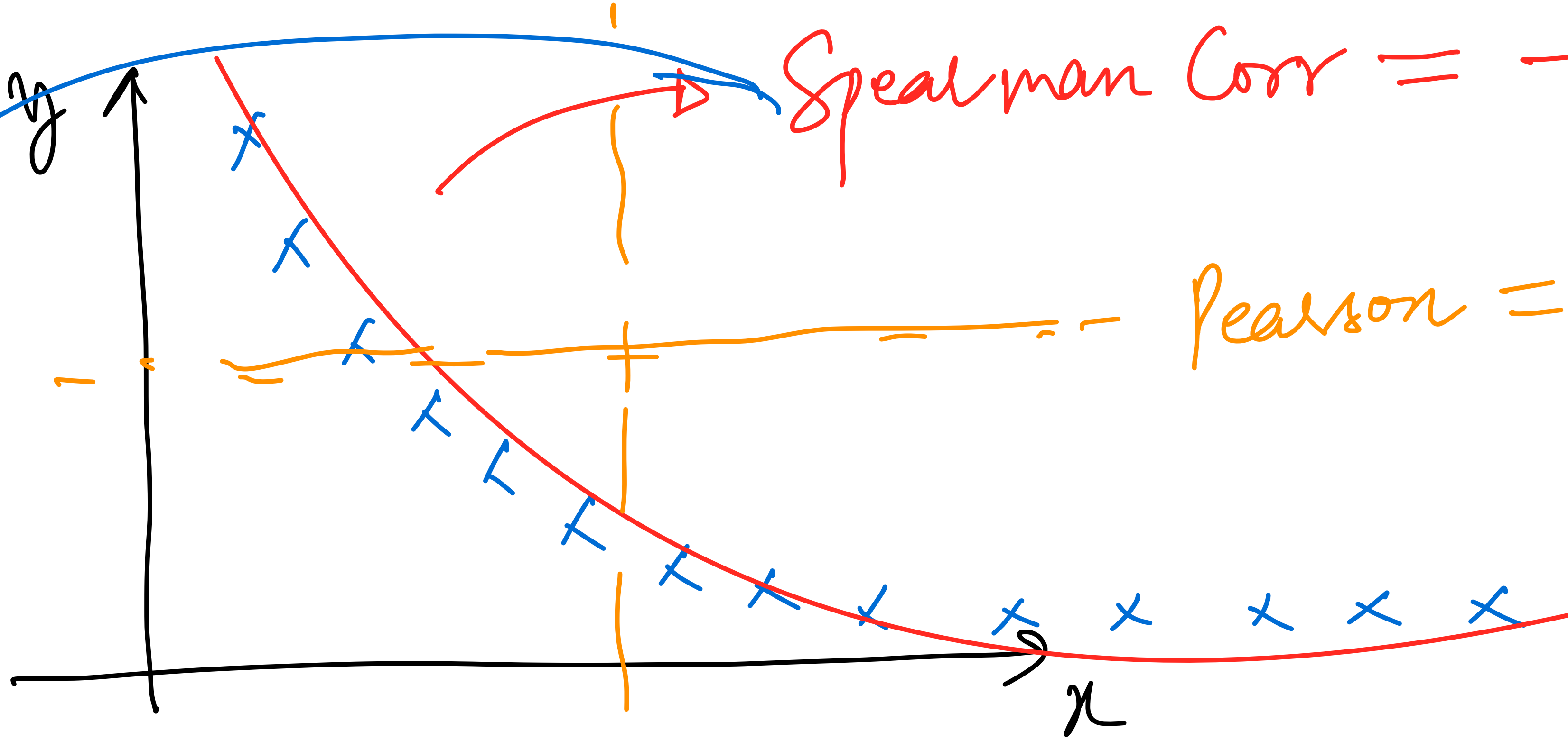
	rank(x)	rank(y)
A	1	1
B	2	2
C	3	3
D	4	4
E	5	5
⋮	⋮	⋮



Pearson Correlation

Rank the data ⟶ Compute

+1 ✓



Spearman Corr = -1

Pearson = -0.8

strength of monotonically ↑ or ↓ relationship b/w x & y.

10:20 *