



H02-30min Correlation and Causation (Higher): 30-minute Pack

EDEXCEL 1ST0 · Calculator allowed · about 30 minutes

Total Marks

Name: _____ Date: ____ / ____ / ____

Answer ALL questions. Show all your working.

- 1** For each pair of variables, write down whether you would expect positive correlation, negative correlation, or no correlation.
- (a) The number of pages in a paperback book and the price of the book. (1)
- _____
- (b) The age of a second-hand car and its resale value. (1)
- _____
- (c) A student's shoe size and their mark in a geography test. (1)
- _____

(Total for Question 1 is 3 marks)

- 2** The formula for the product moment correlation coefficient (PMCC), r , is:

$$r = S_{xy} / \sqrt{S_{xx} * S_{yy}}$$

where $S_{xy} = (\text{sum of } xy) - (\text{sum of } x)(\text{sum of } y)/n$, $S_{xx} = (\text{sum of } x^2) - (\text{sum of } x)^2/n$ and $S_{yy} = (\text{sum of } y^2) - (\text{sum of } y)^2/n$.

For a set of paired data, $S_{xy} = 84$, $S_{xx} = 120$ and $S_{yy} = 98$.

Work out the value of r , giving your answer correct to 3 significant figures.

(Total for Question 2 is 3 marks)

- 3** For 7 employees at a company, x is the number of years they have worked there, and y is the total number of minutes they arrived late to work last month. The scatter graph shows the results.



- (a) Describe the pattern shown by the scatter graph. (1)
- (b) The product moment correlation coefficient for this data is $r = -0.618$. State whether this value is consistent with the pattern shown. (1)
- (c) Explain your answer to part (b). (2)

(Total for Question 3 is 4 marks)

- 4** The table shows the product moment correlation coefficient, r , calculated between two variables in four different studies.

Study	r
A	-0.92
B	0.15
C	0.68
D	-0.40

- (a) Write down which study shows the strongest association between its two variables, regardless of sign. (1)

.....

- (b) Write down which study shows the weakest association between its two variables. (1)

.....

- (c) Write down which studies show a negative association. (1)

.....

- (d) Describe the direction and strength of the association shown in Study C. (1)

(Total for Question 4 is 4 marks)

5 In a seaside town, a student collects data over 12 months and finds a strong positive correlation ($r = 0.89$) between the average number of sunglasses sold per week and the average number of visits to the town's outdoor swimming pool per week.

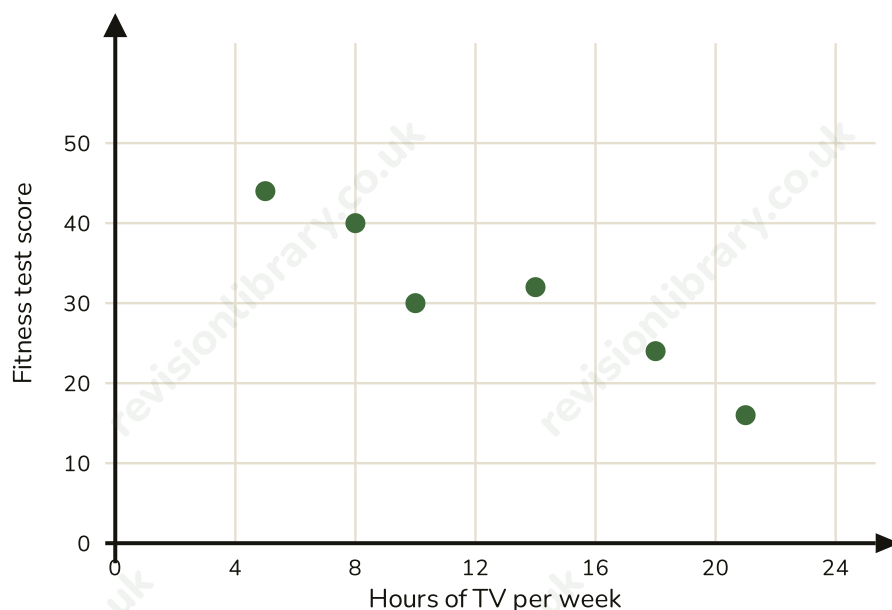
- (a)** State the type of correlation described. (1)
- (b)** Give the name for a variable, such as weather, that could explain a correlation like this without either variable causing the other. (1)
- (c)** Explain why this correlation does not prove that buying sunglasses causes people to visit the swimming pool more often. (2)
- (d)** Suggest one thing the student could do to check whether weather is really responsible for the correlation. (1)

(Total for Question 5 is 5 marks)

- 6 A researcher records, for 6 teenagers, the number of hours of TV watched per week, x , and their score in a fitness test out of 50, y .

Teenager	Hours of TV, x	Fitness score, y
1	5	44
2	8	40
3	10	30
4	14	32
5	18	24
6	21	16

You may use: $S_{xy} = (\text{sum of } xy) - (\text{sum of } x)(\text{sum of } y)/n$, $S_{xx} = (\text{sum of } x^2) - (\text{sum of } x)^2/n$, $S_{yy} = (\text{sum of } y^2) - (\text{sum of } y)^2/n$, and $r = S_{xy} / \sqrt{S_{xx} * S_{yy}}$.



