Understanding Correlation: Simplified Explanation!

Data Science and A.I. Lecture Series

Bindeshwar Singh Kushwaha

PostNetwork Academy

Covariance and Standard Deviation

Definitions:

- Sample Covariance:

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

- Sample Standard Deviations:

$$\sigma_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}, \quad \sigma_Y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

Cauchy-Schwarz Inequality

The Cauchy-Schwarz inequality states:

$$\left(\sum_{i=1}^{n} A_i B_i\right)^2 \leq \left(\sum_{i=1}^{n} A_i^2\right)\left(\sum_{i=1}^{n} B_i^2\right),$$

where $A_i = X_i - \bar{X}$ and $B_i = Y_i - \bar{Y}$.

Cauchy-Schwarz Inequality

The Cauchy-Schwarz inequality states:

$$\left(\sum_{i=1}^{n} A_i B_i\right)^2 \leq \left(\sum_{i=1}^{n} A_i^2\right)\left(\sum_{i=1}^{n} B_i^2\right),$$

where $A_i = X_i - \bar{X}$ and $B_i = Y_i - \bar{Y}$.
This leads to:

$$(\mathrm{Cov}(X, Y))^2 \leq \sigma_X^2 \sigma_Y^2$$

$$|\mathrm{Cov}(X, Y)| \leq \sigma_X \sigma_Y.$$

Standard Deviation

Given data

$$X = \{1, 2, 3, 4, 5\}, \quad Y = \{5, 4, 3, 2, 1\}$$

Step 1: Compute Standard Deviations

$$\bar{X} = 3, \quad \bar{Y} = 3$$

$$\sigma_X = \sqrt{\frac{1}{4}\sum_{i=1}^{5}(X_i - 3)^2}, \quad \sigma_Y = \sqrt{\frac{1}{4}\sum_{i=1}^{5}(Y_i - 3)^2}$$

$$\sigma_X = \sigma_Y = \sqrt{\frac{1}{4}[(-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2]}$$

$$\sigma_X = \sigma_Y = \sqrt{\frac{1}{4}(4 + 1 + 0 + 1 + 4)} = \sqrt{\frac{10}{4}} = \sqrt{2.5} \approx 1.58$$

$$\sigma_X x \sigma_Y = 1.58x1.58 = 2.50$$

Given data

$$X = \{1, 2, 3, 4, 5\}, \quad Y = \{5, 4, 3, 2, 1\}$$

Compute Covariance

$$\mathrm{Cov}(X, Y) = \frac{1}{4}\sum_{i=1}^{5}(X_i - 3)(Y_i - 3)$$

$$= \frac{1}{4}\left[(-2)(2) + (-1)(1) + (0)(0) + (1)(-1) + (2)(-2)\right]$$

$$= \frac{1}{4}\left[-4 - 1 + 0 - 1 - 4\right] = \frac{-10}{4} = -2.50$$

Results:

$$|\mathrm{Cov}(X, Y)| = 2.50, \quad \sigma_X x \sigma_Y = 1.58x1.58 = 2.50$$

$$|\mathrm{Cov}(X, Y)| \leq \sigma_X \sigma_Y.$$

Covariance Calculation

Given Data:

$$X = \{1, 2, 3, 4, 5\}, \quad Y = \{5, 2, 3, 4, 1\}$$

Step 1: Compute the Means

$$\bar{X} = \frac{1+2+3+4+5}{5} = 3, \quad \bar{Y} = \frac{5+2+3+4+1}{5} = 3$$

Step 2: Compute Covariance

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

Substitute $\bar{X} = 3$ and $\bar{Y} = 3$:

$$\mathrm{Cov}(X, Y) = \frac{1}{5-1} [(1-3)(5-3) + (2-3)(2-3) + (3-3)(3-3) + (4-3)(4-3) + (5-3)(1-3)]$$

Covariance Calculation

Simplify each term:

$$= \frac{1}{4}[(-2)(2) + (-1)(-1) + (0)(0) + (1)(1) + (2)(-2)]$$

$$= \frac{1}{4}[-4 + 1 + 0 + 1 - 4]$$

$$= \frac{1}{4}(-6) = -1.50$$

Result:

$$\mathrm{Cov}(X, Y) = -1.5 \quad \sigma_X x \sigma_Y = 1.58x1.58 = 2.50$$

$$|\mathrm{Cov}(X, Y)| \leq \sigma_X \sigma_Y.$$

Step-by-Step Sample Covariance Calculation

- Given:

$$x = \{1, 2, 3, 4, 5\}, \quad y = \{1, 2, 3, 4, 5\}$$

- Sample size: $n = 5$
- Mean of x and y:

$$\bar{x} = \frac{1+2+3+4+5}{5} = 3, \quad \bar{y} = \frac{1+2+3+4+5}{5} = 3$$

- Sample covariance:

$$\mathrm{Cov}(x, y) = \frac{1}{5-1} \sum_{i=1}^{5} (x_i - 3)(y_i - 3)$$

- Substituting values:

$$\mathrm{Cov}(x, y) = \frac{1}{4} [(1-3)(1-3) + (2-3)(2-3) + (3-3)(3-3) + (4-3)(4-3) + (5-3)(5-3)]$$

$$\mathrm{Cov}(x, y) = \frac{1}{4} [4 + 1 + 0 + 1 + 4] = \frac{10}{4} = 2.5$$

$$|\mathrm{Cov}(X, Y)| \leq \sigma_X \sigma_Y.$$

Step-by-Step Sample Covariance Calculation

- Given:

$$x = \{1, 2, 3, 4, 5\}, \quad y = \{1, 2, 3, 4, 5\}$$

Step-by-Step Sample Covariance Calculation

- Given:

$$x = \{1, 2, 3, 4, 5\}, \quad y = \{1, 2, 3, 4, 5\}$$

- Sample size: $n = 5$

Step-by-Step Sample Covariance Calculation

- Given:

$$x = \{1, 2, 3, 4, 5\}, \quad y = \{1, 2, 3, 4, 5\}$$

- Sample size: $n = 5$
- Mean of x and y:

$$\bar{x} = \frac{1+2+3+4+5}{5} = 3, \quad \bar{y} = \frac{1+2+3+4+5}{5} = 3$$

Step-by-Step Sample Covariance Calculation

- Given:

$$x = \{1, 2, 3, 4, 5\}, \quad y = \{1, 2, 3, 4, 5\}$$

- Sample size: $n = 5$
- Mean of x and y:

$$\bar{x} = \frac{1+2+3+4+5}{5} = 3, \quad \bar{y} = \frac{1+2+3+4+5}{5} = 3$$

- Sample covariance:

$$\mathrm{Cov}(x, y) = \frac{1}{5-1}\sum_{i=1}^{5}(x_i - 3)(y_i - 3)$$

Step-by-Step Sample Covariance Calculation

- Given:

$$x = \{1, 2, 3, 4, 5\}, \quad y = \{1, 2, 3, 4, 5\}$$

- Sample size: $n = 5$
- Mean of x and y:

$$\bar{x} = \frac{1+2+3+4+5}{5} = 3, \quad \bar{y} = \frac{1+2+3+4+5}{5} = 3$$

- Sample covariance:

$$\text{Cov}(x, y) = \frac{1}{5-1} \sum_{i=1}^{5} (x_i - 3)(y_i - 3)$$

- Substituting values:

$$\text{Cov}(x, y) = \frac{1}{4} [(1-3)(1-3) + (2-3)(2-3) + (3-3)(3-3) + (4-3)(4-3) + (5-3)(5-3)]$$

Step-by-Step Sample Covariance Calculation

- Given:

$$x = \{1, 2, 3, 4, 5\}, \quad y = \{1, 2, 3, 4, 5\}$$

- Sample size: $n = 5$
- Mean of x and y:

$$\bar{x} = \frac{1+2+3+4+5}{5} = 3, \quad \bar{y} = \frac{1+2+3+4+5}{5} = 3$$

- Sample covariance:

$$\mathrm{Cov}(x, y) = \frac{1}{5-1} \sum_{i=1}^{5} (x_i - 3)(y_i - 3)$$

- Substituting values:

$$\mathrm{Cov}(x, y) = \frac{1}{4} [(1-3)(1-3) + (2-3)(2-3) + (3-3)(3-3) + (4-3)(4-3) + (5-3)(5-3)]$$

- $$\mathrm{Cov}(x, y) = \frac{1}{4} [4 + 1 + 0 + 1 + 4] = \frac{10}{4} = 2.5$$

Step-by-Step Sample Covariance Calculation

- Given:

$$x = \{1, 2, 3, 4, 5\}, \quad y = \{1, 2, 3, 4, 5\}$$

- Sample size: $n = 5$
- Mean of x and y:

$$\bar{x} = \frac{1+2+3+4+5}{5} = 3, \quad \bar{y} = \frac{1+2+3+4+5}{5} = 3$$

- Sample covariance:

$$\mathrm{Cov}(x, y) = \frac{1}{5-1} \sum_{i=1}^{5} (x_i - 3)(y_i - 3)$$

- Substituting values:

$$\mathrm{Cov}(x, y) = \frac{1}{4} [(1-3)(1-3) + (2-3)(2-3) + (3-3)(3-3) + (4-3)(4-3) + (5-3)(5-3)]$$

- $$\mathrm{Cov}(x, y) = \frac{1}{4} [4 + 1 + 0 + 1 + 4] = \frac{10}{4} = 2.5$$

- $$|\mathrm{Cov}(X, Y)| \le \sigma_X \sigma_Y.$$