

2.3 Extracting Information from Data

Name: _____ Class: _____ Date: _____

Total: 9 marks

Objective

Build the skills to answer exam questions on **extracting information from data**.

You must be able to:

- explain how programs find **patterns** 模式和 **trends** 趋势 in large data sets
- use **filtering** and **cleaning** 清洗 to prepare data
- distinguish **correlation** 相关 from **causation** 因果
- describe how **metadata** 元数据 and a **visualization** 可视化 help

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ Patterns and cleaning

Programs process large data sets to reveal **patterns and trends**. Raw data is first **filtered** and **cleaned** (removing errors and duplicates).

■ Correlation vs causation

Two things moving together is **correlation**; it does **not** prove one **causes** the other—a hidden factor may drive both.

■ Metadata and visualization

Metadata is data about data (e.g. a photo's date and location). A **visualization** (chart or graph) communicates insight. How data is **collected** can introduce bias.

2 Practice

2.1 State the difference between correlation and causation. [2]

2.2 Define metadata. [1]

2.3 State one reason to clean data before analysing it. [1]

3 Exam-style questions

3.1 Data about data, such as a photo's date and location, is called [1]

- **A** metadata
 - **B** a variable
 - **C** a byte
 - **D** a pixel
-

3.2 Two variables that rise together show [1]

- **A** causation
 - **B** correlation
 - **C** an error
 - **D** nothing at all
-

3.3 A chart shows that ice-cream sales and drowning incidents both rise in summer.

(a) Name the relationship the chart shows. [1]

(b) State whether one directly causes the other. [1]

(c) Name a likely hidden factor behind both. [1]

4 Go further

- work through the **2.3 Extracting Information from Data** lesson on the **Learn** page;
- read the **Data** section of the AP Computer Science Principles handout on the **Know** page.

Solutions

2.1 correlation is two things changing together; causation is one thing actually making the other happen.

2.2 data about data (e.g. the date, size, or location of a file).

2.3 to remove errors, duplicates, or missing values so the analysis is reliable.

3.1 A.

3.2 B.

3.3 (a) correlation. (b) no. (c) hot weather (more people swim and buy ice cream).