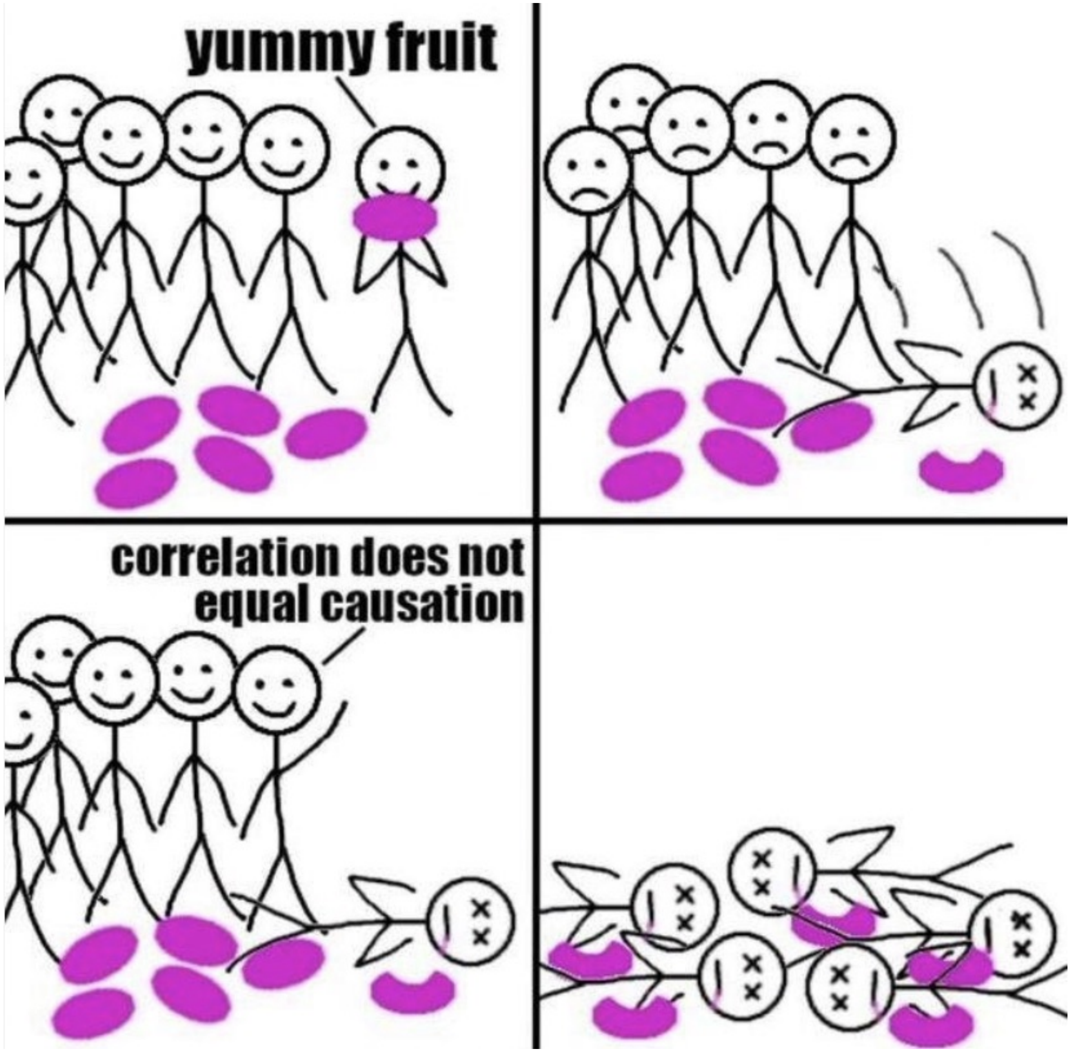
6th july 2023

Correlations



Let's start @ 9:05

① Num. v/s Cat. → t test, z test

Gender	Income
M	80 K
F	90K
M	80K
F	75 K.

M F

μ_m μ_f.

② Education level Income

id.	
1	70K
2	65K
3	59K
4	⋮

ANOVAA

$\alpha = 0.05$

③ Cat vs Cat. → Chi^2.

④ Num. v/s Numerical. :

App. ID

unique to every cust.

