HIGHER

4.3 Correlation and Causation (Higher): 1-hour Pack

Total Marks

EDEXCEL 1ST0 · Calculator allowed · about 60 minutes

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

1 A study finds that the product moment correlation coefficient between the number of ice creams sold and the number of sunburn cases is r = -0.35.

(a) Describe the correlation. Use these descriptions: strong is |r| of 0.7 or more, moderate is from 0.4 up to 0.7, weak is below 0.4.

(b) A student says that a change in the number of ice creams sold causes a change in the number of sunburn cases. Comment on this claim.

(Total for Question 1 is 4 marks)

2 The data below could be plotted on a scatter graph. A teacher records hours spent revising (x) and test mark (y).

x	2	6	8	10	14	15	18	19
y	98	89	79	74	69	64	54	52

A line of best fit is drawn so that it passes through the points (6, 89) and (18, 54).

(a) Work out the gradient of the line.

(b) Find the equation of the line in the form y = mx + c.

(Total for Question 2 is 4 marks)

© Revision Library | revisionlibrary.co.uk

3 A study finds that the product moment correlation coefficient between the price of a house and the number of bedrooms is r = 0.45.

(a) Describe the correlation. Use these descriptions: strong is |r| of 0.7 or more, moderate is from 0.4 up to 0.7, weak is below 0.4.

(b) A student says that a change in the price of a house causes a change in the number of bedrooms. Comment on this claim.

(Total for Question 3 is 4 marks)

---

4 For 7 pairs of data (x, y): sum of x = 40, sum of y = 296, sum of xy = 2112, sum of x^2 = 372, sum of y^2 = 13992.

(a) Work out S_{xy} = (sum of xy) - (sum of x)(sum of y)/n and S_{xx} = (sum of x^2) - (sum of x)^2/n.

(b) Work out the gradient b = S_{xy} / S_{xx} of the regression line y = a + bx, to 3 significant figures.

(c) The mean of x is 5.71 and the mean of y is 42.29. Find a, then use the line to estimate y when x = 7. Give each answer to 3 significant figures, using unrounded values in your working.

(Total for Question 4 is 6 marks)

© Revision Library | revisionlibrary.co.uk

5 For 6 pairs of data (x, y): sum of x = 62, sum of y = 330, sum of xy = 3816, sum of x^2 = 728, sum of y^2 = 20608.

(a) Work out S_{xy} = (sum of xy) - (sum of x)(sum of y)/n and S_{xx} = (sum of x^2) - (sum of x)^2/n.

(b) Work out the gradient b = S_{xy} / S_{xx} of the regression line y = a + bx, to 3 significant figures.

(c) The mean of x is 10.33 and the mean of y is 55. Find a, then use the line to estimate y when x = 6. Give each answer to 3 significant figures, using unrounded values in your working.

(Total for Question 5 is 6 marks)

---

6 The data below could be plotted on a scatter graph. A teacher records hours spent revising (x) and test mark (y).

x	2	7	10	11	13	16	17	19
y	72	57	46	40	26	17	15	10

A line of best fit is drawn so that it passes through the points (7, 57) and (17, 15).

(a) Work out the gradient of the line.

(b) Find the equation of the line in the form y = mx + c.

(c) Use your equation to estimate y when x = 12.

(d) State whether your estimate in (c) is interpolation or extrapolation, and whether it is likely to be reliable.

(Total for Question 6 is 7 marks)

© Revision Library | revisionlibrary.co.uk

7 For 6 pairs of data (x, y): sum of x = 55, sum of y = 315, sum of xy = 3233, sum of x^2 = 609, sum of y^2 = 17987.

(a) Work out S_{xy} = (sum of xy) - (sum of x)(sum of y)/n and S_{xx} = (sum of x^2) - (sum of x)^2/n.

(b) Work out the gradient b = S_{xy} / S_{xx} of the regression line y = a + bx, to 3 significant figures.

(c) The mean of x is 9.17 and the mean of y is 52.5. Find a, then use the line to estimate y when x = 18. Give each answer to 3 significant figures, using unrounded values in your working.

(Total for Question 7 is 6 marks)

---

8 A study finds that the product moment correlation coefficient between the number of fire engines at a fire and the damage caused is r = -0.47.

(a) Describe the correlation. Use these descriptions: strong is |r| of 0.7 or more, moderate is from 0.4 up to 0.7, weak is below 0.4.

(b) A student says that a change in the number of fire engines at a fire causes a change in the damage caused. Comment on this claim.

(Total for Question 8 is 4 marks)

© Revision Library | revisionlibrary.co.uk

9 For 6 pairs of data (x, y): sum of x = 49, sum of y = 263, sum of xy = 2467, sum of x^2 = 503, sum of y^2 = 12789.

(a) Work out S_{xy} = (sum of xy) - (sum of x)(sum of y)/n and S_{xx} = (sum of x^2) - (sum of x)^2/n.

(b) Work out the gradient b = S_{xy} / S_{xx} of the regression line y = a + bx, to 3 significant figures.

(c) The mean of x is 8.17 and the mean of y is 43.83. Find a, then use the line to estimate y when x = 5. Give each answer to 3 significant figures, using unrounded values in your working.

