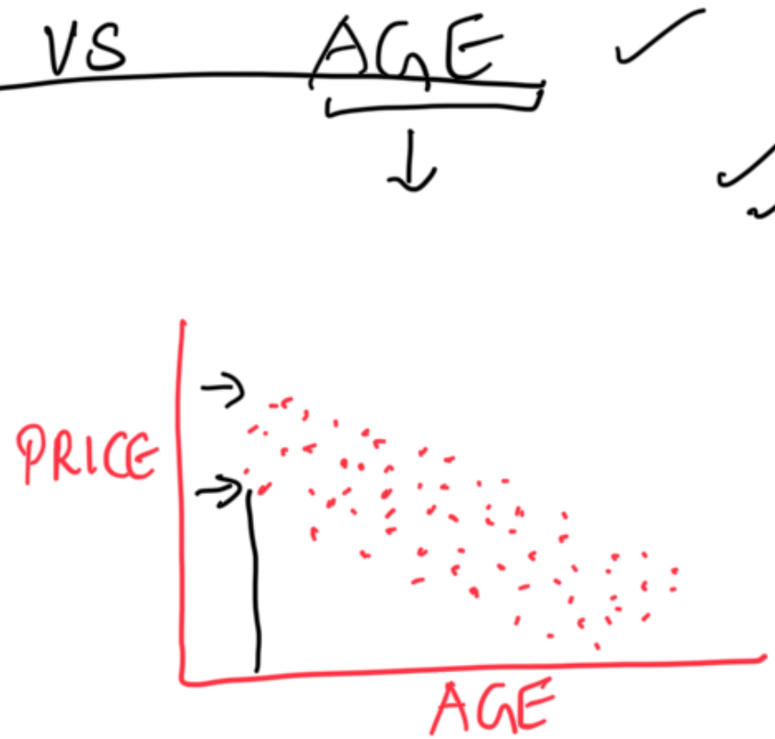
CORRELATION

CAR-PRICE VS AGE

relationship :



CAR-PRICE VS HORSE-POWER

CAR 2
3
4 1

scatter

CORRELATION

↓

2 - numeric variables.

Parametric vs Non - Parametric test

↓

[underlying assumption about the distribution]

Z - test

↓

fewer or no assumption

-> KS - test

-> ch - sq

-> Kruskal - Wallis

t-test

ANOVA

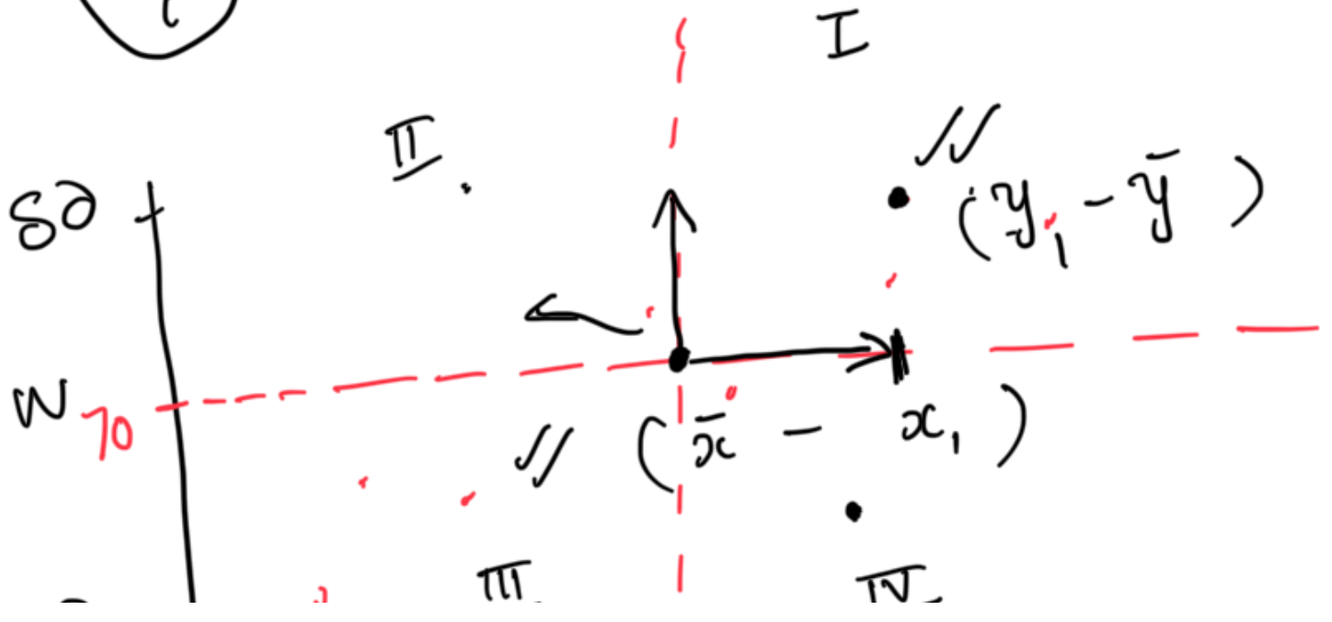
→ Leveme

Correlation

Correlation does not imply causation

// **Covariance** :