Education

Income

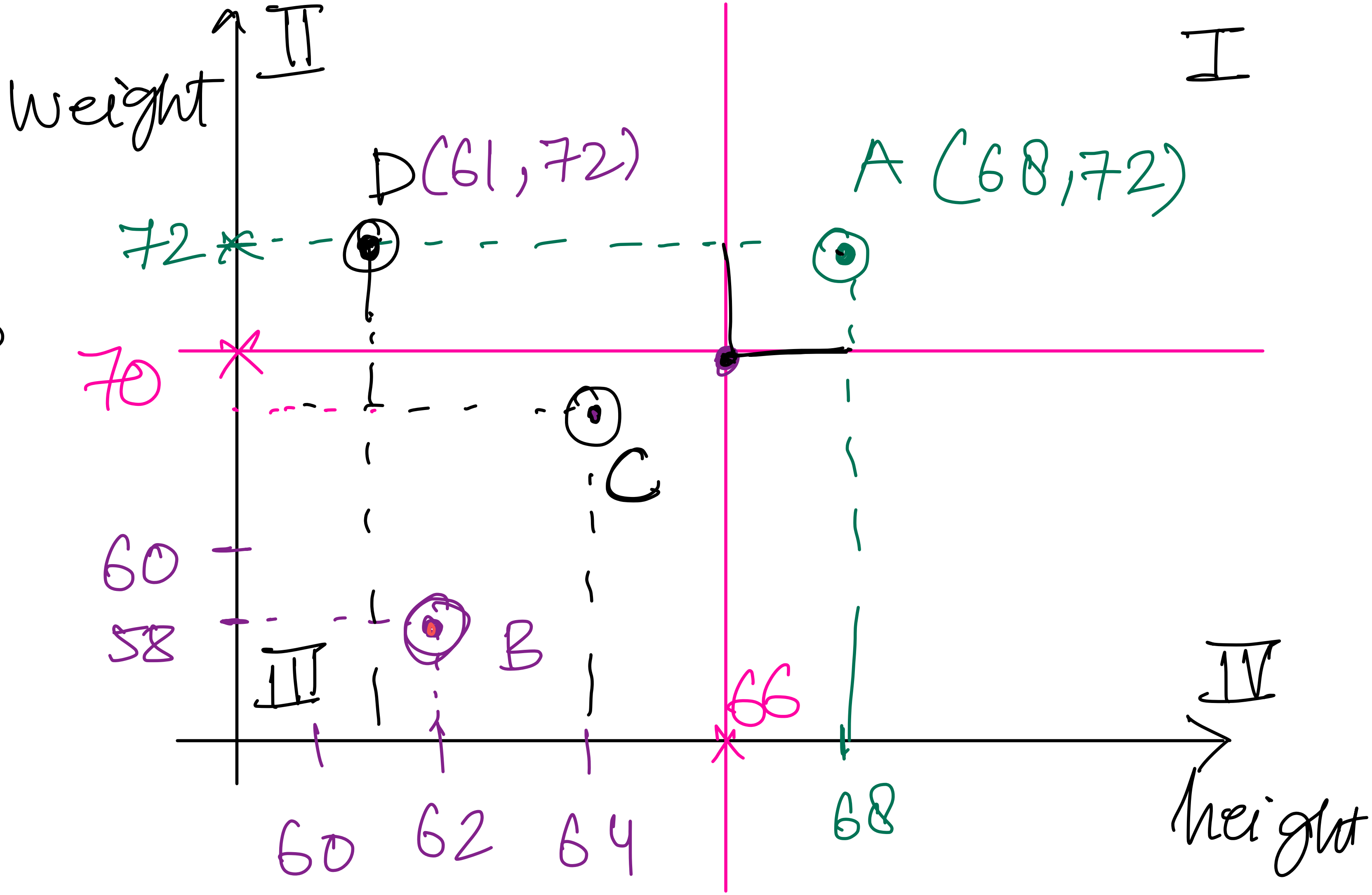
(Numerical var.)

Loan Amt

(Numerical (Var))

If associated how much ?

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



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

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

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

$(61-66)\times(72-70) = (-5)(+2) = -10$

$(4+48+6-10+ \dots \dots +60)$

$$\frac{(4 + 48 + 6 - 10 - - - - -)}{10}$$

Covariance **

trend : +ve

↑ly relationship b/w var

-ve

↓ly trend.

$$\text{Cov}(x, y) = \frac{\sum_{i=1}^{10} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n}$$

$1 \rightarrow (x_1, y_1)$

$2 \rightarrow (x_2, y_2)$

3

$\vdots$

10

$$\left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} (x_i, y_i)$$

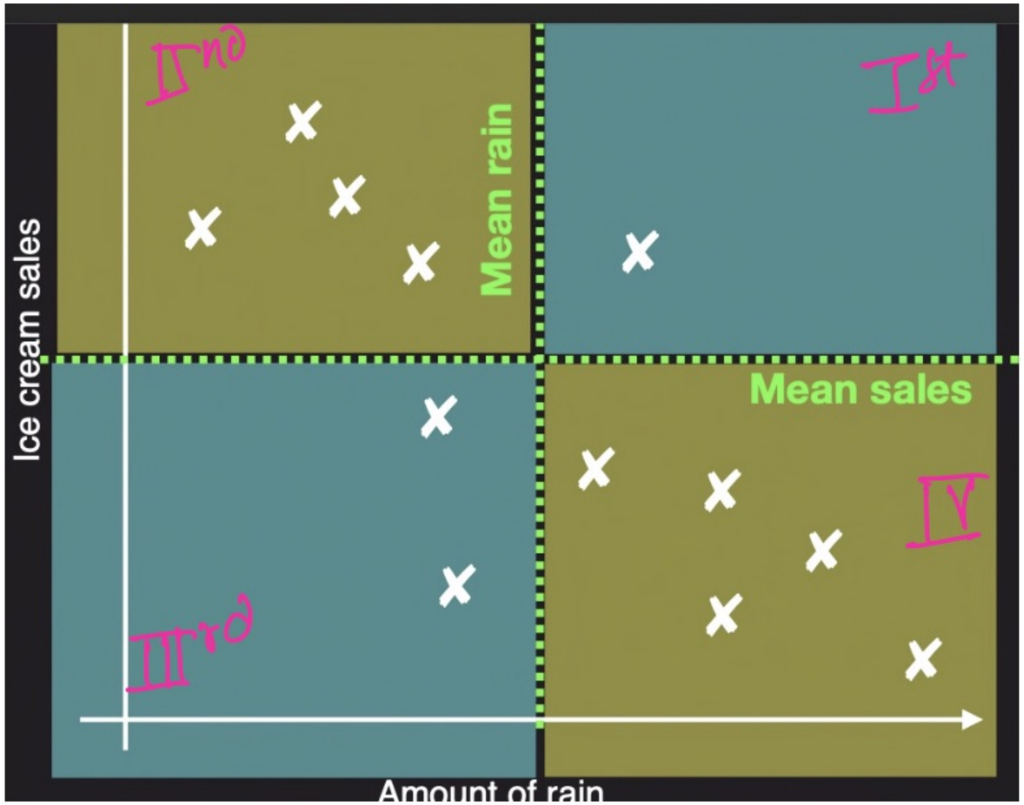
$n = 10$

Cov +ve

$\rightarrow$ I & III

more pts

Icecream v/s Rain :



