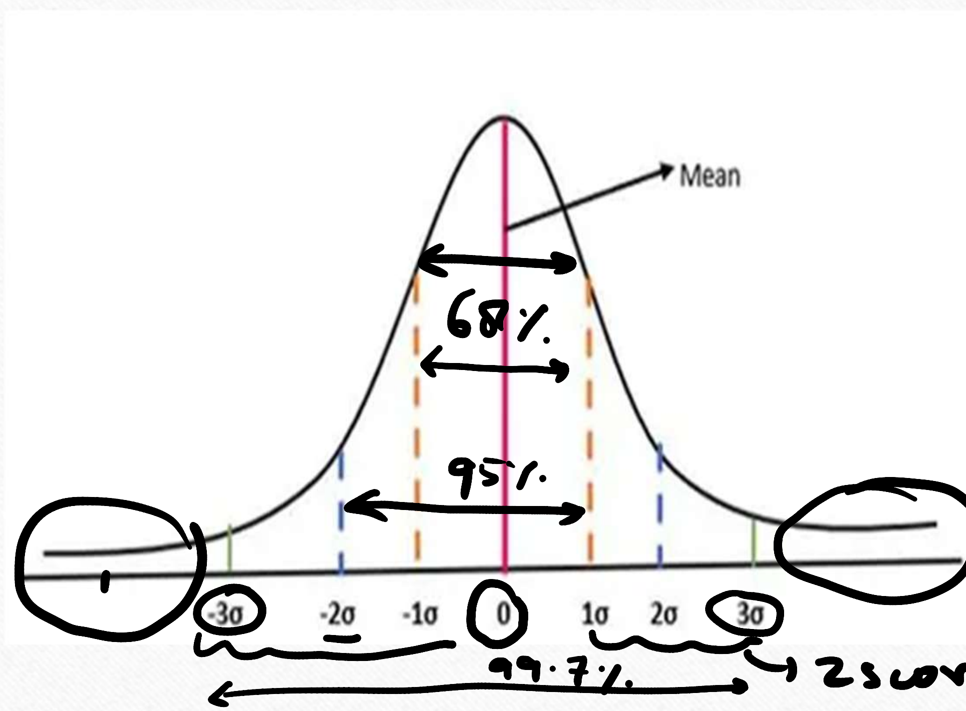
3σ Normal +3σ

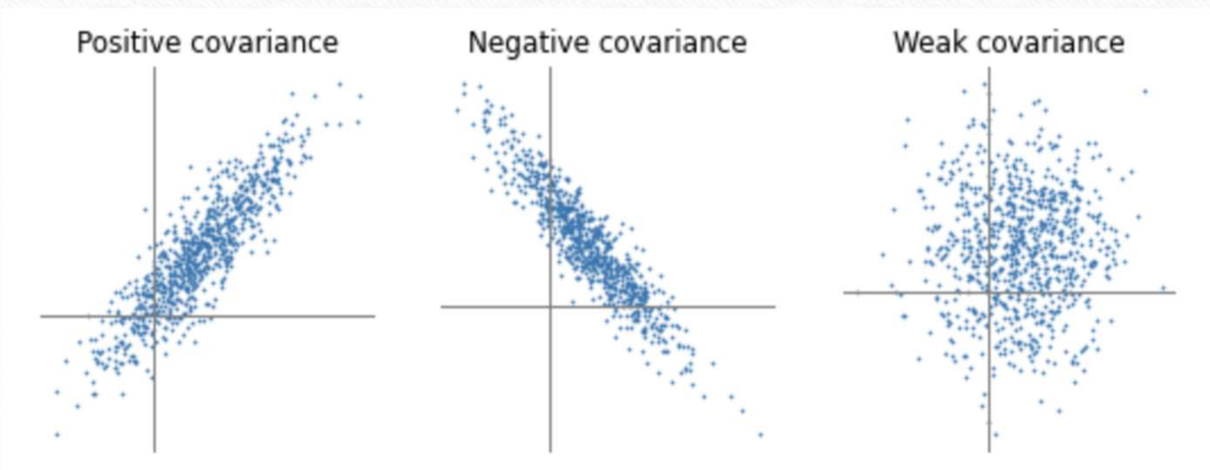
Z-Score Or Standard Score

- Standard scores work by transforming sets of data into a new, theoretical distribution with a mean of 0 and a standard deviation of 1.
- Standard scores can take any value, and they indicate position relative to the mean. Positive z-scores mean that your value is above the mean, and negative z-scores mean that your value is below it. If your z-score is 0, then your value is the mean itself. The size of the number shows how far away the value is from the mean.