$$\sum_i (h_i - \bar{h})\,(w_i - \bar{w})$$

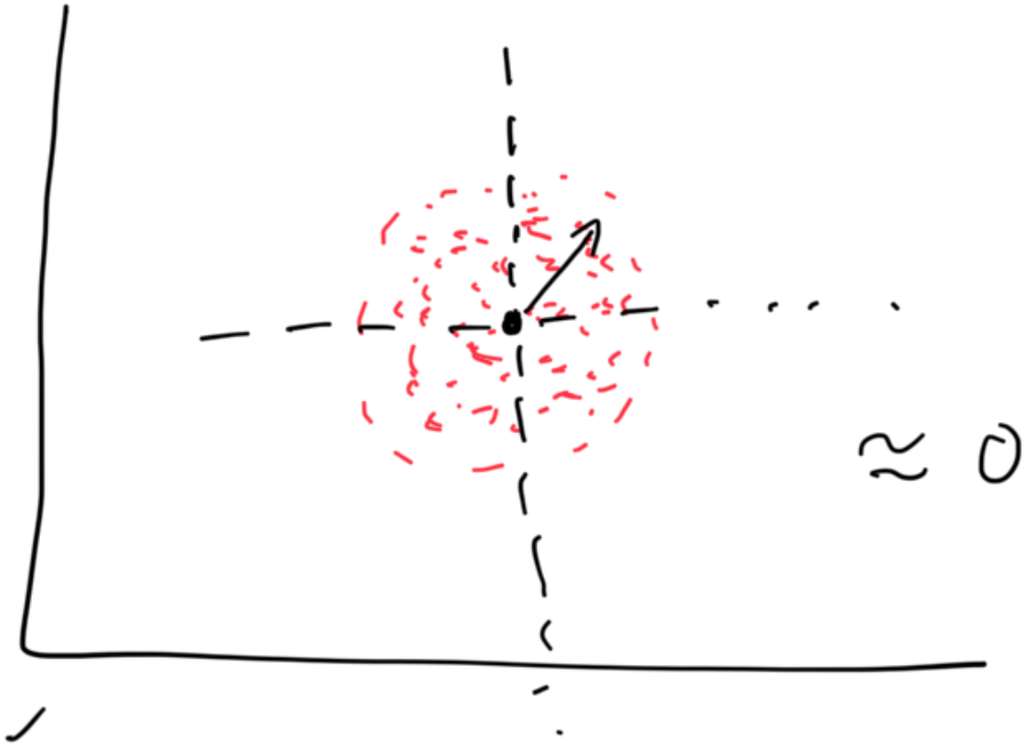


60

60 64 68 72

+ve correlated

-ve correlated



1000 1000

$$\text{cov} \quad \sum \frac{(x_i - \bar{x})}{\sigma_x} \frac{(y_i - \bar{y})}{\sigma_y}$$

$$\text{Pearson - correlation} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}$$

cov: +ve individual with above avg height seem to have above average weight.

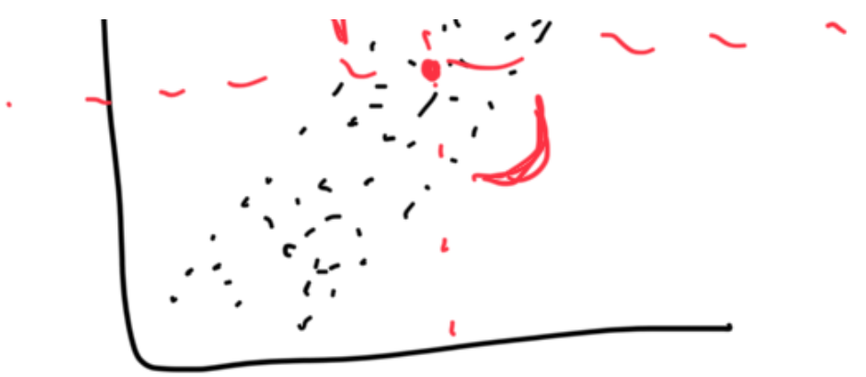
-ve individual with above avg height seem to