Covariance

+ve

-ve

more pts is IInd & IVth

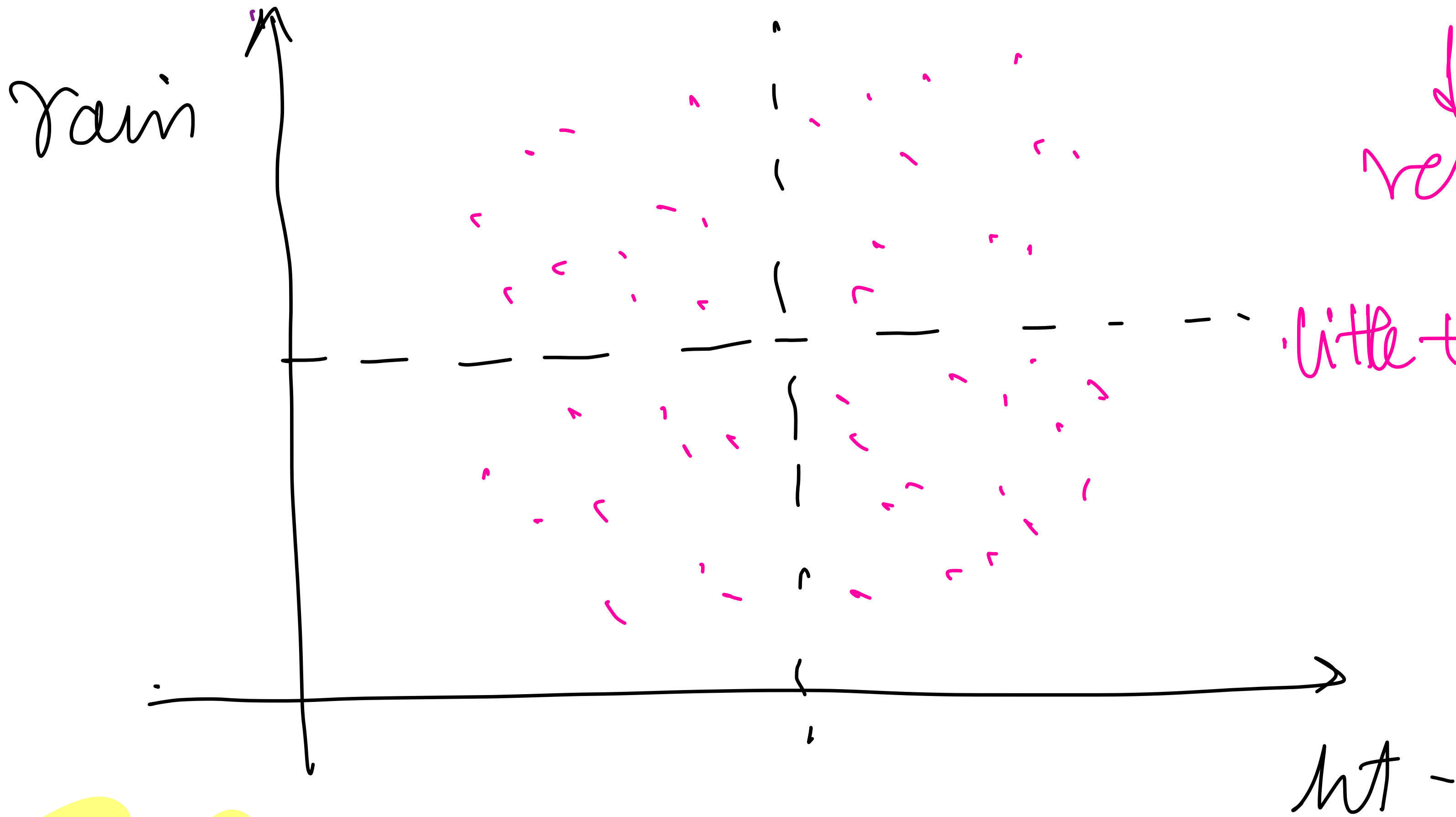
Magnitude of Cov. $= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$

height v/s Amt. of Rain

Cov. $\simeq 0$

↓

very low



little +ve or −ve

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

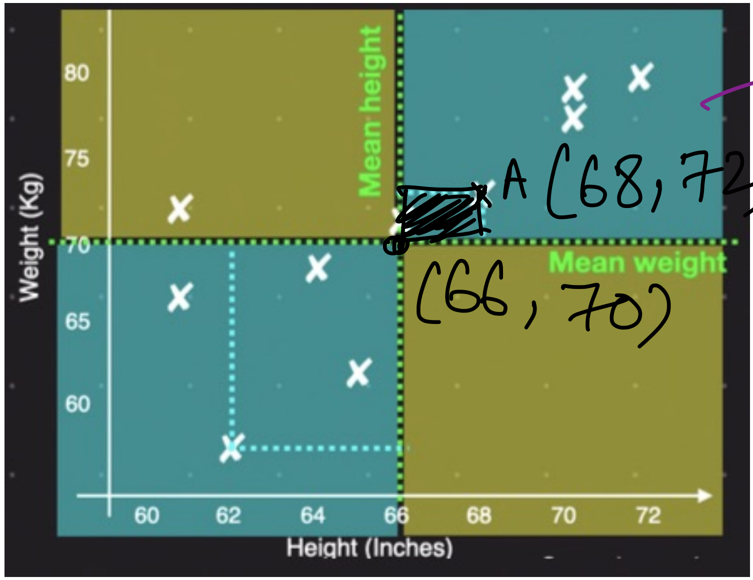
$$\frac{1}{n}\sum (x_i-\bar{x})(y_i-\bar{y})$$

Cov. → Scale or units of the variables

−1000 → Bharat.

−2000 → Vibhav.

→ Not the std. Qty. to quantify linear trend.



