The Definition and Calculation of The Correlation Coefficient

Data Science and A.I. Lecture Series

Bindeshwar Singh Kushwaha

PostNetwork Academy

Definition of Correlation Coefficient

- Measures the strength and direction of a linear relationship between two variables.

Definition of Correlation Coefficient

- Measures the strength and direction of a linear relationship between two variables.
- Denoted by r, it ranges from -1 to $+1$:

Definition of Correlation Coefficient

- Measures the strength and direction of a linear relationship between two variables.
- Denoted by r, it ranges from -1 to $+1$:
  - $r = +1$: Perfect positive correlation.

Definition of Correlation Coefficient

- Measures the strength and direction of a linear relationship between two variables.
- Denoted by r, it ranges from -1 to $+1$:
  - $r = +1$: Perfect positive correlation.
  - $r = -1$: Perfect negative correlation.

Definition of Correlation Coefficient

- Measures the strength and direction of a linear relationship between two variables.
- Denoted by r, it ranges from -1 to $+1$:
  - $r = +1$: Perfect positive correlation.
  - $r = -1$: Perfect negative correlation.
  - $r = 0$: No linear correlation.

Definition of Correlation Coefficient

- Measures the strength and direction of a linear relationship between two variables.
- Denoted by r, it ranges from -1 to $+1$:
  - $r = +1$: Perfect positive correlation.
  - $r = -1$: Perfect negative correlation.
  - $r = 0$: No linear correlation.
- Formula:

$$r = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}$$

Assumptions for Correlation Coefficient

1. **Linearity:** The relationship between the variables must be linear.

Assumptions for Correlation Coefficient

1. **Linearity:** The relationship between the variables must be linear.
2. **Normality:** Both variables should follow a normal distribution.

Assumptions for Correlation Coefficient

1. **Linearity:** The relationship between the variables must be linear.
2. **Normality:** Both variables should follow a normal distribution.
3. **Cause-and-Effect Relationship:**Correlation does not imply causation; it only measures association.

Examples of Causation

Causation Examples:

- Flipping a Switch and Turning on a Light:

Examples of Causation

Causation Examples:

- Flipping a Switch and Turning on a Light:
  - Flipping the switch (cause) results in the light turning on (effect).

Examples of Causation

Causation Examples:

- Flipping a Switch and Turning on a Light:
  - Flipping the switch (cause) results in the light turning on (effect).
- Smoking and Lung Cancer:

Examples of Causation

Causation Examples:

- Flipping a Switch and Turning on a Light:
  - Flipping the switch (cause) results in the light turning on (effect).
- Smoking and Lung Cancer:
  - Smoking (cause) significantly increases the risk of lung cancer (effect).

Examples of Causation

Causation Examples:

- Flipping a Switch and Turning on a Light:
  - Flipping the switch (cause) results in the light turning on (effect).
- Smoking and Lung Cancer:
  - Smoking (cause) significantly increases the risk of lung cancer (effect).
- Studying and Grades:

Examples of Causation

Causation Examples:

- Flipping a Switch and Turning on a Light:
  - Flipping the switch (cause) results in the light turning on (effect).
- Smoking and Lung Cancer:
  - Smoking (cause) significantly increases the risk of lung cancer (effect).
- Studying and Grades:
  - More study time (cause) leads to better exam grades (effect).

Examples of Causation

Causation Examples:

- Flipping a Switch and Turning on a Light:
  - Flipping the switch (cause) results in the light turning on (effect).
- Smoking and Lung Cancer:
  - Smoking (cause) significantly increases the risk of lung cancer (effect).
- Studying and Grades:
  - More study time (cause) leads to better exam grades (effect).
- Exercise and Fitness:

Examples of Causation

Causation Examples:

- Flipping a Switch and Turning on a Light:
  - Flipping the switch (cause) results in the light turning on (effect).
- Smoking and Lung Cancer:
  - Smoking (cause) significantly increases the risk of lung cancer (effect).
- Studying and Grades:
  - More study time (cause) leads to better exam grades (effect).
- Exercise and Fitness:
  - Regular exercise (cause) improves physical fitness (effect).

Example 1: Positive Correlation

Data:

Example 1: Positive Correlation

Data:

$$X = [10, 20, 30, 40, 50], \quad Y = [15, 30, 45, 60, 75]$$

Steps:

1. Compute means:

$$\bar{X} = \frac{\sum X}{n} = 30, \quad \bar{Y} = \frac{\sum Y}{n} = 45.$$

Example 1: Positive Correlation

Data:

$$X = [10, 20, 30, 40, 50], \quad Y = [15, 30, 45, 60, 75]$$

Steps:

1. Compute means:

$$\bar{X} = \frac{\sum X}{n} = 30, \quad \bar{Y} = \frac{\sum Y}{n} = 45.$$

2. Find deviations:

$$d_x = X - \bar{X}, \quad d_y = Y - \bar{Y}.$$

Example 1: Positive Correlation

Data:

$$X = [10, 20, 30, 40, 50], \quad Y = [15, 30, 45, 60, 75]$$

Steps:

1. Compute means:

$$\bar{X} = \frac{\sum X}{n} = 30, \quad \bar{Y} = \frac{\sum Y}{n} = 45.$$

2. Find deviations:

$$d_x = X - \bar{X}, \quad d_y = Y - \bar{Y}.$$

3. Compute r:

$$r = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \cdot \sum d_y^2}}.$$

Example 1: Computation Table

Step-by-Step Table:

X	Y	d_x	d_y	d_x^2	d_y^2	$d_x d_y$
10	15	-20	-30	400	900	600
20	30	-10	-15	100	225	150
30	45	0	0	0	0	0
40	60	10	15	100	225	150
50	75	20	30	400	900	600
$\sum$	$\sum$	0	0	1000	2250	1500

Final Calculation:

Example 1: Computation Table

Step-by-Step Table:

X	Y	d_x	d_y	d_x^2	d_y^2	$d_x d_y$
10	15	-20	-30	400	900	600
20	30	-10	-15	100	225	150
30	45	0	0	0	0	0
40	60	10	15	100	225	150
50	75	20	30	400	900	600
$\sum$	$\sum$	0	0	1000	2250	1500

Final Calculation:

$$r = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \cdot \sum d_y^2}} = \frac{1500}{\sqrt{1000 \cdot 2250}} = \frac{1500}{1500} = 1$$

Result: $r = 1$ indicates a perfect positive correlation.

Conclusion

- The correlation coefficient r measures the linear relationship between two variables.

Conclusion

- The correlation coefficient r measures the linear relationship between two variables.
- Causation is stronger, requiring evidence that one variable directly affects the other.

Conclusion

- The correlation coefficient r measures the linear relationship between two variables.
- Causation is stronger, requiring evidence that one variable directly affects the other.
- Examples:

Conclusion

- The correlation coefficient r measures the linear relationship between two variables.
- Causation is stronger, requiring evidence that one variable directly affects the other.
- Examples:
  - $r = 1$: Perfect positive correlation.

Conclusion

- The correlation coefficient r measures the linear relationship between two variables.
- Causation is stronger, requiring evidence that one variable directly affects the other.
- Examples:
  - $r = 1$: Perfect positive correlation.
  - $r = -1$: Perfect negative correlation.

Reach PostNetwork Academy

Website

PostNetwork Academy | www.postnetwork.co

YouTube Channel

www.youtube.com/@postnetworkacademy

Facebook Page

www.facebook.com/postnetworkacademy

LinkedIn Page

www.linkedin.com/company/postnetworkacademy