Co-Variance

- Covariance measures the directional relationship between the returns on two assets. A positive covariance means asset returns move together, while a negative covariance means they move inversely.



$$COV(X,Y) = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{n-1}$$

$X : [2 \quad 4 \quad 6 \quad 8 \quad 10] \qquad \bar{X} = 6 \qquad X - \bar{X}$

$Y : [1 \quad 2 \quad 3 \quad 4 \quad 5] \qquad \bar{Y} = 3$

$(2-6) \quad (4-6) \quad (6-6) \quad (8-6) \quad (10-6)$

$X - \bar{X} : \quad \underbrace{-4 \quad -2} \quad 0 \quad +2 \quad +4$

$Y - \bar{Y} : \quad -2 \quad -1 \quad 0 \quad +1 \quad +2$

$(1-3) \quad (2-3) \quad (3-3) \quad (4-3) \quad (5-3)$

$8 \quad 2 \quad 0 \quad 2 \quad 8$

$$\frac{(X-\bar{X})(Y-\bar{Y})}{N-1} = \frac{8+2+0+2+8}{n-1\,(5-1)}$$

$$= \frac{20}{4} = (+)\,5^{x}$$

$(-)$

$X = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \end{bmatrix} \quad \bar{X} = 3$

$= \begin{bmatrix} 10 & 8 & 6 & 4 & 2 \end{bmatrix} \quad \bar{Y} = 6$

$X - \bar{X} : \quad \underbrace{-2 \quad -1} \quad 0 \quad \underbrace{+1 \quad +2}$

$Z - \bar{Z} : \quad \underbrace{+4 \quad +2} \quad 0 \quad \underbrace{-2 \quad -4}$

$-8 \quad -2 \quad 0 \quad -2 \quad -8 \quad =$

$\frac{\Sigma}{n-1}$

$= (\frac{1}{\ } \ 5$

$$\frac{\sum (x-\bar{x})(z-\bar{z})}{n-1} = \frac{-8-2-0-2\quad 8}{4}$$

$$= -5$$

$$Cov(X,Y) = cov(Y,X)$$

square matrix

$cov(F_1, F_1)$

metric
matrix

COVARIANCE MATRIX

$$= \frac{\Sigma (f_1 - \bar{f}_1)(f_1 \ldots}{N-1}$$

$$= \frac{\Sigma (\ldots}{N \ldots}$$

$$
\begin{array}{c|c}
 & \begin{array}{ccc} f_1 & f_2 & f_3 \end{array} \\
\hline
\begin{array}{c} f_1 \\ f_2 \\ f_3 \end{array} &
\begin{bmatrix}
Var(f_1) & cov(f_2\, f_1) & cov(f_3\, f_1) \\
cov(f_1\, f_2) & Var(f_2) & cov(f_3\, f_2) \\
cov(f_1, f_3) & cov(f_2\, f_3) & Var(f_3)
\end{bmatrix}
\end{array}
$$

CORRELATION NEVER MEANS CAUSATION

Correlation

→ kms
y→ hours

(−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.

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

$$\frac{km \ hours}{km \ hours} = 1$$

Correlation Coefficient

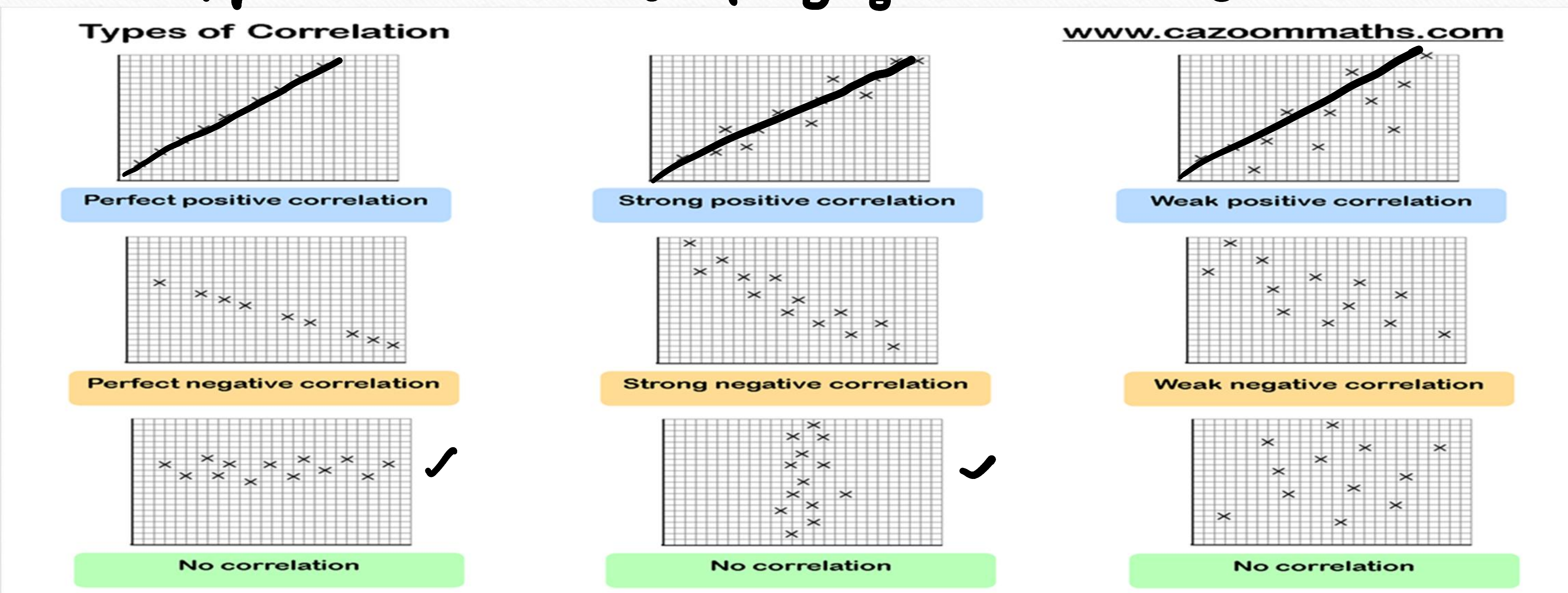
- Pearson's Correlation Coefficient
- Spearman's Correlation Coefficient

