

E9.5 Scatter diagrams

Name: _____ Class: _____ Date: _____ **Total: 11 marks**

KEY WORDS correlation 相关性 • line of best fit 最佳拟合线 • extrapolation 外推
• causation 因果关系 • scatter diagram 散点图

Objective

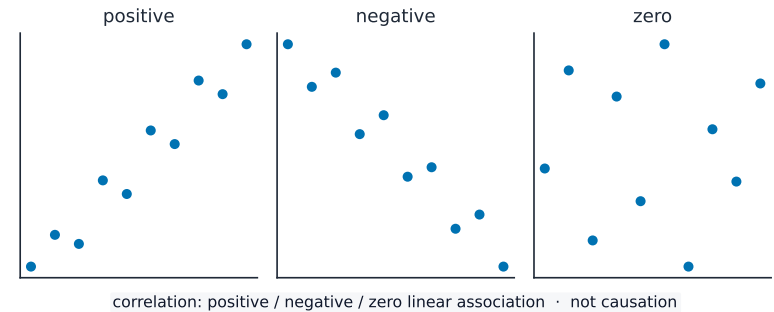
Build the skills to answer Extended exam questions on **scatter diagrams** 散点图 — describing **correlation** 相关性, drawing and using a **line of best fit** 最佳拟合线, and explaining why correlation is not **causation** 因果关系.

You must be able to:

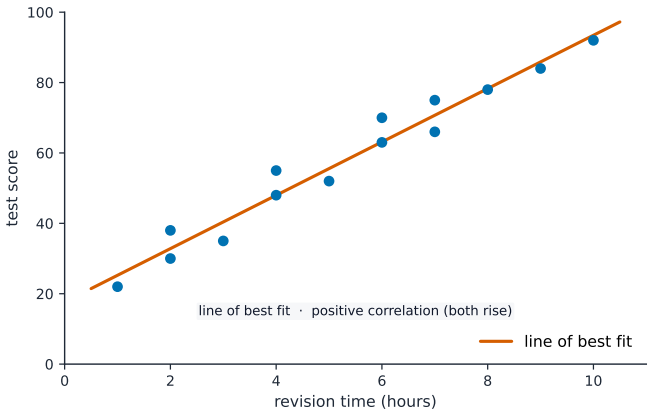
- describe the type and strength of correlation shown by a scatter diagram
- use a line of best fit to estimate a value, and say when that estimate is unreliable
- explain that correlation does not prove one variable causes the other

1 Worked examples

■ Name the type, then the strength



Up to the right is positive; down to the right is negative



The line should follow the trend with roughly equal numbers of points either side

- **Positive correlation:** as one goes up, so does the other. **Negative:** as one goes up, the other goes down. **No correlation:** no pattern.
- **Strength:** “strong” when the points lie close to a line, “weak” when they are scattered. A full answer gives BOTH words — “strong positive”.
- **Line of best fit:** drawn through the middle of the points, with roughly as many above as below. Use it to read an estimate off the graph.
- **Two warnings:** reading far outside the data (**extrapolation**) is unreliable because the pattern may not continue; and correlation does NOT prove cause — a third factor may explain both.

Answer shapes (marked by **method marks** — M1 for a correct method, A1 for the correct answer): these are B marks for the right words, and a “why” question always wants a reason, not a restatement.

2 Practice

2.1 On a scatter diagram the points rise from the bottom left to the top right. Describe the correlation. [1]

2.2 State what a line of best fit is used for. [1]

2.3 State one reason why an estimate read far outside the range of the data may be unreliable. [1]

3 Exam-style questions

3.1 A scatter diagram shows the number of hours of revision and the test score for 12 students. The points lie close to a line rising to the right.

(a) Describe the correlation, giving both its type and its strength. [2]

(b) Explain how a line of best fit could be used to estimate the score of a student who revised for 6 hours. [2]

(c) Explain why this scatter diagram does not prove that revising causes a higher score. [2]

3.2 State, with a reason, the type of correlation you would expect between the age of a car and its value. [2]

Solutions

Marking: **M1** a correct method, **A1** the correct answer.

Each answer shows the steps, then the **result**.

2.1

- as one variable increases so does the other

positive correlation

2.2

- to estimate the value of one variable from a known value of the other

2.3

- the pattern shown by the data may not continue outside the range that was measured
— this is extrapolation

3.1

(a)

- rising to the right is positive
- the points lying CLOSE to the line makes it strong, so: **strong positive correlation**

(b)

- find 6 hours on the horizontal axis and read straight up to the line of best fit
- then read across to the score axis; the estimate is the value there

(c)

- correlation shows only that the two vary together, not that one produces the other
- a third factor could explain both — for example a student who is more motivated may both revise more AND pay more attention in class

3.2

- **negative correlation**
- as a car gets older its value generally falls, so the two move in opposite directions