Bivariate Distribution Made Simple: From Definition to Covariance Calculation

Data Science and A.I. Lecture Series

Bindeshwar Singh Kushwaha

PostNetwork Academy

Introduction

- Statistics involves analyzing relationships between variables.

Introduction

- Statistics involves analyzing relationships between variables.
- The relationship between two variables can be studied using bivariate distribution.

Introduction

- Statistics involves analyzing relationships between variables.
- The relationship between two variables can be studied using bivariate distribution.
- Bivariate distribution explores joint behavior and dependence between variables.

Bivariate Distribution

- A bivariate distribution describes the probability distribution of two random variables.

Bivariate Distribution

- A bivariate distribution describes the probability distribution of two random variables.
- It defines how the two variables, say X and Y, behave jointly.

Bivariate Distribution

- A bivariate distribution describes the probability distribution of two random variables.
- It defines how the two variables, say X and Y, behave jointly.
- Example: Height and weight of individuals in a population.

Bivariate Distribution

- A bivariate distribution describes the probability distribution of two random variables.
- It defines how the two variables, say X and Y, behave jointly.
- Example: Height and weight of individuals in a population.
- Visualization techniques include scatter plots and contour plots.

Covariance

Definition

- Covariance measures how two variables vary together.

Covariance

Definition

- Covariance measures how two variables vary together.
- Let (x_i, y_i) be pairs of data points. The covariance formula is:

$$Cov(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Covariance

Definition

- Covariance measures how two variables vary together.
- Let (x_i, y_i) be pairs of data points. The covariance formula is:

$$Cov(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

- Here, $\bar{x}$ and $\bar{y}$ are the mean values of X and Y.

Covariance

Definition

- Covariance measures how two variables vary together.
- Let (x_i, y_i) be pairs of data points. The covariance formula is:

$$Cov(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

- Here, $\bar{x}$ and $\bar{y}$ are the mean values of X and Y.
- Covariance indicates the direction of the relationship between X and Y.

Key Insights on Covariance

- If $Cov(X, Y) > 0$: X and Y tend to increase together (positive relationship).

Key Insights on Covariance

- If $Cov(X, Y) > 0$: X and Y tend to increase together (positive relationship).
- If $Cov(X, Y) < 0$: X increases as Y decreases (negative relationship).

Key Insights on Covariance

- If $Cov(X, Y) > 0$: X and Y tend to increase together (positive relationship).
- If $Cov(X, Y) < 0$: X increases as Y decreases (negative relationship).
- If $Cov(X, Y) = 0$: X and Y are uncorrelated (no linear relationship).

Key Insights on Covariance

- If $Cov(X, Y) > 0$: X and Y tend to increase together (positive relationship).
- If $Cov(X, Y) < 0$: X increases as Y decreases (negative relationship).
- If $Cov(X, Y) = 0$: X and Y are uncorrelated (no linear relationship).
- Covariance is sensitive to the scale of variables.

Illustration

- Consider the dataset:

$$(x_i, y_i) = \{(1,2),(2,4),(3,6),(4,8),(5,10)\}$$

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$
1	2	-2	-4
2	4	-1	-2
3	6	0	0
4	8	1	2
5	10	2	4

Illustration

- Consider the dataset:

$$(x_i, y_i) = \{(1,2),(2,4),(3,6),(4,8),(5,10)\}$$

- Organize the data in a table:

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$
1	2	-2	-4
2	4	-1	-2
3	6	0	0
4	8	1	2
5	10	2	4

Step-by-Step Calculation (Part 1)

- **Step 1: Compute the Means**

$$\bar{x} = \frac{\sum x_i}{n} = \frac{15}{5} = 3$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{30}{5} = 6$$

Step-by-Step Calculation (Part 1)

- **Step 1: Compute the Means**

$$\bar{x} = \frac{\sum x_i}{n} = \frac{15}{5} = 3$$
$$\bar{y} = \frac{\sum y_i}{n} = \frac{30}{5} = 6$$

- **Step 2: Find Deviations**

$$x_i - \bar{x} : -2, -1, 0, 1, 2$$
$$y_i - \bar{y} : -4, -2, 0, 2, 4$$

Step-by-Step Calculation (Part 2)

- **Step 3: Multiply Deviations and Sum**

$$(x_i - \bar{x})(y_i - \bar{y}) : 8, 2, 0, 2, 8$$

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 20$$

Step-by-Step Calculation (Part 2)

- **Step 3: Multiply Deviations and Sum**

$$(x_i - \bar{x})(y_i - \bar{y}) : 8, 2, 0, 2, 8$$

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = 20$$

- **Step 4: Apply Covariance Formula**

$$Cov(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$
$$= \frac{1}{5} \times 20 = 4$$

Analysis and Conclusion

- The covariance between X and Y is 4, indicating a positive linear relationship.

Analysis and Conclusion

- The covariance between X and Y is 4, indicating a positive linear relationship.
- This implies that as X increases, Y also tends to increase proportionally.

Analysis and Conclusion

- The covariance between X and Y is 4, indicating a positive linear relationship.
- This implies that as X increases, Y also tends to increase proportionally.
- Covariance does not measure the strength of the relationship (correlation coefficient is used for that).

Summary

- Bivariate distribution helps analyze the joint behavior of two variables.

Summary

- Bivariate distribution helps analyze the joint behavior of two variables.
- Covariance measures how two variables change together.

Summary

- Bivariate distribution helps analyze the joint behavior of two variables.
- Covariance measures how two variables change together.
- Positive covariance: Variables move in the same direction.

Summary

- Bivariate distribution helps analyze the joint behavior of two variables.
- Covariance measures how two variables change together.
- Positive covariance: Variables move in the same direction.
- Negative covariance: Variables move in opposite directions.

Summary

- Bivariate distribution helps analyze the joint behavior of two variables.
- Covariance measures how two variables change together.
- Positive covariance: Variables move in the same direction.
- Negative covariance: Variables move in opposite directions.
- Covariance is a building block for correlation analysis and regression.

Thank You!