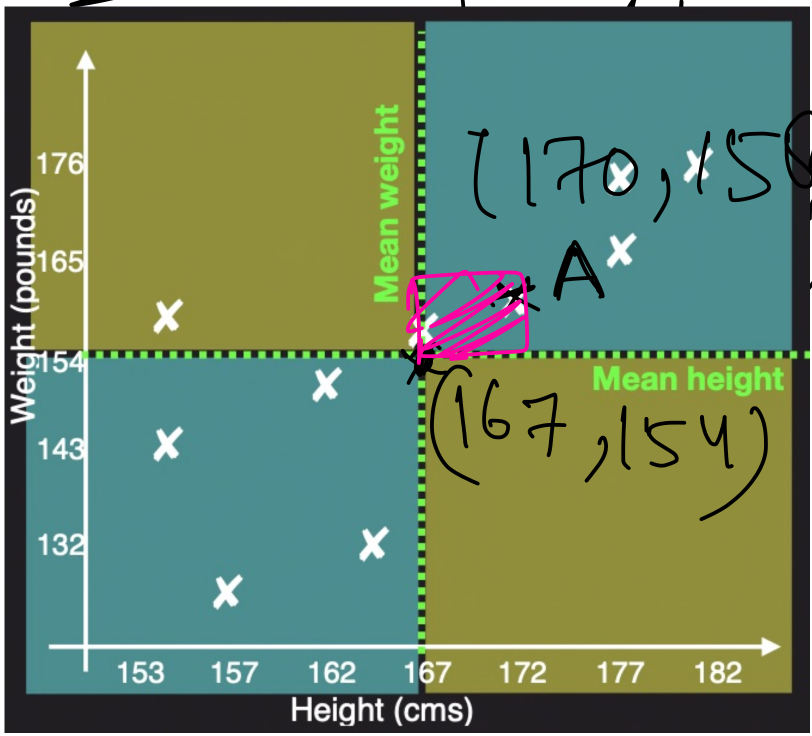
Bharat → Cov_1

$A: (68-66)\cdot(72-70)$

$= +2 \times +2 = \underline{+4}$

Sagar → weight → Pounds → 1 Kg = 2.2 pounds

→ cm → 1 inch = 2.5 cm.



Sagar → Cov_2

$A = (68 \times 2.5\ cm,$
$72 \times 2.2\ pounds)$

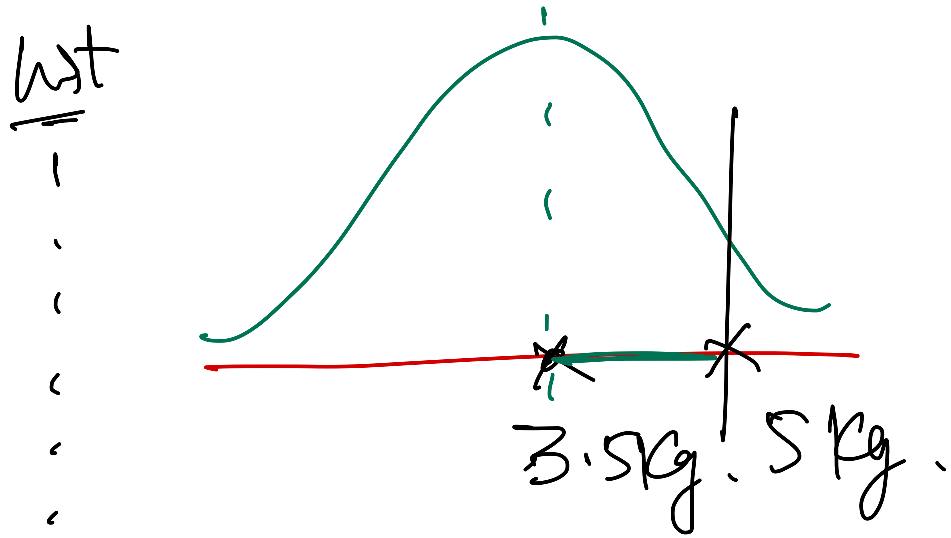
$A = (170\ cm, 158.4\ pounds)$

$$\text{Area}(A) = (170 - 167) \times (158.4 - 154)$$

$$= 3 \times 4.4$$

$$= 13.2$$

* **Pediatretian** : (wt, ht, health)



5kg

Z score

$\downarrow$

+1.1

Z score = 0

Z score = −1.2

$$Z_w = \left(\frac{w - \bar{w}}{\sigma_w}\right) \rightarrow \boxed{Z_w = 0}$$

$w_{child} = \bar{w}_{avg.}$

$\downarrow$ 3.5 $\quad$ $\downarrow$ 3.5kg

avg. wt. baby

$Z_w = 0 \rightarrow$ avg. wt. baby

$Z_w \rightarrow +1.1 \rightarrow$ overweight.

$Z_w \rightarrow -1.5 \rightarrow$ underweight.

std. qty

*

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

not std. qty.