- (a) Calculate S_{xy} , S_{xx} and S_{yy} for this data. (3)
-
- (b) Hence calculate the value of r , giving your answer correct to 3 significant figures. (2)
-
- (c) Interpret the value of r in the context of hours of TV watched and fitness test score. (2)

(Total for Question 6 is 7 marks)

- 7 A national newspaper reports: 'New study finds $r = 0.97$ correlation between a country's average chocolate consumption per person and its number of universities that produce prize-winning scientific research. Eating chocolate boosts brainpower, say researchers.'

Evaluate this claim.

(Total for Question 7 is 6 marks)

Mark scheme · H02-30min Correlation and Causation (Higher): 30-minute Pack

Question 1

- (a) **B1** Positive cao
- (a) **Answer: Positive correlation**
- (b) **B1** Negative cao
- (b) **Answer: Negative correlation**
- (c) **B1** No correlation cao
- (c) **Answer: No correlation**

Question 2

- **M1** substitutes correctly into $r = S_{xy} / \sqrt{S_{xx} * S_{yy}}$
- **A1** $r = 0.774596...$ (or equivalent unrounded value)
- **A1** 0.775 (3 sf), ft
- **Answer: 0.775**

Question 3

- (a) **B1** a generally downward/falling pattern with a fair amount of scatter, oe
- (a) **Answer: The points generally fall from left to right, but with a fair amount of scatter.**
- (b) **B1** Yes cao
- (b) **Answer: Yes, it is consistent.**
- (c) **B1** the negative sign matches the falling pattern of the points
- (c) **B1** the size of r (moderate, not close to -1) matches the amount of scatter around the trend rather than a perfect straight line
- (c) **Answer: The negative sign matches the overall falling trend, and because -0.618 is only moderate in size (not close to -1), this matches the noticeable scatter around that trend rather than a perfect line.**

Question 4

- (a) **B1** A cao
- (a) **Answer: A**
- (b) **B1** B cao
- (b) **Answer: B**
- (c) **B1** A and D cao
- (c) **Answer: A and D**
- (d) **B1** positive, moderate to fairly strong, oe
- (d) **Answer: A moderate to fairly strong positive association.**

Question 5

- (a) **B1** Positive (strong) cao
- (a) **Answer: Strong positive correlation**
- (b) **B1** confounding variable / third variable, oe
- (b) **Answer: A confounding (or third) variable**
- (c) **B1** identifies that a third factor (warm/sunny weather) could increase both sunglasses sales and pool visits independently
- (c) **B1** correct conclusion, e.g. the correlation could be explained by weather rather than one variable causing the other
- (c) **Answer: Warm, sunny weather could independently increase both sunglasses sales and pool visits, so the correlation may simply reflect the weather rather than one thing causing the other.**
- (d) **B1** e.g. also record temperature/hours of sunshine each week and compare weeks with similar weather to see whether the link remains, oe
- (d) **Answer: Record the weather (e.g. temperature) each week too, and compare weeks with similar weather to see if the link between sunglasses sales and pool visits weakens.**

Question 6

- (a) **B1** $S_{xy} = -300$
- (a) **B1** $S_{xx} = 187.3$ (awrt), or $562/3$
- (a) **B1** $S_{yy} = 526$
- **(a) Answer: $S_{xy} = -300$, $S_{xx} = 187.3$ (3 sf), $S_{yy} = 526$**
- (b) **M1** $r = -300 / \sqrt{187.33... * 526}$, oe, ft from (a)
- (b) **A1** -0.956 (3 sf)
- **(b) Answer: -0.956**
- (c) **B1** identifies a strong negative correlation/association
- (c) **B1** correct context, e.g. teenagers who watched more TV tended to get lower fitness scores, oe
- **(c) Answer: There is a strong negative correlation - teenagers who watched more hours of TV per week tended to get lower fitness test scores.**

Question 7

- **B1** correctly reads $r = 0.97$ as a very strong positive correlation between the two variables
- **B1** states that even a very strong correlation does not by itself establish causation
- **B1** identifies a plausible confounding variable, e.g. national wealth/investment in research, which could independently raise both chocolate consumption and university research funding
- **B1** notes that the data is collected at the level of whole countries, so it says nothing about whether an individual person eating more chocolate improves their own brainpower
- **B1** suggests what further evidence or research design would be needed to test a causal claim properly, e.g. controlling for national wealth, or a controlled experiment on individuals
- **B1** reaches a balanced overall conclusion that the newspaper's causal claim is not supported by the correlation alone
- **Answer: The claim is not supported: $r = 0.97$ shows a very strong link between the two country-level variables, but wealthier countries can likely afford both more chocolate and better-funded universities, so national wealth is a plausible confounding variable; the correlation also says nothing about individual people, so it does not show that eating chocolate boosts anyone's own brainpower.**