Exercise walk-through

6.2.2
Scatter Predictions

AQA · GCSE · Mathematics

Samples and data

- **Primary data** is collected by you; **secondary data** was collected by someone else.
- **Categorical data** is in words, such as favourite colour.
- **Discrete** data can take only certain values, such as the number of brothers and sisters.
- **Continuous data** can take any value in a **range**, such as a time or a height, even if it is rounded when it is recorded.

Histograms, cumulative frequency and box plots (H)

- **Histograms.** In a **histogram** the area of each bar, not its height, shows the frequency.
- When the **class intervals** have different widths, the height is the **frequency density**:
- **Worked example 6 (H).** Classes $0 \leq x < 10$, $10 \leq x < 20$ and $20 \leq x < 40$ have frequencies 20, 30 and 20.
- The **class width** of the last class is 20, so the frequency densities are $20 \div 10 = 2$, $30 \div 10 = 3$ and $20 \div 20 = 1$.

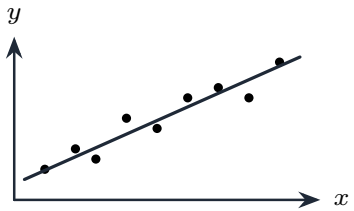
$$\text{frequency density} = \frac{\text{frequency}}{\text{class width}}$$

Sources for this lesson

- AQA, *GCSE Mathematics (8300) specification*, subject content section 3.6 (Statistics).
- "Landon, 1,293,669; Roosevelt, 972,897", *The Literary Digest* (31 October 1936).
- P. Squire, "Why the 1936 Literary Digest poll failed", *Public Opinion Quarterly* 52 (1988), pages 125 to 133.
- AQA, *GCSE Mathematics* Higher Papers 1, 2 and 3 (June 2024), the **pie chart**, frequency table and **box plot** questions, adapted.

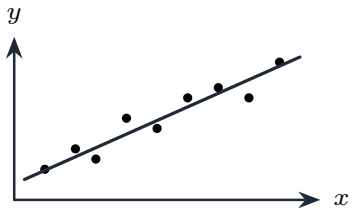
Exercise 2.1

2.1 Use the worked example. Predict the score for 2.5 hours of study. [1]



Exercise 2.1

2.1 Use the worked example. Predict the score for 2.5 hours of study. [1]

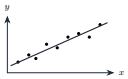


2.1

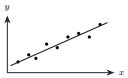
- $40 + 4.5 \times 2.5 = 40 + 11.25 = \mathbf{51.25}$, about 51

Exercise 2.2

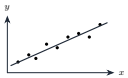
2.2 Describe the type of correlation you would expect between the age of a car and its value. [1]



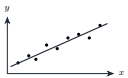
2.3 Describe the type of correlation you would expect between temperature and ice cream sales, and between shoe size and test score. [2]



2.4 A student studied for 2 hours and scored 85. Explain why this point is an outlier on the worked example's graph. [1]

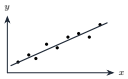


2.5 Use the line to predict the score for 15 hours of study. Explain why this prediction is not reliable. [2]

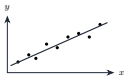


Exercise 2.2

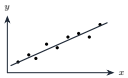
2.2 Describe the type of correlation you would expect between the age of a car and its value. [1]



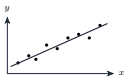
2.3 Describe the type of correlation you would expect between temperature and ice cream sales, and between shoe size and test score. [2]



2.4 A student studied for 2 hours and scored 85. Explain why this point is an outlier on the worked example's graph. [1]



2.5 Use the line to predict the score for 15 hours of study. Explain why this prediction is not reliable. [2]



2.2

- **negative**: older cars tend to be worth less

2.3

- temperature and ice cream sales: **positive**
- shoe size and test score: **no** correlation

Full solution on next slide

Full solution 2.2

2.2

- **negative**: older cars tend to be worth less

2.3

- temperature and ice cream sales: **positive**
- shoe size and test score: **no** correlation

2.4

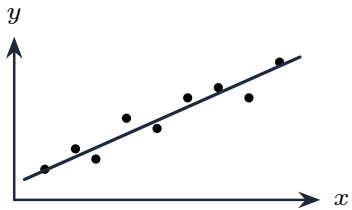
- the line predicts $40 + 4.5 \times 2 = 49$ for 2 hours; 85 is far above the line and the pattern

2.5

- $40 + 4.5 \times 15 = \mathbf{107.5}$
- 15 hours is outside the data (extrapolation), and 107.5 is more than the maximum mark of 100, so it is not reliable

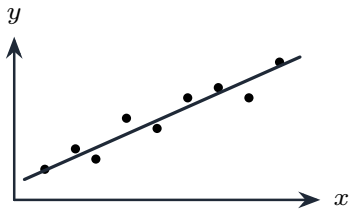
Exercise 3.1

3.1 A town finds that ice cream sales and sunburn cases are positively correlated. Explain why ice cream does not cause sunburn. [1]



Exercise 3.1

3.1 A town finds that ice cream sales and sunburn cases are positively correlated. Explain why ice cream does not cause sunburn. [1]

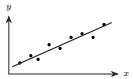


3.1

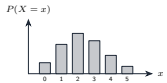
- both increase in hot, sunny weather: a third factor, not one causing the other

Exercise 3.2

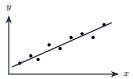
3.2 Sam draws his line of best fit through the first and last points only. Explain why this may not be a good line of best fit. [1]



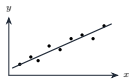
3.3 Explain what the numbers 4.5 and 40 mean in the line score $= 40 + 4.5 \times$ hours. [2]



3.4 Use the line to work out how many hours of study give a predicted score of 58. [2]

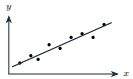


3.5 For another class, a line of best fit passes through $(2, 49)$ and $(10, 85)$, where x is hours and y is score. Find its equation in the form $y = a + bx$, and use it to predict the score for 5 hours. [4]

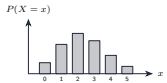


Exercise 3.2

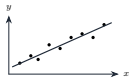
3.2 Sam draws his line of best fit through the first and last points only. Explain why this may not be a good line of best fit. [1]



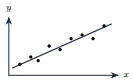
3.3 Explain what the numbers 4.5 and 40 mean in the line score $= 40 + 4.5 \times$ hours. [2]



3.4 Use the line to work out how many hours of study give a predicted score of 58. [2]



3.5 For another class, a line of best fit passes through $(2, 49)$ and $(10, 85)$, where x is hours and y is score. Find its equation in the form $y = a + bx$, and use it to predict the score for 5 hours. [4]



3.2

- a line of best fit should follow all the points, with about as many above as below; two points may be unusual

3.3

- 4.5: each extra hour of study adds about **4.5 marks**

Full solution on next slide

Full solution 3.2

3.2

- a line of best fit should follow all the points, with about as many above as below; two points may be unusual

3.3

- 4.5: each extra hour of study adds about **4.5 marks**
- 40: the predicted score for **0 hours** of study

3.4

- $40 + 4.5h = 58$, so $4.5h = 18$
- $h = \mathbf{4}$ hours

3.5

- gradient $b = \dfrac{85 - 49}{10 - 2} = \dfrac{36}{8} = 4.5$
- $49 = a + 4.5 \times 2$, so $a = 40$
- $\boldsymbol{y = 40 + 4.5x}$
- for 5 hours: $y = 40 + 4.5 \times 5 = \mathbf{62.5}$