Pearson's Correlation

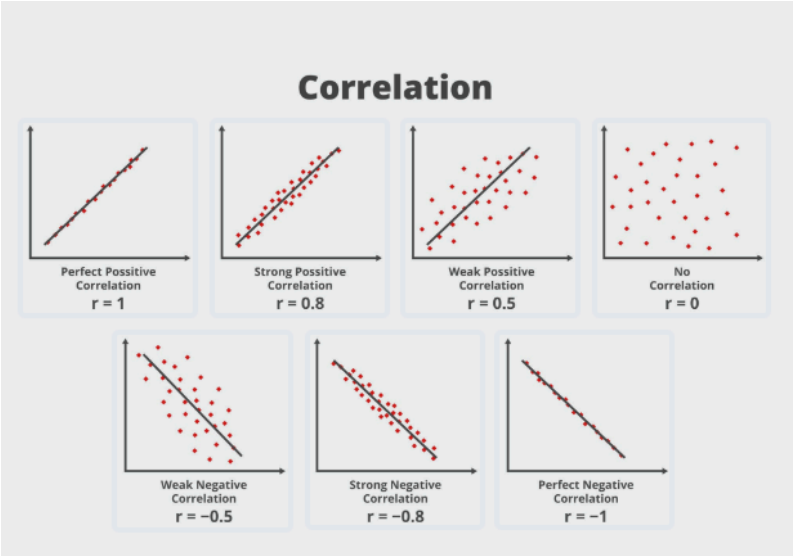
The Pearson Correlation Coefficient (*r*) is a statistical measure that tells you two things about the relationship between two sets of continuous data:

1. **Strength:** How close the data points are to forming a straight line.
2. **Direction:** Whether the variables move in the same direction (positive) or opposite directions (negative).

Basically, it talks about if one thing changes, does the other thing also change and how much?

The value of r is always between **-1** and **+1**.

Value	Meaning	Interpretation
+1	Perfect positive	If A increases, B always increases
Close to +1 (e.g., 0.8)	Strong Positive Correlation	The variables generally move together
0	No relation	A changes, B doesn't care
Close to -1 (e.g., -0.8)	Strong Negative Correlation	As one goes up, the other goes down strongly
-1	Perfect negative	If A increases, B always decreases



How to find r (Pearson correlation coefficient)?

Deviation from mean formula

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \cdot \sum(Y - \bar{Y})^2}}$$

Product Moment Formula

$r_{xy} = \frac{S_{xy}}{S_x S_y}$

r_{xy} = sample correlation coefficient

S_{xy} = sample covariance

S_x = sample standard deviation of x

S_y = sample standard deviation of y

Let’s take an easy example

X	Y
1	2
2	4
3	6

First find out the mean of X and Y

$\overline{X}$ = 2 $\overline{Y}$ = 4

X	Y	X - $\overline{X}$	$(X - \overline{X})^2$	Y - $\overline{Y}$	$(Y - \overline{Y})^2$	(X - $\overline{X}$) (Y - $\overline{Y}$)
1	2	1-2 = -1	1	2-4=-2	4	2
2	4	2-2=0	0	4-4=0	0	0
3	6	3-2=1	1	6-4=2	4	2

Using the formula,

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \cdot \sum(Y - \bar{Y})^2}}$$

$\sum$ (X - $\overline{X}$) (Y - $\overline{Y}$) = 2 + 0 + 2 = 4

$\sum(X - \overline{X})^2$ = 1 + 0 + 1 = 2

$\sum (Y - \overline{Y})^2$ = 4 + 0 + 4 = 8

$= \frac{4}{\sqrt{2 \times 8}} = 1$

r = +1

Using Product Moment Formula

S_{xy} = sample covariance = $\frac{\sum(X - \bar{X})\ (Y - \bar{Y})}{n-1} = \frac{4}{3-1} = 2$

sample standard deviations,

$S_x = \sqrt{\frac{\sum(X - \bar{X})^2}{n-1}} = \sqrt{\frac{2}{2}} = 1.$

$S_y = \sqrt{\frac{\sum(Y - \bar{Y})^2}{n-1}} = \sqrt{\frac{8}{2}} = \sqrt{4} = 2.$

$$r = \frac{s_{xy}}{s_x s_y} = \frac{2}{(1)(2)} = 1$$

Example of Positive Correlation

Hours of studying	Marks
1	50
2	60
3	70

As the number of hours invested in studying increases, marks also increase.

r = +1 (perfect positive correlation)

Example of Positive Correlation

Exercise	Body fat
1	30
2	25
4	20

As you exercise more, your body fat goes down.

$r \approx -0.98$

QUIZ

If the value of Pearson's correlation coefficient rrr is **–0.95**, what does it indicate?

A. Weak positive relationship
B. Strong positive relationship
C. Weak negative relationship
D. Strong negative relationship

ANS : D

Which of the following is a **real-life example** of negative correlation?

A. Height and weight
B. Income and spending
C. Temperature and heater usage
D. Time studied and marks scored

ANS : C

Which of the following is a possible value of Pearson's correlation coefficient rrr?

A. 1.2
B. –0.8
C. 2.5
D. 3

ANS : B

Identify correlation direction

| X | 2 | 4 | 6 |
| Y | 10 | 8 | 6 |

The correlation is:

A. Positive
B. Negative
C. Zero

ANS : B

True/False

Pearson’s correlation works best for **linear** relationships.

ANS : True