X: [1, 2, 3, 4, 5]

Z: [10, 8, 6, 4, 2]

$$\mathrm{COV}(X, Z) = \frac{\sum (x-\bar{x})(z-\bar{z})}{N-1}$$

$$= \frac{(1-3)(10-6)}{5-1} = \frac{-2 \times 4}{4} = -2$$

negative, variables are moving in opp. direction square $\leftarrow$ COVARIANCE MATRIX $\rightarrow$ symmetric

Data: f_1 f_2 f_3

$$cov(f_1, f_1) = \frac{\sum (f_1 - \bar{f}_1)(f_1 - \bar{f}_1)}{N-1}$$

$$= \frac{\sum (f_1 - \bar{f}_1)^2}{N-1} = var(f_1)$$

$$3 \left\{ \begin{array}{c|ccc} & f_1 & f_2 & f_3 \\ \hline f_1 & var(f_1) & cov(f_1, f_2) & cov(f_1, f_3) \\ f_2 & cov(f_2, f_1) & var(f_2) & cov(f_2, f_3) \\ f_3 & cov(f_3, f_1) & cov(f_3, f_2) & var(f_3) \end{array} \right.$$

3

* $cov(A, B) = cov(B, A)$

$$cov(A, B) = \frac{\sum (A - \bar{A})(B - \bar{B})}{N-1}$$

$$cov(B, A) = \frac{\sum (B - \bar{B})(A - \bar{A})}{N-1}$$

$$\text{corr} = \frac{\text{cov}}{\sigma_x \sigma_y} \Rightarrow \frac{\sum (x-\bar{x})(y-\bar{y})}{\sigma_x \sigma_y} \rightarrow \text{PCC} \Rightarrow +1 \text{ or } -1$$

Correlation

$[+1, -1]$

- While covariance measures the direction of a relationship between two variables, correlation measures the strength of that relationship. This is usually expressed through a correlation coefficient(r), which can range from -1 to +1.

$$Correlation = \frac{Cov\,(x, y)}{\sigma x * \sigma y}$$

linear

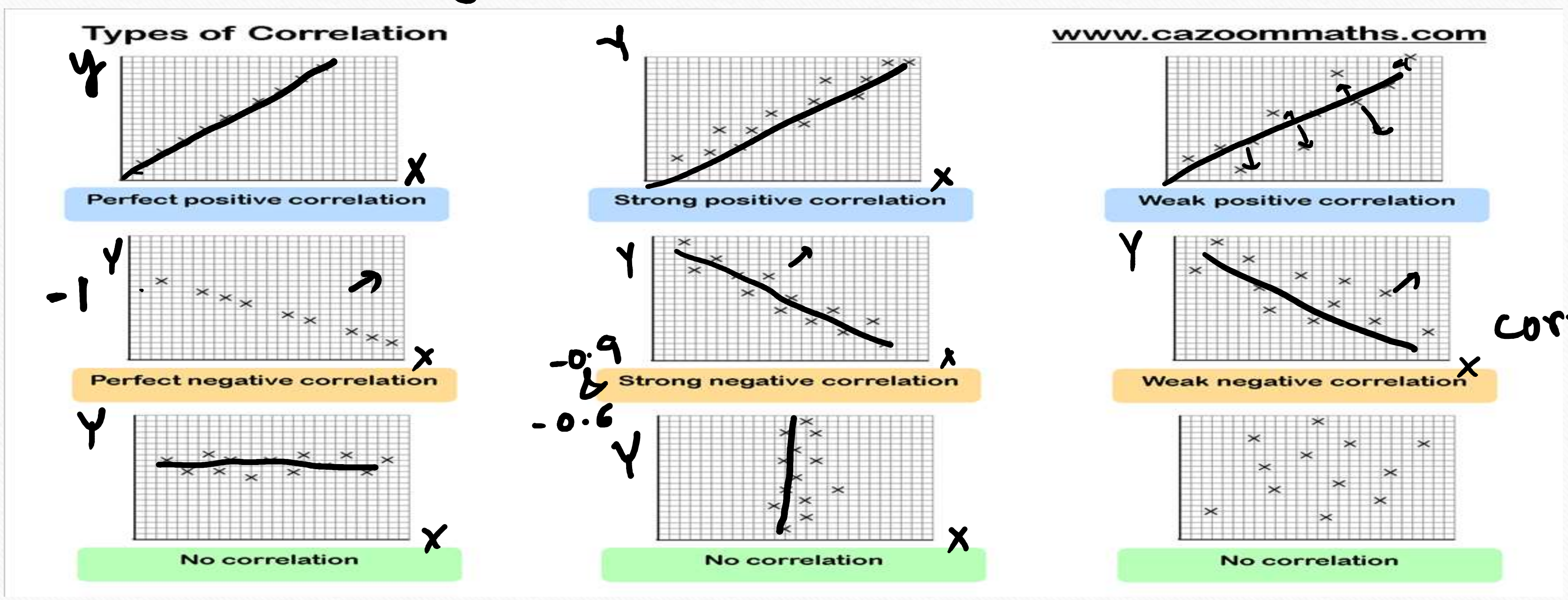
Correlation Coefficient

Pearson's Correlation Coefficient

Spearman's Correlation Coefficient → nonlinear

Correlation

corr= +1 corr= 0.9-0.6 corr<0.6



corr> -0.6

"CORRELATION NEVER MEANS CAUSATION"