(Total for Question 9 is 6 marks)

---

10 The data below could be plotted on a scatter graph. A shop records the number of rainy days in a month (x) and umbrellas sold (y).

x	1	3	4	7	8	10	19	20
y	111	100	100	90	89	81	62	63

A line of best fit is drawn so that it passes through the points (3, 100) and (19, 62).

(a) Work out the gradient of the line.

(b) Find the equation of the line in the form y = mx + c.

(Total for Question 10 is 4 marks)

© Revision Library | revisionlibrary.co.uk

11 The data below could be plotted on a scatter graph. A dealer records the age in years of a car (x) and its value in hundreds of pounds (y).

x	5	10	13	16	17	20
y	57	48	36	38	29	28

A line of best fit is drawn so that it passes through the points (10, 48) and (17, 29).

(a) Work out the gradient of the line.

(b) Find the equation of the line in the form $y = mx + c$.

(Total for Question 11 is 4 marks)

---

12 A study finds that the product moment correlation coefficient between the number of hours of TV watched and a person's weight is $r = 0.15$.

(a) Describe the correlation. Use these descriptions: strong is $|r|$ of 0.7 or more, moderate is from 0.4 up to 0.7, weak is below 0.4.

(b) A student says that a change in the number of hours of TV watched causes a change in a person's weight. Comment on this claim.

(Total for Question 12 is 4 marks)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 4.3 Correlation and Causation (Higher): 1-hour Pack

Question 1

negative	**B1**
weak	**B1**
correlation does not prove causation	**B1**
a third factor, such as hot weather, may affect both	**B1**

Answer: (a) weak negative correlation; (b) correlation does not imply causation; a third factor such as hot weather may be the cause

Question 2

change in y divided by change in x	**M1**
gradient -2.9166666667 cao	**A1**
substitutes a point into y = mx + c	**M1**
y = -2.9166666667x + 106.5 cao	**A1**

Answer: (a) -2.9166666667; (b) y = -2.9166666667x + 106.5

Question 3

positive	**B1**
moderate	**B1**
correlation does not prove causation	**B1**
a third factor, such as the size of the house, may affect both	**B1**

Answer: (a) moderate positive correlation; (b) correlation does not imply causation; a third factor such as the size of the house may be the cause

Question 4

S_{xy} or S_{xx} formula used	**M1**
S_{xy} = 420.57 and S_{xx} = 143.43	**A1**
b = 2.93	**A1**
a = mean of y minus b x mean of x	**M1**
a = 25.5	**A1**
46.1 cao	**A1**

Answer: (a) 420.57, 143.43; (b) 2.93; (c) a = 25.5, y = 46.1

Question 5

S_{xy} or S_{xx} formula used	**M1**
S_{xy} = 406 and S_{xx} = 87.33	**A1**
b = 4.65	**A1**
a = mean of y minus b x mean of x	**M1**
a = 6.96	**A1**
34.9 cao	**A1**

Answer: (a) 406, 87.33; (b) 4.65; (c) a = 6.96, y = 34.9

Question 6

change in y divided by change in x	**M1**
-4.2 cao	**A1**
substitutes a point to find c	**M1**
y = -4.2x + 86.4 cao	**A1**
substitutes x = 12	**M1**
36 cao	**A1**
interpolation and more reliable as within the data range	**B1**

Answer: (a) -4.2; (b) y = -4.2x + 86.4; (c) 36; (d) interpolation; likely to be reasonably reliable because x = 12 is within the observed range 2 to 19 and the data follows a clear linear pattern.

Question 7

S_{xy} or S_{xx} formula used	**M1**
S_{xy} = 345.5 and S_{xx} = 104.83	**A1**
b = 3.30	**A1**
a = mean of y minus b x mean of x	**M1**
a = 22.3	**A1**
81.6 cao	**A1**

Answer: (a) 345.5, 104.83; (b) 3.30; (c) a = 22.3, y = 81.6

Question 8

negative	**B1**
moderate	**B1**
correlation does not prove causation	**B1**
a third factor, such as the size of the fire, may affect both	**B1**

Answer: (a) moderate negative correlation; (b) correlation does not imply causation; a third factor such as the size of the fire may be the cause

Question 9

S_{xy} or S_{xx} formula used	**M1**
S_{xy} = 319.17 and S_{xx} = 102.83	**A1**
b = 3.10	**A1**
a = mean of y minus b x mean of x	**M1**
a = 18.5	**A1**
34.0 cao	**A1**

Answer: (a) 319.17, 102.83; (b) 3.10; (c) a = 18.5, y = 34.0

Question 10

change in y divided by change in x	**M1**
gradient -2.375 cao	**A1**
substitutes a point into y = mx + c	**M1**
y = -2.375x + 107.125 cao	**A1**

Answer: (a) -2.375; (b) y = -2.375x + 107.125

Question 11

change in y divided by change in x	**M1**
gradient -2.7142857143 cao	**A1**
substitutes a point into y = mx + c	**M1**
y = -2.7142857143x + 75.1428571429 cao	**A1**

Answer: (a) -2.7142857143; (b) y = -2.7142857143x + 75.1428571429

Question 12

positive	**B1**
weak	**B1**
correlation does not prove causation	**B1**
a third factor, such as overall lifestyle, may affect both	**B1**

Answer: (a) weak positive correlation; (b) correlation does not imply causation; a third factor such as overall lifestyle may be the cause

© Revision Library | revisionlibrary.co.uk