individuals ... have below average weight

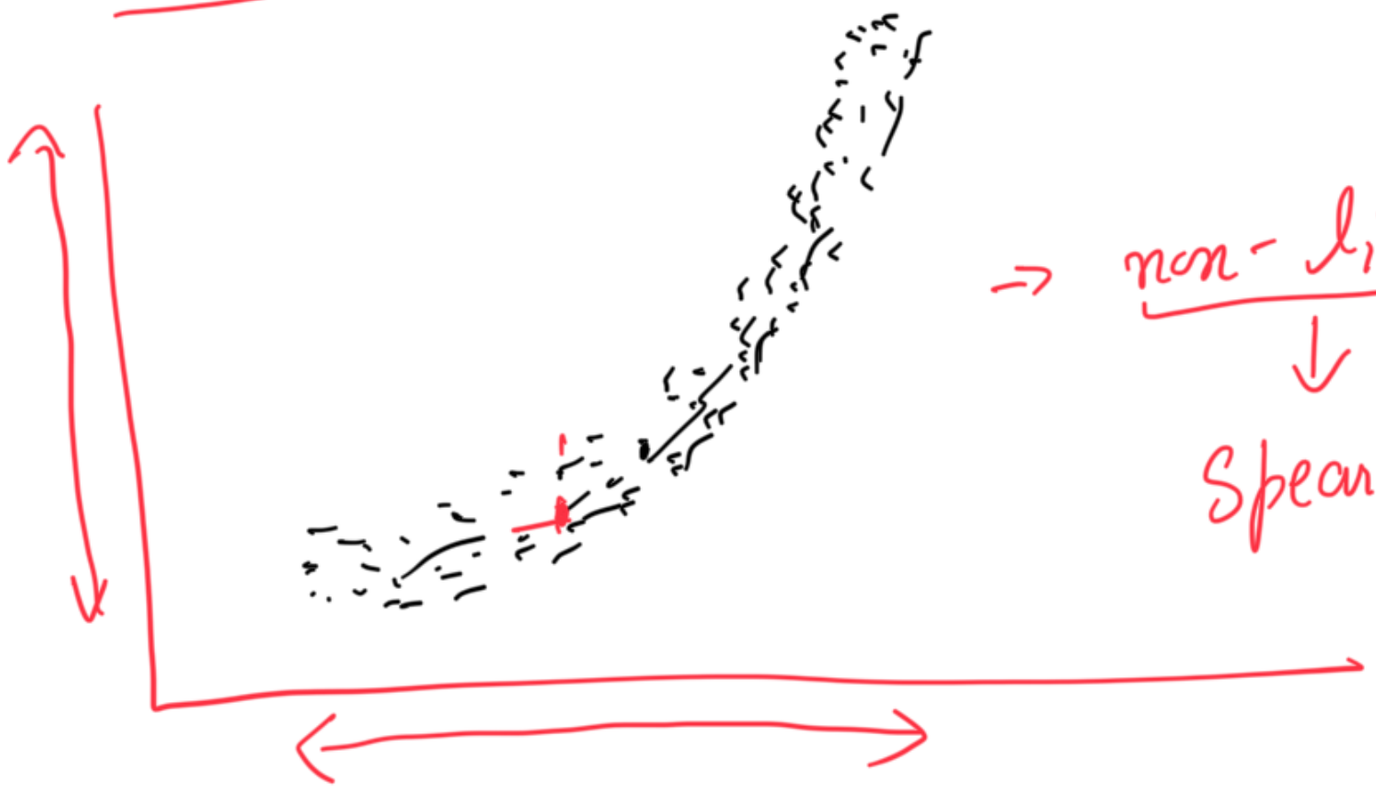
cov ~ profit : cannot be compared.

correlation : strong +ve correlation

means as height increases, weight increases proportionally.



Pearson - Correlation : linear correlation



→ non-linear relationship.

↓

Spearman's correlation

Spearman :

monotonically increasing.