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

Correlation

$\rightarrow S_{xy}$

Pearson Correlation

Cov. $\rightarrow \in (-\infty \text{ to } \infty)$

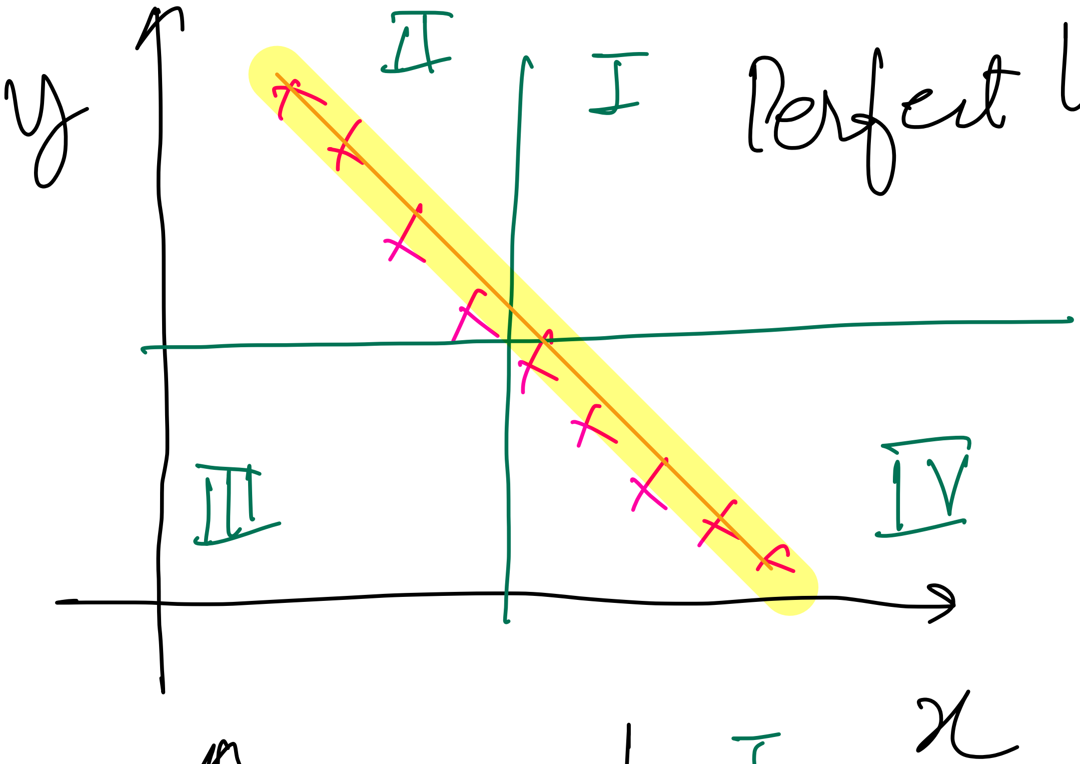
Correlation $\in (-1 \text{ to } +1)$

Corr. → ht & wt → +0.8

Corr. → ht & IQ → +0.7

more stronger linear relationship

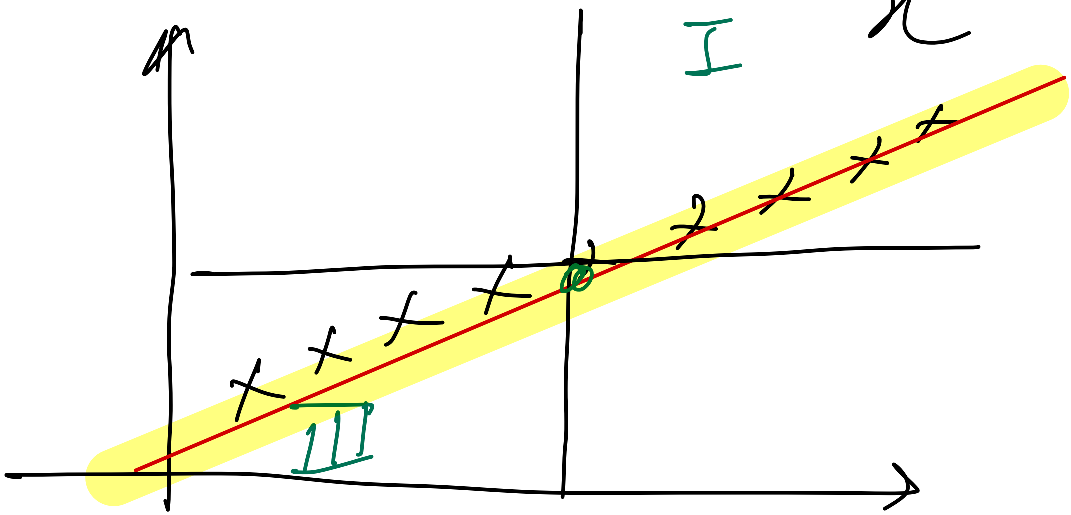
Strength of linear relationship : Correlation