Pearson's Correlation Coefficient

- It is used in linear applications.
- Dataframe.corr(method = 'pearson')

cov(x,y)

$$r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1) S_X S_Y}$$

$\frac{\text{Total}}{(N-1)\, S_x S_y}$

$6-1$ (5)

$S_x = 8.28$

$S_y = 32.91$

$\Sigma (x_i - \bar{x}_i) \times (y_i - \bar{y}_i)$

↳ Total

Question to Play!

$\bar{x}_i = 12.5$ $\bar{y}_i = 68$

A TEACHER WANTS TO DETERMINE THE CORRELATION BETWEEN THE NUMBER OF HOURS SPENT STUDYING AND TEST SCORES.

STUDENT NAME	x_i	y_i	$\bar{x}_i$	$\bar{y}_i$	$\bar{x}-x_i$	$\bar{y}-y_i$
JOHN	13	53				
ALLIE	15	69				
MARK	7	92				
SAMANTHA	3	10				
JESSICA	10	85				
JOSEPH	27	99				
	S_x	S_y				

$$S_x = \sqrt{\frac{1}{N-1} \Sigma (x_i - \bar{x}_i)^2}$$

Solution!!

0.9 and 0.6

0.9–0.7 – 0.6

strong

moderately strong

$$r = \frac{1}{(6-1)s_x s_y}\left[\; 821 \;\right] = \frac{1 \times 821}{5 \times 8.28 \times 32.91} = 0.6$$

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
13	53	0.5	– 15	– 7.5
15	69	2.5	1	2.5
7	92	– 5.5	24	– 132
3	10	– 9.5	– 58	551
10	85	– 2.5	17	– 42.5
27	99	14.5	31	449.5
$\bar{x} = 12.5$	$\bar{y} = 68$			SUM = 821
$s_x = 8.28$	$s_y = 32.91$			

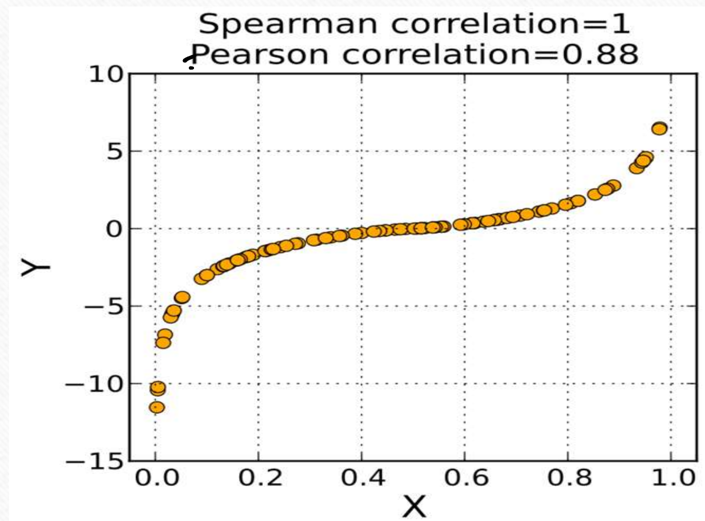
d = diff. b/w ranks Ranks → contextual

Spearman's Correlation Coefficient

- It is used in non-linear application.
- Dataframe.corr(method = 'spearman')

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

d



$d \rightarrow d^2$

Question to Play!

The scores for 10 students in English and Maths are as follows. Compute Spearman's Coefficient.

Ranks

	Marks									
English	56	75	45	71	62	64	58	80	76	61
Maths	66	70	40	60	65	56	59	77	67	63

Solution!!

n=10

English (mark)	Maths (mark)	Rank (English)	Rank (maths)	d	d^2
56	66	9	4	5	25
75	70	3	2	1	1
45	40	10	10	0	0
71	60	4	7	3	9
62	65	6	5	1	1
64	56	5	9	4	16
58	59	8	8	0	0
80	77	1	1	0	0
76	67	2	3	1	1
61	63	7	6	1	1

$$\sum d_i^2 = 25 + 1 + 9 + 1 + 16 + 1 + 1 = 54$$

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

$$\rho = 1 - \frac{6 \times 54}{10(10^2 - 1)}$$

$$\rho = 1 - \frac{324}{990}$$

$$\rho = 1 - 0.33$$

$$\rho = 0.67$$