rank : 1 2 3 4 5 6 - - - - -

H : 60 50

W : 45 47

1 2 3

	Rank H	Rank W
(H_1, W_1)	2	4
(H_2, W_2)	3	2

$$(r_h - r_w)^2$$

14 x 14

13 x 13

12 x 12

11 x 11

10 x 10

9 x 9

8 x 8

7 x 7

6 x 6

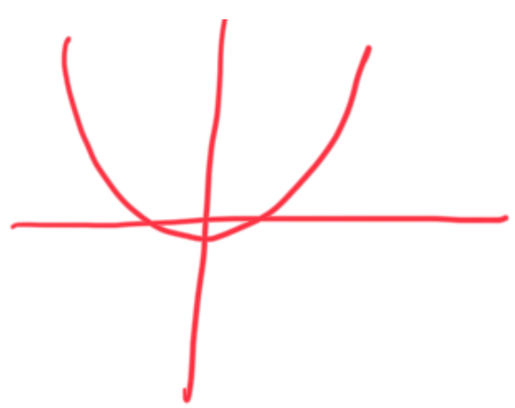
5 x 5

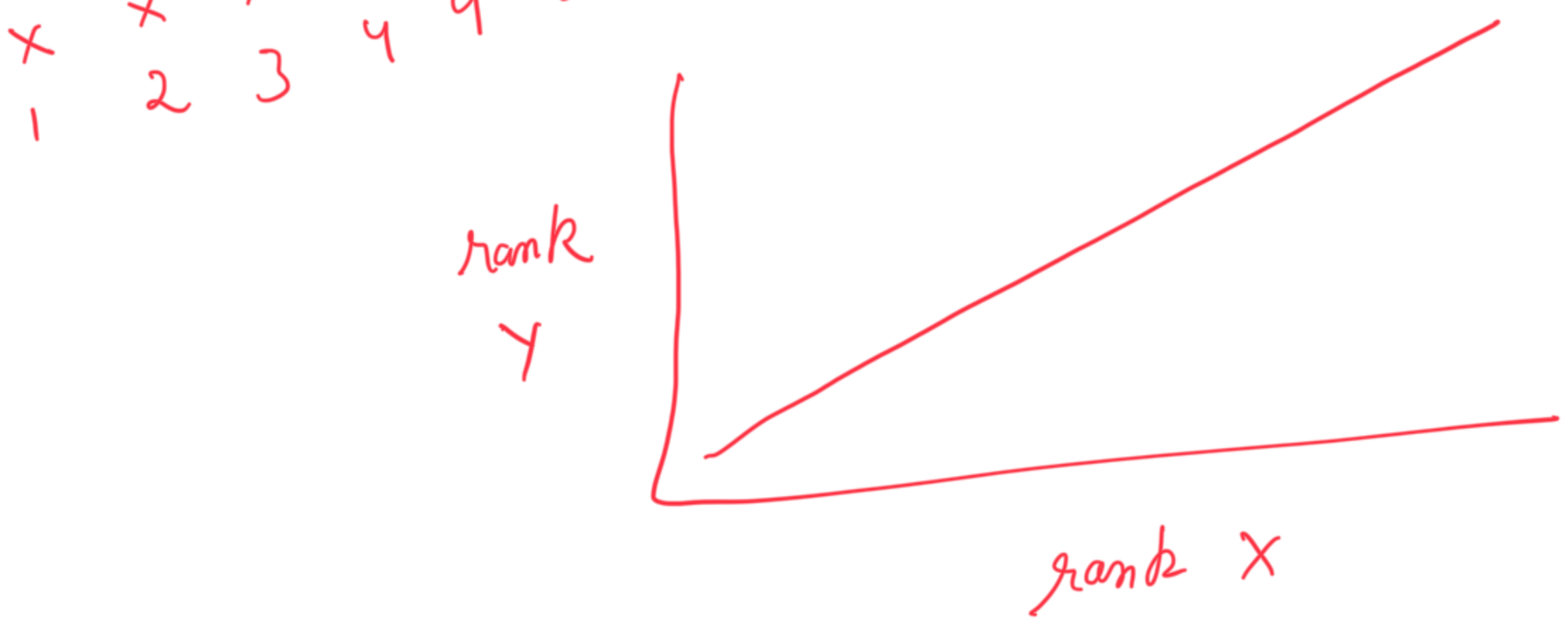
4 x 4

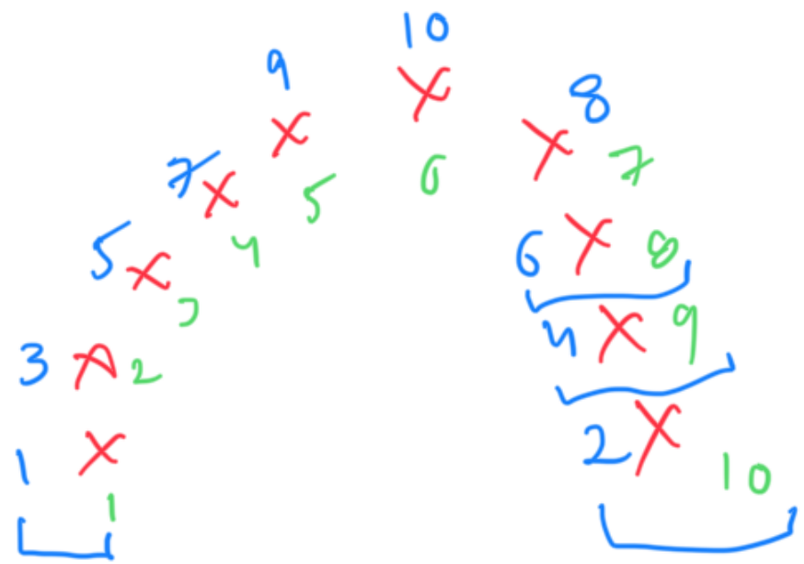
3 x 3

2 x 2

1 x 1







$$\sum d_i^2$$

OKAM'S RAZOR

paper cutter kitchen-knife sword