

Identifying Trends and Patterns in Large Datasets

Part 1: Office Roleplay Dialogue

Scenario: A Data Analyst, Rina, is discussing data analysis techniques with her colleague, David, to identify trends and patterns in a large dataset.

David: Hey Rina, I've been going through the sales data for the past year. What's the best way to summarize the key figures?

Rina: You can start with **descriptive statistics** to calculate averages, medians, and standard deviations. That will give us an overall picture of the data.

David: That makes sense. I also want to look at long-term changes. Would **trend analysis** help with that?

Rina: Absolutely! **Trend analysis** will help us see whether sales are increasing, decreasing, or staying stable over time.

David: Good to know. I also noticed that higher advertising spending seems to be linked to increased sales. Could there be a **correlation**?

Rina: Possibly! We can measure the strength of the **correlation** between ad spending and sales using a correlation coefficient.

David: Great! I also found a few extreme values that don't match the overall pattern. Could those be **outliers**?

Rina: Yes, **outliers** can sometimes be errors, but they can also indicate unusual but important business events.

David: Got it. And should we check the **distribution** of the data to see if it's skewed?

Rina: Definitely. Checking the **distribution** helps us understand whether most sales are concentrated around a certain value or spread out widely.

David: Thanks, Rina. I'll summarize the findings, and we can discuss the next steps tomorrow.

Rina: Sounds good! Let's see what insights we can uncover.

Part 2: Comprehension Questions

1. What does Rina suggest using to summarize key figures in the dataset?

- (A) Changing the sales data manually
- (B) Removing all missing values
- (C) Descriptive statistics
- (D) Adding more columns to the dataset

2. What does trend analysis help with?

- (A) Identifying patterns in data over time
- (B) Removing outliers from the dataset
- (C) Sorting data alphabetically
- (D) Predicting future stock prices

3. Why is correlation important in data analysis?

- (A) It increases the size of the dataset
- (B) It deletes unnecessary records

- (C) It converts text into numerical values
- (D) It measures the relationship between two variables

4. What do outliers in a dataset indicate?

- (A) They always need to be deleted
 - (B) They could be errors or important unusual events
 - (C) They make trend analysis impossible
 - (D) They are automatically removed in Excel
-

Part 3: Key Vocabulary Definitions in Japanese

1. **Descriptive Statistics (記述統計)** – 平均、中央値、標準偏差などを計算してデータの全体像を把握する統計手法。
 2. **Trend Analysis (トレンド分析)** – データの時間的な変化を分析し、増加、減少、または安定しているかを判断する方法。
 3. **Correlation (相関関係)** – 2 つの変数の関係を測定し、一方の変化がもう一方の変化とどのように関連しているかを示す指標。
 4. **Outliers (外れ値)** – 他のデータポイントと大きく異なる異常値。エラーの場合もあるが、重要な傾向を示すこともある。
 5. **Distribution (分布)** – データポイントの分散状況を示し、平均値の周りにどのようにデータが分布しているかを確認する方法。
-

Part 4: Questions & Correct Answers

1. **What does Rina suggest using to summarize key figures in the dataset?**

(C) Descriptive statistics

2. **What does trend analysis help with?**

(A) Identifying patterns in data over time

3. **Why is correlation important in data analysis?**

(D) It measures the relationship between two variables

4. **What do outliers in a dataset indicate?**

(B) They could be errors or important unusual events