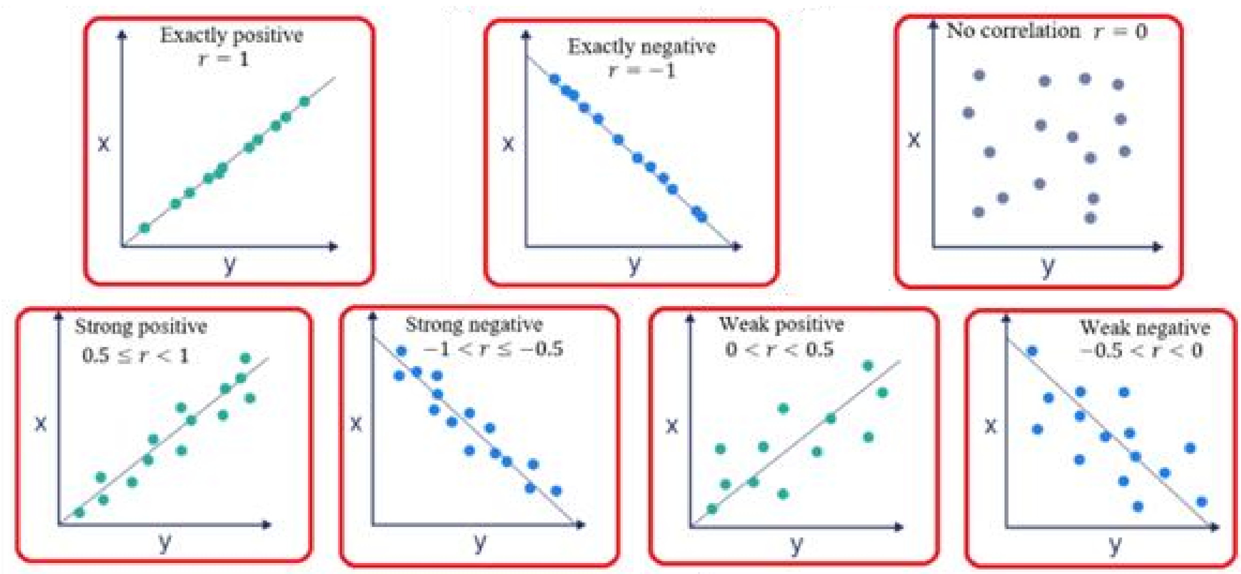
Pearson Correlation Coefficient (معامل ارتباط بيرسون)

Pearson's correlation coefficient r is the most common way to measure the linear relationship between two variables. Its value ranges from -1 to 1, and it indicates the strength and direction of the relationship between the two variables. The table below explains how to interpret the strength and direction of the relationship based on the correlation coefficient value.

r	Correlation type	Interpretation
$r = 1$	Exactly positive	When one variable increases, the other variable tends to increase too, and when one decreases, the other tends to decrease. They're
$0.5 \le r < 1$	Strong positive	
$0 < r < 0.5$	Weak positive	
$r = 0$	No correlation	No relationship
$-0.5 < r < 0$	Weak negative	When one variable goes up, the other tends to go down, and vice versa.
$-1 < r \le -0.5$	Strong negative	
$r = -1$	Exactly negative	

The Pearson correlation coefficient shows you the direction of the slope: if it's downward, r is negative, and if it's upward, r is positive.



Here's the formula to calculate the correlation coefficient between two variables x and y, each with n values:

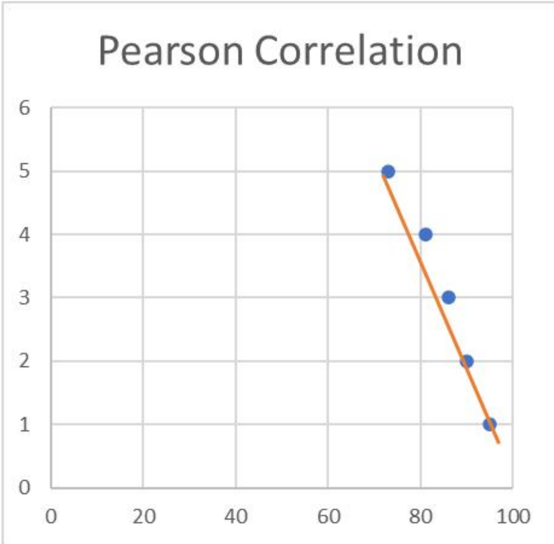
$$r = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2 \sum (y_i - \overline{y})^2}}$$

Example 1: Calculate the Pearson correlation coefficient (r) for the following data on oxygen levels (x) and severity of illness (y) for 5 patients:

x_i	95	90	86	81	73
y_i	1	2	3	4	5

Solution:

$$\overline{x} = \frac{\sum x_i}{n} = \frac{425}{5} = 85 \text{ and } \quad \overline{y} = \frac{\sum y_i}{n} = \frac{15}{5} = 3$$



						Sum
x_i	95	90	86	81	73	425
y_i	1	2	3	4	5	15
$(x_i - \overline{x})$	10	5	1	-4	-12	
$(x_i - \overline{x})^2$	100	25	1	16	144	286
$(y_i - \overline{y})$	-2	-1	0	1	2	
$(y_i - \overline{y})^2$	4	1	0	1	4	10
$(x_i - \overline{x})(y_i - \overline{y})$	-20	-5	0	-4	-24	-53

$$r = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2 \sum (y_i - \overline{y})^2}} = \frac{-53}{\sqrt{286 \times 10}} = \frac{-53}{53.48} = -0.99$$

Strong negative linear relationship (close to -1).

As one variable increases, the other decreases significantly.