+1 → perfect +ve   -1 → perfect -ve

+1   -1

Correlation

+1   0.9 – 0.6   0.6 <



Pearson's Correlation Coefficient

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

cov.

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

$$r = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{(N-1)\, S_x\, S_y}$$

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

Question to Play!

"10 minute"

std dev. of sample

$S_x = 8.29$ $S_y = 32.91$

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

STUDENT NAME	x_i	y_i
JOHN	$(0.5)^2$ 13 -12.5	53
ALLIE	$(2.5)^2$ 15 -12.5	69
MARK	$(-5.5)^2$ 7 -12.5	92
SAMANTHA	$(-9.5)^2$ 3 -12.5	10
JESSICA	$(-2.5)^2$ 10 -12.5	85
JOSEPH	$(14.5)^2$ 27 -12.5	99

$(x_i - \bar{x}_i)$

$\frac{\quad}{5}$

$\frac{75}{6} = 12.5$

$13+15+\overbrace{7+3}^{10}+10+27$

$47+13+15$

62

$+13$

75

Solution!!

0.60

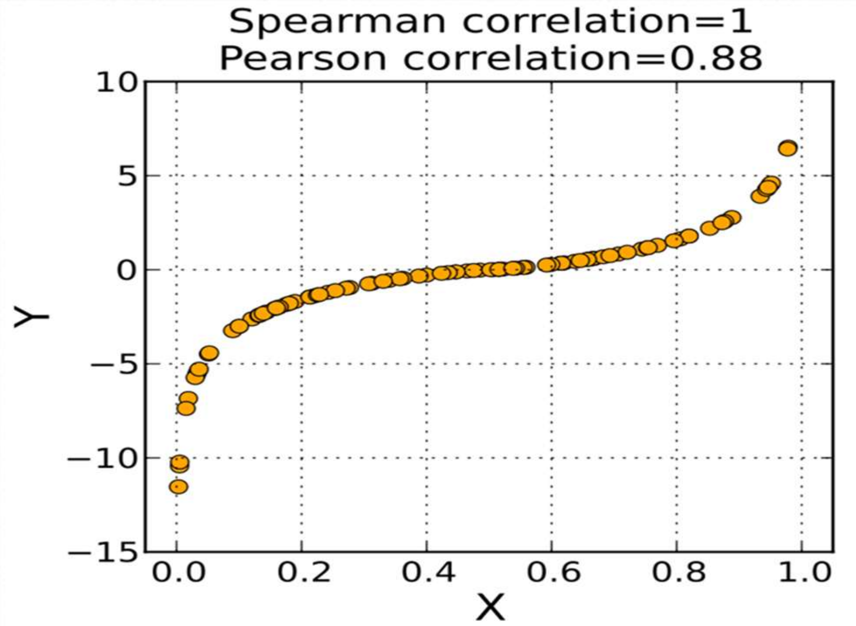
$$r = \frac{1}{(6-1)s_x s_y}\left[\ 821\ \right]$$

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
$\bar{x} = 12.5$ $s_x = 8.28$	$\bar{y} = 68$ $s_y = 32.91$			SUM = 821

Spearman's Correlation Coefficient

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

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



Rank

Question to Play!

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

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

Solution!!

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

54

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

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

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

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

$$\rho = 1 - 0.33$$

$$\rho = 0.67$$

The End