Perfect linear relationship

→ 0 Covar → min

Corr. −1



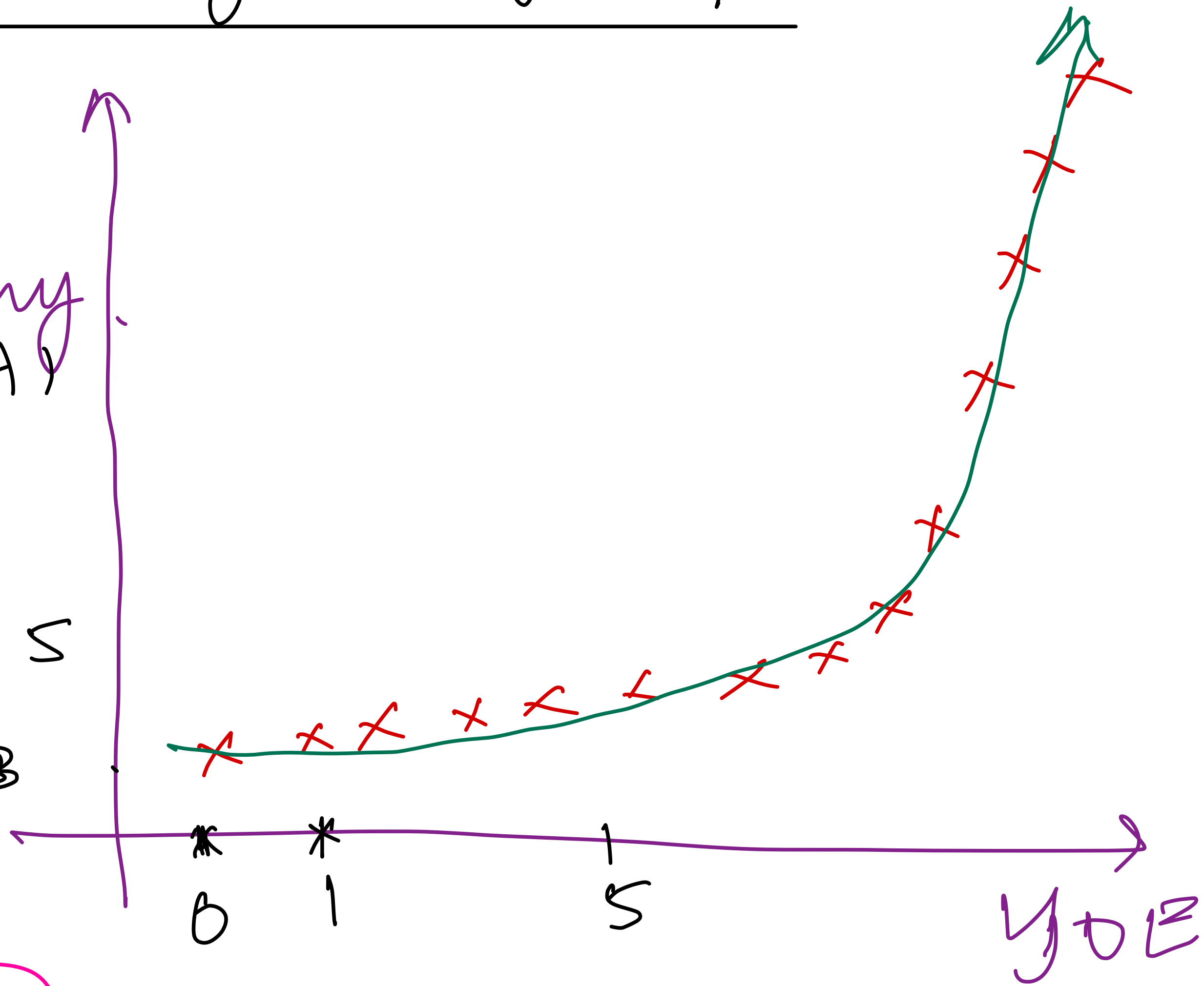
Covar → max

Correlation +1

* Salary grows wrt. years of Exp. :

linear : Y / N

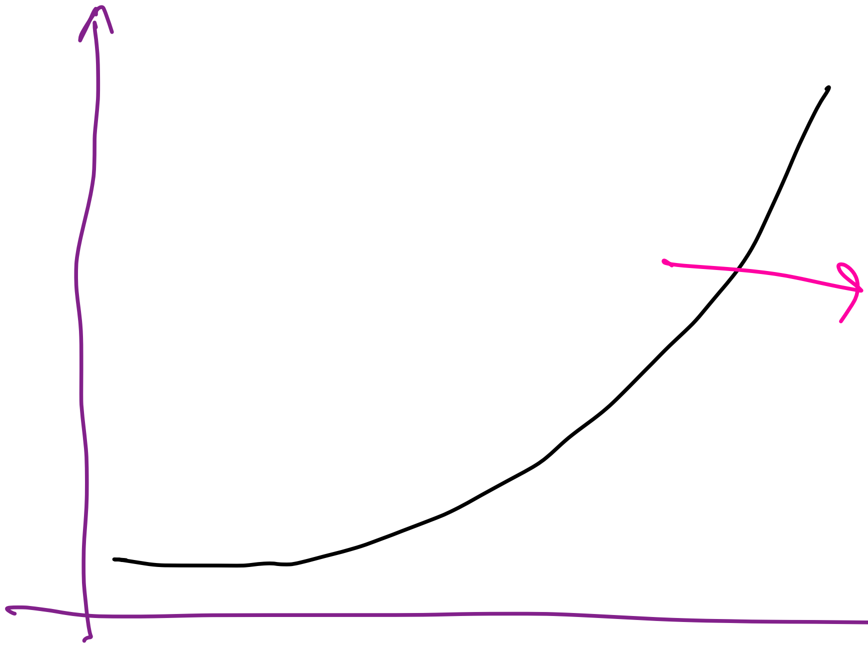
monotonically increasing.



* Correlation → +1 ?

Strength of linear relationship.

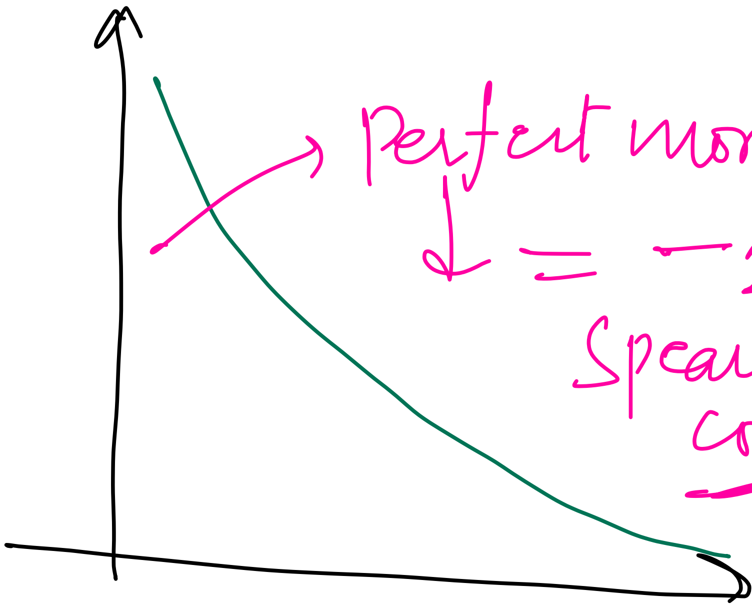
Perfectly monotonic.



