Statistical and Matrix Analysis of Sales Data

Your Name

May 6, 2025

Problem Description

You are given the following dataset represented as a matrix:

$$D = \begin{bmatrix} 4 & 8 & 12 \\ 6 & 10 & 15 \\ 5 & 7 & 9 \\ 3 & 6 & 9 \end{bmatrix}$$

Each row represents a different region, and each column represents a specific product category (A, B, C). The values in the matrix are sales quantities.

Part 1: Understanding the Data

Data Types

The elements in the matrix D are integers, representing sales quantities. Therefore, the data type of the matrix elements is **integer**.

Descriptive Statistics

We calculate the mean, median, and standard deviation for each column (Product A, B, and C).

Product A (first column: $[4, 6, 5, 3]$**)**

- **Mean:**
$$\text{Mean}_A = \frac{(4+6+5+3)}{4} = \frac{18}{4} = 4.5$$
- **Median:** Sorted values for Product A: $[3, 4, 5, 6]$. The median is the average of the two middle values:
$$\text{Median}_A = \frac{(4+5)}{2} = 4.5$$

- **Standard Deviation:**

$$\text{Standard Deviation}_A = \sqrt{\frac{(0.25 + 2.25 + 0.25 + 2.25)}{4}} = \sqrt{1.25} \approx 1.118$$

Product B (second column: $[8, 10, 7, 6]$)

- **Mean:**

$$\text{Mean}_B = \frac{(8 + 10 + 7 + 6)}{4} = \frac{31}{4} = 7.75$$

- **Median:** Sorted values for Product B: $[6, 7, 8, 10]$. The median is the average of the two middle values:

$$\text{Median}_B = \frac{(7 + 8)}{2} = 7.5$$

- **Standard Deviation:**

$$\text{Standard Deviation}_B = \sqrt{\frac{(0.0625 + 5.0625 + 0.5625 + 3.0625)}{4}} = \sqrt{2.1875} \approx 1.48$$

Product C (third column: $[12, 15, 9, 9]$)

- **Mean:**

$$\text{Mean}_C = \frac{(12 + 15 + 9 + 9)}{4} = \frac{45}{4} = 11.25$$

- **Median:** Sorted values for Product C: $[9, 9, 12, 15]$. The median is the average of the two middle values:

$$\text{Median}_C = \frac{(9 + 12)}{2} = 10.5$$

- **Standard Deviation:**

$$\text{Standard Deviation}_C = \sqrt{\frac{(0.5625 + 14.0625 + 5.0625 + 5.0625)}{4}} = \sqrt{6.1875} \approx 2.49$$

Variability

For each column, we calculate the range and variance.

Product A

- **Range:**

$$\text{Range}_A = 6 - 3 = 3$$

- **Variance:**

$$\text{Variance}_A = 1.25$$

Product B

- **Range:**

$$\text{Range}_B = 10 - 6 = 4$$

- **Variance:**

$$\text{Variance}_B = 2.1875$$

Product C

- **Range:**

$$\text{Range}_C = 15 - 9 = 6$$

- **Variance:**

$$\text{Variance}_C = 6.1875$$

Part 2: Linear Algebra

Matrix Operations

We compute the transpose of the matrix D.

The transpose of matrix D is denoted as D^T:

$$D^T = \begin{bmatrix} 4 & 6 & 5 & 3 \\ 8 & 10 & 7 & 6 \\ 12 & 15 & 9 & 9 \end{bmatrix}$$

Transforming the Data

Now, let's multiply the matrix D by the price vector $P = \begin{bmatrix} 5 & 10 & 15 \end{bmatrix}$.

The matrix multiplication of D by P gives the revenue for each region as:

$$R = D \times P = \begin{bmatrix} 4 & 8 & 12 \\ 6 & 10 & 15 \\ 5 & 7 & 9 \\ 3 & 6 & 9 \end{bmatrix} \times \begin{bmatrix} 5 \\ 10 \\ 15 \end{bmatrix} = \begin{bmatrix} (4 \times 5) + (8 \times 10) + (12 \times 15) \\ (6 \times 5) + (10 \times 10) + (15 \times 15) \\ (5 \times 5) + (7 \times 10) + (9 \times 15) \\ (3 \times 5) + (6 \times 10) + (9 \times 15) \end{bmatrix} = \begin{bmatrix} 190 \\ 290 \\ 240 \\ 225 \end{bmatrix}$$

Thus, the revenue generated by each region is:

$$R = \begin{bmatrix} 190 \\ 290 \\ 240 \\ 225 \end{bmatrix}$$

Identify Patterns

- The region that generates the highest revenue is the second region with a revenue of $290.
- The total revenue for each product across all regions can be computed as:

$$\text{Total Revenue}_A = (4 \times 5) + (6 \times 5) + (5 \times 5) + (3 \times 5) = 70$$

$$\text{Total Revenue}_B = (8 \times 10) + (10 \times 10) + (7 \times 10) + (6 \times 10) = 310$$

$$\text{Total Revenue}_C = (12 \times 15) + (15 \times 15) + (9 \times 15) + (9 \times 15) = 690$$

Part 3: Conceptual Questions

Interpretation

- The **mean** tells us the average sales quantity for each product across all regions.
- The **coefficient of variation** (CV), calculated as the ratio of standard deviation to the mean, helps us understand the consistency of sales. A lower CV indicates more consistency in sales across regions.

Application

Matrix multiplication is used in real-world data science problems to compute total revenue, model customer preferences, or predict future trends. It is particularly useful in scenarios like:

- Calculating revenue across different products and regions, as shown in this example.
- Recommender systems, where matrix multiplication helps predict user-item preferences.