

Scenario 1:

Delivery Time Analysis for an E-commerce Company

An e-commerce company tracks delivery times (in minutes) for 15 orders:

[25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95]

The company wants to analyze the delivery performance using percentiles and detect if there are any unusual delivery times.

✓ Question 1: Calculate Q1 and Q3

- Q1 (25th Percentile) → 25% of the data is below this value
- Q3 (75th Percentile) → 75% of the data is below this value

📌 Formula:

$$Q1 = 0.25 \times (n+1) \quad Q1 = 0.25 \times (n+1) \quad Q3 = 0.75 \times (n+1) \quad Q3 = 0.75 \times (n+1) \quad Q3 = 0.75 \times (n+1)$$

- Q1 = 4th value → 40
- Q3 = 11th value → 75

✓ Question 2: Find the IQR

- Interquartile Range (IQR) measures the spread of the middle 50% of data.

📌 Formula:

$$IQR = Q3 - Q1 = 75 - 40 = 35 \quad IQR = Q3 - Q1 = 75 - 40 = 35 \quad IQR = Q3 - Q1 = 75 - 40 = 35$$

✓ Question 3: Detect Outliers

- Outliers are extreme values.
- Use these formulas to find boundaries:

📌 Formulas:

$$\begin{aligned} \text{Lower Bound} &= Q1 - 1.5 \times IQR \\ \text{Upper Bound} &= Q3 + 1.5 \times IQR \end{aligned}$$

- Lower Bound = $40 - 1.5 \times 35 = -12.5$ → No lower outliers
- Upper Bound = $75 + 1.5 \times 35 = 127.5$ → No upper outliers. ✓ Conclusion: No outliers in this dataset.

Scenario 2:

A teacher is analyzing the mathematics scores of students in her class. The scores are:

[45, 50, 55, 60, 60, 62, 63, 65, 90, 95]

Question:

- Calculate the mean, median, and mode of the scores.
- Explain why the median might be a better representation than the mean in this case.

Answer:

- Mean: $\frac{45+50+55+60+60+62+63+65+90+95}{10} = 65.5$
- Median: Middle value = $\frac{60+62}{2} = 61$
- Mode: 60 (Occurs twice)

Explanation:

- The median is a better measure here because the high outliers (90 and 95) skew the mean. The median is not affected by extreme values, providing a better reflection of student performance.

Scenario 3:

A grocery store manager tracks how many customers visit the store daily for a month:
[5, 10, 8, 15, 20, 5, 12, 14, 10, 18]

Question 1:

- Create a frequency distribution table for this data.

Answer:

Number of Customers	Frequency
5	2
10	2
8	1
12	1
14	1
15	1
18	1
20	1

Scenario 4:

A real estate model has three variables:

- House Size
- Number of Rooms

- Number of Bathrooms

Question 1:

- How can you detect multicollinearity?

Answer:

- Logic:
 - Calculate the Variance Inflation Factor (VIF).
 - $VIF > 10$ indicates multicollinearity.
- Answer: High VIF means the variables are correlated, impacting model accuracy.

Scenario 5:

A company made a new medicine to lower blood pressure. They gave it to one group and gave a fake pill (placebo) to another group.

Question 1:

How can the company check if the new medicine works?

Solution:

- Make a guess (Hypothesis):
 - H_0 : The medicine doesn't lower blood pressure.
 - H_1 : The medicine lowers blood pressure.
- Do a T-Test:
 - Find the p-value (a number that shows how likely the result happened by chance).
 - If $p\text{-value} < 0.05$, it means the medicine likely works.
- Final Answer:
 - If the p-value is small, the medicine is effective.

Scenario 6: Identifying Outliers in Sales Data

A company wants to find any unusual spikes in sales.

Question 1:

How can the company detect outliers in their sales data?

Logic:

- Step 1: Calculate the Interquartile Range (IQR).

- Step 2: Identify outliers using the formula:

$$\text{Outliers} = (\text{Data} < Q1 - 1.5 \times \text{IQR}) \text{ or } (\text{Data} > Q3 + 1.5 \times \text{IQR})$$

Scenario 7: Understanding Customer Satisfaction

A restaurant conducted a survey to rate customer satisfaction on a scale of 1 to 5:
 [5, 4, 4, 5, 3, 4, 5, 2, 4, 3]

 Question 1: How can the restaurant summarize the overall satisfaction?

Answer:

- Find the Mode to see the most common rating.
- Calculate the Mean and Median for further insights.
- If most ratings are 4 or 5, satisfaction is generally high.