PEARSON'S r: CORRELATION AND CAUSATION

Class XI – Artificial Intelligence

CBSE | Simple Notes with a Solved Example

Reference: *Decoding Artificial Intelligence* by Dr. Sachin Gupta, along with CBSE Class XI AI learning material.

---

1. What is Correlation?

Correlation means finding out whether two things are related and how they change together.

Example: We want to know whether students who study for more hours generally score higher marks.

- Variable X = Hours spent studying
- Variable Y = Marks obtained

If students who study more hours generally get higher marks, the two variables have a positive correlation.

In simple words: Correlation tells us how two things are related.

2. What is Pearson's r?

Pearson's r is a mathematical method used to measure the **direction and strength of the linear relationship between two numerical variables**.

The answer is represented by the letter **r**.

The value of Pearson's r always lies between **−1 and +1**.

Understanding the values of r

Value of r	Meaning	Simple example
+1	Perfect positive correlation	As X increases, Y increases in a perfectly straight-line pattern.
Close to +1	Strong positive correlation	More study hours are generally associated with higher marks.
Close to 0	Weak or no linear correlation	There is little or no straight-line relationship.
Close to −1	Strong negative correlation	As one variable increases, the other generally decreases.

Value of r	Meaning	Simple example
−1	Perfect negative correlation	The variables follow a perfectly straight downward line.

Remember:

- Positive (+): Both variables generally move in the same direction.
- Negative (−): The variables generally move in opposite directions.
- Zero (0): No linear relationship is indicated. A curved relationship may still exist.

The value of r is **+0.90**.

- The positive sign means the relationship is positive.
- The value 0.90, being close to +1, indicates a strong positive linear relationship in this small dataset.

Final conclusion: In these observations, students who studied for more hours generally obtained higher marks.

Important: This result shows a relationship. It does not prove that studying longer hours alone caused the higher marks.

---

5. What is Positive Correlation?

Positive correlation occurs when two variables generally move in the same direction.

Example: Study hours and test marks.

As study hours increase, marks tend to increase.

Other possible examples:

- Advertising expenditure and sales.
- Height and weight, in many populations.

A positive correlation has an r value greater than zero.

6. What is Negative Correlation?

Negative correlation occurs when two variables generally move in opposite directions.

Example: As the price of a product increases, the quantity people buy may decrease.

Other example:

- Outside temperature and the amount of heating required in a building.

A negative correlation has an r value less than zero.

7. What is Zero Correlation?

Zero correlation means that there is no apparent linear relationship between the two variables.

Example: In a particular dataset, a student's shoe size might have almost no linear relationship with their marks in an examination.

A value of r close to zero indicates little or no linear relationship.

Important: A zero Pearson correlation does not always mean that the variables have no relationship at all. They might have a curved or other non-linear relationship.

---

8. What is Causation?

Causation means that one thing directly contributes to or produces a change in another thing.

Example: If a student practises solving mathematics questions, the practice can help improve the student's performance.

Here:

- Cause: Practising questions.
- Possible effect: Improved performance.

Causation is about whether one factor actually helps bring about a change in another factor.

In real life, other factors may also affect the result, such as the student's previous knowledge, sleep and quality of practice.

---

9. Difference Between Correlation and Causation

Correlation	Causation
Shows that two variables are related.	Means one factor helps cause a change in another.

Correlation	Causation
Describes how two variables change together.	Explains a cause-and-effect relationship.
Does not prove that one variable causes the other.	Requires evidence that supports a causal relationship.
Can be measured using Pearson's r when appropriate.	Cannot be established by Pearson's r alone.
Example: Ice-cream sales and swimming activity may both increase in summer.	Example: Heating water causes its temperature to rise when sufficient heat is transferred.

example: Ice cream and swimming

Suppose we observe that ice-cream sales increase when more people go swimming.

Does eating ice cream cause people to go swimming?

No, not necessarily.

A third factor, **hot weather**, can help explain both observations:

- Hot weather can increase ice-cream sales.
- Hot weather can encourage more people to swim.

Therefore, ice-cream sales and swimming activity may be correlated, but this does not mean one causes the other.

This is an example of why correlation does not automatically establish causation.

The most important rule

Correlation does not necessarily mean causation.

Even a very high Pearson correlation coefficient does not, by itself, prove cause and effect.

To establish causation, researchers need additional evidence and an appropriate study design. They must consider other possible explanations and factors that may influence the result.

---

10. Why Is Pearson's r Useful in Artificial Intelligence?

In Artificial Intelligence and Machine Learning, data often contain many variables.

Pearson's r can help us:

1. Identify relationships between numerical variables.
2. Explore patterns in a dataset.
3. Understand whether variables tend to move together or in opposite directions.
4. Support data analysis before building a machine-learning model.

For example, a student-performance dataset may contain study hours, attendance and marks. Correlation can help us explore how these variables are related.

However, **correlation alone cannot tell us which variable causes a student's marks to change**.

---

11. Quick Revision

- Correlation describes the direction and strength of a relationship between variables.
- Pearson's r measures linear correlation between two numerical variables.
- The value of r ranges from −1 to +1.
- Positive correlation means the variables generally move in the same direction.
- Negative correlation means they generally move in opposite directions.
- A value near zero indicates little or no linear relationship.
- Causation means one factor contributes to a change in another.
- Correlation alone does not prove causation.

One-line exam answer

Pearson's correlation coefficient (r) is a statistical measure that indicates the strength and direction of the linear relationship between two numerical variables. Its value ranges from −1 to +1. Correlation does not necessarily imply causation.

---

12. Practice Question

A teacher records the following data:

Student	Study hours (X)	Marks (Y)
A	1	2

Student	Study hours (X)	Marks (Y)
B	2	4
C	3	3
D	4	5

Calculate Pearson's r using the computational formula.

Answer for checking: r ≈ 0.80, indicating a strong positive linear relationship in this dataset.

Reference

- *Decoding Artificial Intelligence*, by Dr. Sachin Gupta, Sultan Chand & Sons.
- CBSE Class XI Artificial Intelligence learning material, Pearson's r section: https://cbseacademic.nic.in/web_material/Curriculum21/publication/srsec/AI_Curriculum_HandbookClassXI_Level2.pdf