In other words, in this example:

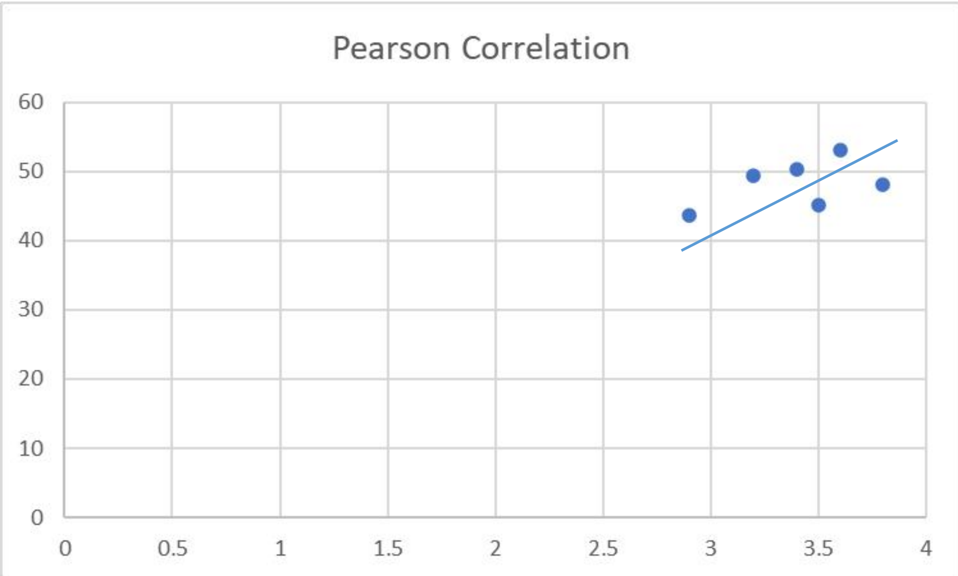
As oxygen levels decrease, illness severity increases significantly.

Example 2: The following data represent the weights and heights of 6 babies born last month in a maternity hospital. Calculate the Pearson correlation coefficient.

Weight (kg)	2.9	3.2	3.4	3.5	3.6	3.8
Length (cm)	43.7	49.3	50.4	45.2	53.1	48.1

Solution:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{20.4}{6} = 3.4 \quad \text{and} \quad \bar{y} = \frac{\sum y_i}{n} = \frac{289.8}{6} = 48.3$$



							Sum
Weight x_i	2.9	3.2	3.4	3.5	3.6	3.8	20.4
Length y_i	43.7	49.3	50.4	45.2	53.1	48.1	289.8
$(x_i - \bar{x})$	- 0.5	-0.2	0	0.1	0.2	0.4	
$(x_i - \bar{x})^2$	0.25	0.04	0	0.01	0.04	0.16	0.5
$(y_i - \bar{y})$	- 4.6	1	2.1	-3.1	4.8	- 0.2	
$(y_i - \bar{y})^2$	21.16	1	4.41	9.61	23.04	0.04	59.26
$(x_i - \bar{x})(y_i - \bar{y})$	2.3	-0.2	0	-0.31	0.96	-0.08	2.65

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{2.65}{\sqrt{0.5 \times 59.26}} = 0.487$$

Simple Equation Regression Linear (معادلة الانحدار الخطي البسيط)

The linear regression equation is expressed as: y = mx + a where x represents the independent variable (plotted on the X-axis), and y represents the dependent variable (plotted on the Y-axis). The slope of the regression line is denoted by m, and the intercept of the regression line, which is the value of y when x equals 0, is represented by a. The values of m and a are calculated using the formulas:

$$m = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$a = \bar{y} - m\bar{x}$$

Example 3: A study revealed the effect of radiation dose on inhibiting the growth of cancer cells as follows:

Radiation Dose (Gy)	2	4	6	8	10
Growth Inhibiting(%)	20	40	55	70	85

1. Find the linear regression equation.

2. If the radiation dose is 7 (Gy), what effect do you expect it to have on cancer cell growth?

Solution:

						Sum
Radiation Dose (Gy) x	2	4	6	8	10	30
Growth Inhibiting(%) y	20	40	55	70	85	270
xy	40	160	330	560	850	1940
x^2	4	16	36	64	100	220

$$m = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{5 \times 1940 - 30 \times 270}{5 \times 220 - (30)^2}$$

$$m = \frac{1600}{200} = 8$$

$$a = \bar{y} - m\bar{x} = \frac{270}{5} - 8 \times \frac{30}{5} = 54 - 48 = 6$$

1. $y = mx + a = 8x + 6$

2. $y = 8 \times 7 + 6 = 62\%$

Example 4: In one study, absorbed radiation was measured according to tissue thickness, and the results were as follows:

Tissue Thickness (cm)	1	2	3	4	5
Absorbed Radiation (%)	80	62	51	41	30

1. Determine the linear regression line.
2. If the tissue thickness is 3.5cm, what percentage of radiation do you expect to be absorbed?

Solution:

						Sum
Tissue Thickness x	1	2	3	4	5	15
Absorbed Radiation y	80	62	51	41	30	264
xy	80	124	153	164	150	671
x^2	1	4	9	16	25	55

$$m = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{5 \times 671 - 15 \times 264}{5 \times 55 - (15)^2} = -12.1$$

$$a = \bar{y} - m\bar{x} = \frac{264}{5} - (-12.1) \times \frac{15}{5} = 89.1$$

1. $y = -12.1\,x + 89.1$
2. At 3.5 cm: Absorbed Radiation $y = -12.1 \times 3.5 + 89.1 = 46.75\%$

H.W.

1. Calculate the Pearson correlation coefficient, in examples 3 and 4.
2. Determine the linear regression line, in examples 1 and 2.