



H11-test Regression and the PMCC: Test Pack

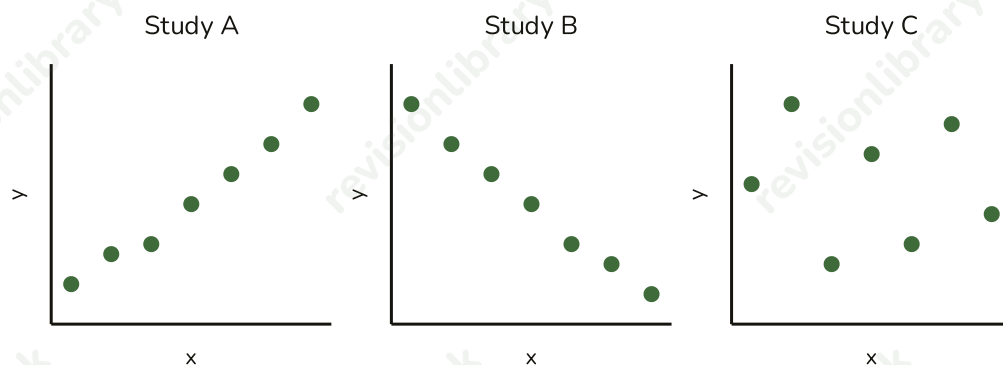
EDEXCEL 1ST0 · Calculator allowed · about 20 minutes

Total Marks

Name: _____ Date: ____ / ____ / ____

Answer ALL questions. Show all your working.

- 1** Each of the following scatter graphs shows paired data collected from a different study.



- (a) State the type of correlation shown in Study A. (1)
- (b) State the type of correlation shown in Study B. (1)
- (c) State the type of correlation shown in Study C. (1)

(Total for Question 1 is 3 marks)

2 Four studies calculate the product moment correlation coefficient, r , between two variables:

A: $r = 0.89$

B: $r = -0.97$

C: $r = 0.15$

D: $r = -0.42$

Match each description below to the correct study.

(i) A very strong negative correlation. (1)

(ii) A strong positive correlation. (1)

(iii) A weak (or negligible) correlation. (1)

(iv) A moderate negative correlation. (1)

(Total for Question 2 is 4 marks)

3 For a sample of 12 days, a museum manager records the number of visitors and the average time (minutes) spent in the gift shop. She calculates $S_{xy} = 245$, $S_{xx} = 410$ and $S_{yy} = 198$.

You may use: $r = S_{xy} / \sqrt{S_{xx} * S_{yy}}$.

(a) Calculate the value of r , giving your answer correct to 3 significant figures. (3)

(b) Interpret the value of r in the context of the number of visitors and time spent in the gift shop. (2)

(Total for Question 3 is 5 marks)

- 4** A student wants to find out whether there is a relationship between the number of hours Year 11 students spend using a revision app each week and their score in a mock maths exam. She plans her investigation using the four stages of the statistical enquiry cycle: plan, collect, process and interpret.
- (a) (Plan) Explain why she needs to record app-usage hours and mock exam score as paired data from the same students, rather than collecting each separately from different groups. (1)
- (b) (Collect) She only surveys students in the top set for maths. Give one reason why this could make her results unreliable for Year 11 as a whole. (1)
- (c) (Process) For her sample of 20 students, she calculates $S_{xy} = 180$, $S_{xx} = 50$ and $S_{yy} = 900$. Calculate the value of r , giving your answer correct to 3 significant figures. (3)
-
- (d) (Interpret) Interpret this value of r , and explain why a strong correlation would not prove that using the revision app causes higher exam scores. (2)

(Total for Question 4 is 7 marks)

Mark scheme · H1 1-test Regression and the PMCC: Test 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

- (i) **B1** B cao
- (i) **Answer: B**
- (ii) **B1** A cao
- (ii) **Answer: A**
- (iii) **B1** C cao
- (iii) **Answer: C**
- (iv) **B1** D cao
- (iv) **Answer: D**

Question 3

- (a) **M1** substitutes correctly into $r = S_{xy} / \sqrt{S_{xx} * S_{yy}}$
- (a) **A1** $r = 0.8599...$ (or equivalent unrounded value)
- (a) **A1** 0.860 (3 sf), ft
- (a) **Answer: 0.860**
- (b) **B1** identifies a strong positive correlation/association
- (b) **B1** correct context, e.g. days with more visitors tend to see more average time spent in the gift shop, oe
- (b) **Answer: There is a strong positive correlation - days with more visitors tend to see more time spent, on average, in the gift shop.**

Question 4

- (a) **B1** a correlation can only be found between two variables recorded for the same individuals, so both measurements must come from the same students, oe
- (a) **Answer: Correlation can only be measured between two variables recorded for the same individuals, so both values must come from the same students.**
- (b) **B1** the top set is not representative of all Year 11 students (e.g. generally higher attainers), so the results may not apply to other students, oe
- (b) **Answer: Top-set students are not representative of all Year 11 students, so the relationship found may not hold for the wider year group.**
- (c) **M1** substitutes correctly into $r = S_{xy} / \sqrt{S_{xx} * S_{yy}}$
- (c) **A1** $r = 0.8485...$ (or equivalent unrounded value)
- (c) **A1** 0.849 (3 sf)
- (c) **Answer: 0.849**
- (d) **B1** correctly interprets $r = 0.849$ as a strong positive correlation, e.g. students who use the app more tend to score higher
- (d) **B1** identifies that this does not prove causation, e.g. a confounding variable such as general motivation or overall time spent revising (not just on the app) could affect both, oe
- (d) **Answer: $r = 0.849$ shows a strong positive correlation - students who used the app more tended to score higher - but this does not prove the app causes higher scores, since a confounding factor (e.g. a generally more motivated student both uses the app more and revises more overall) could explain the link instead.**