Y

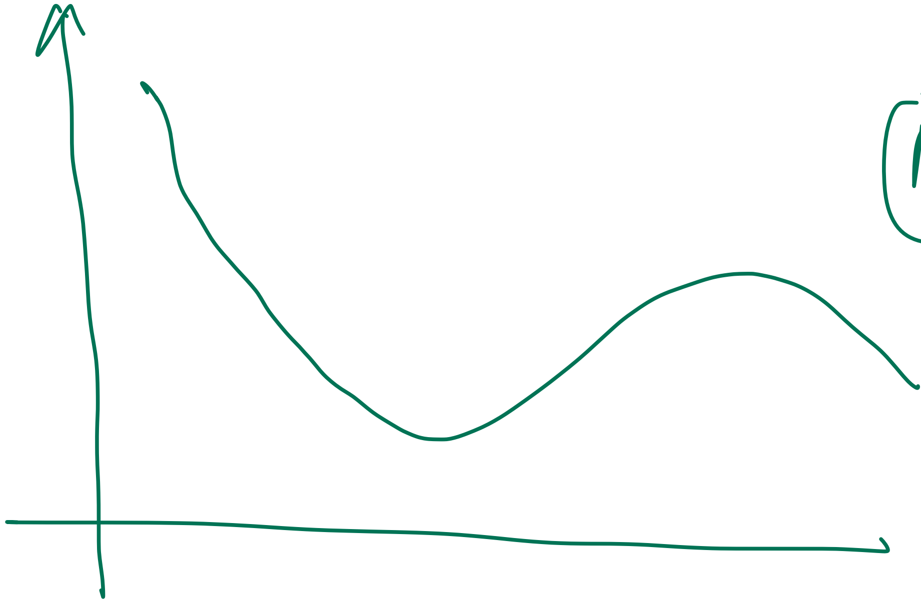
spearman corr = +1

Y

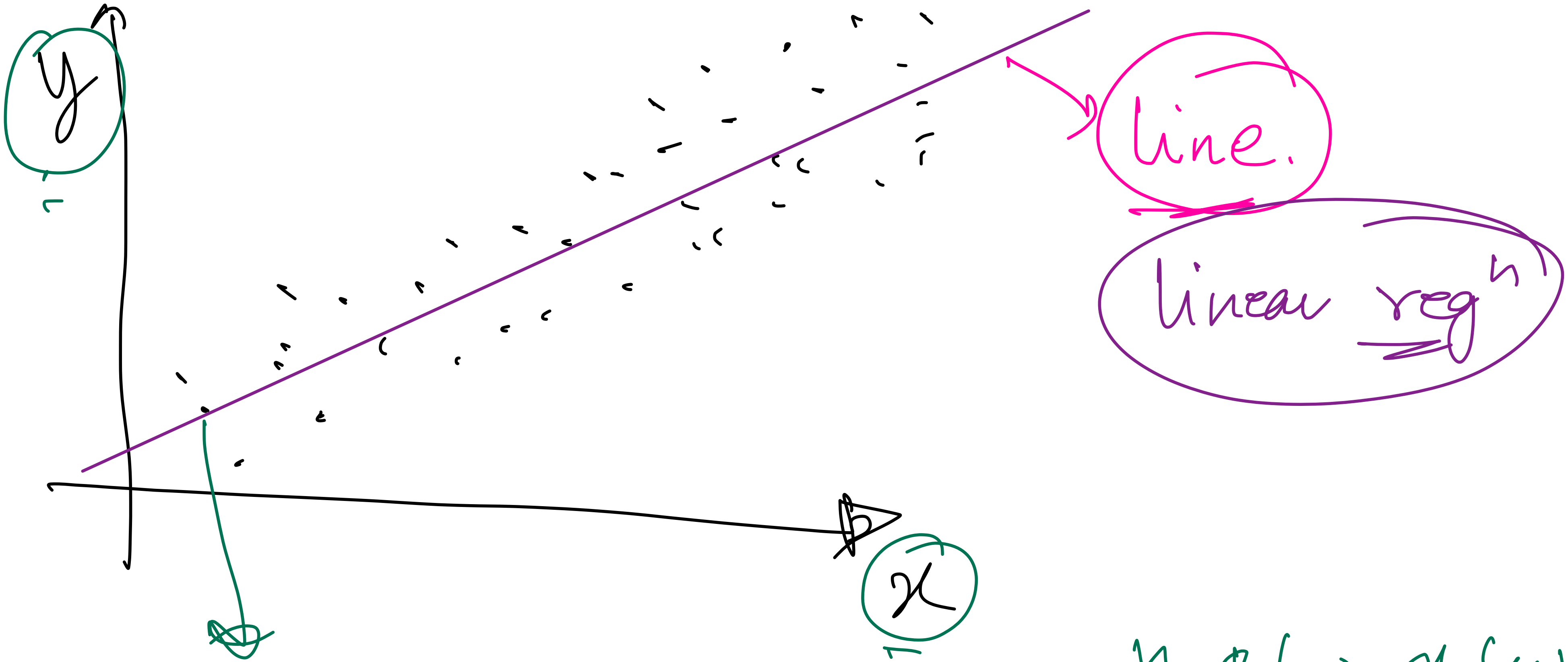


Perfect monotonic = −1

Spearman corr



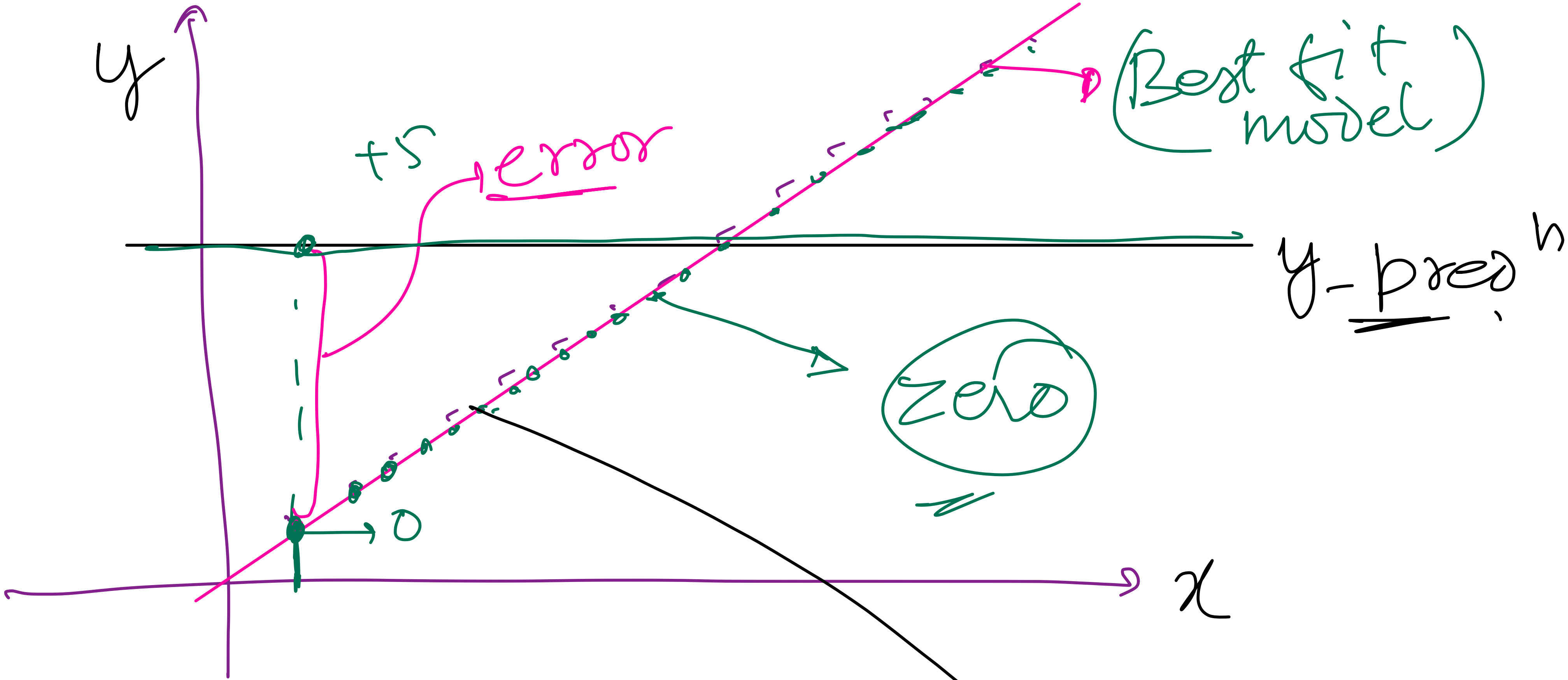
N



Strength of linear relp^n b/w x & y.

0.82

$\Rightarrow R^2$

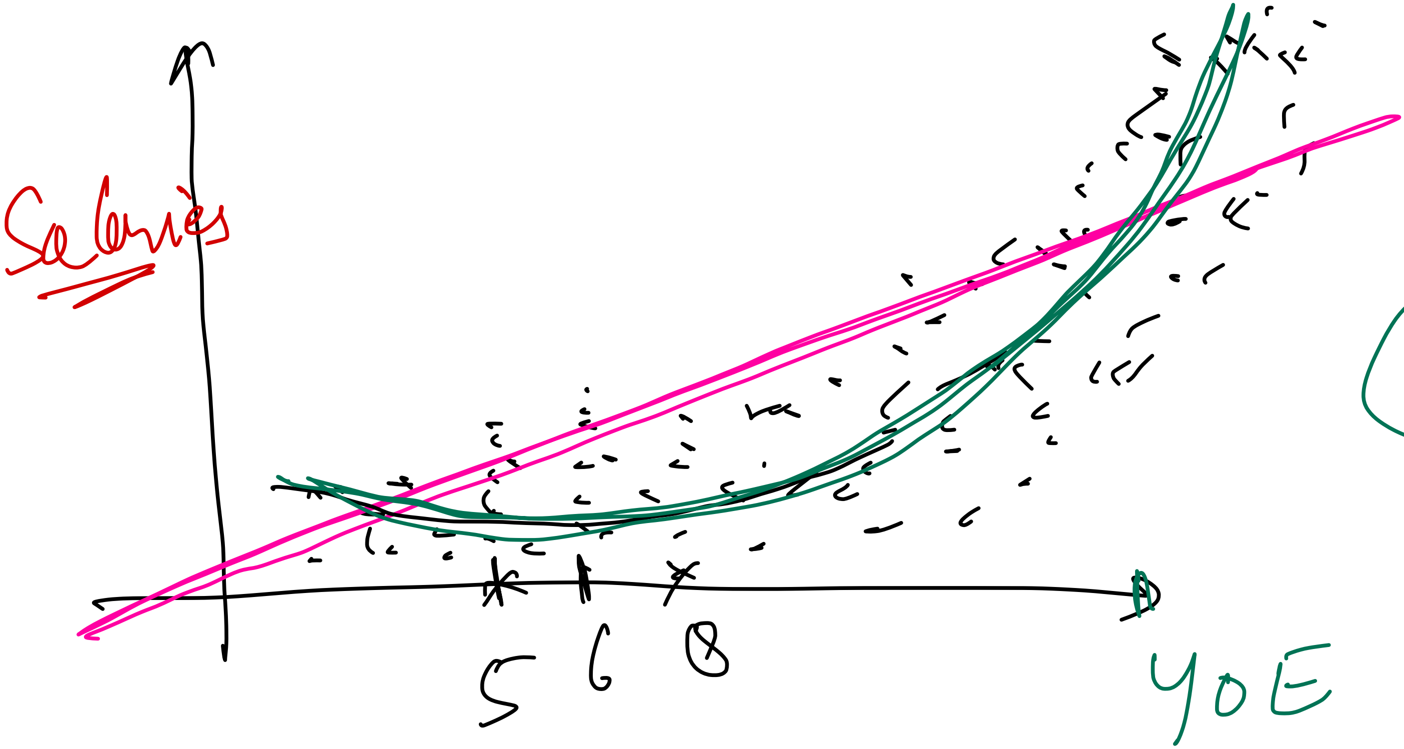


$x \longrightarrow$ predict y

Correlation for Best fit line

+1

100%



Naukri.com

Pearson ← Spearman

New profile (YoE) → MODEL → Expected Salary