

1. On a particular day the height above sea level, x metres, and the mid-day temperature, y °C, were recorded in 8 north European towns. These data are summarised below

$$S_{xx} = 3\,535\,237.5 \quad \sum y = 181 \quad \sum y^2 = 4305 \quad S_{xy} = -23\,726.25$$

- (a) Find S_{yy} . (2)

- (b) Calculate, to 3 significant figures, the product moment correlation coefficient for these data. (2)

- (c) Give an interpretation of your coefficient. (1)

A student thought that the calculations would be simpler if the height above sea level, h , was measured in kilometres and used the variable $h = \frac{x}{1000}$ instead of x .

- (d) Write down the value of S_{hh} . (1)

- (e) Write down the value of the correlation coefficient between h and y . (1)

(Total 7 marks)

2. A teacher asked a random sample of 10 students to record the number of hours of television, t , they watched in the week before their mock exam. She then calculated their grade, g , in their mock exam. The results are summarised as follows.

$$\sum t = 258 \quad \sum t^2 = 8702 \quad \sum g = 63.6 \quad S_{gg} = 7.864 \quad \sum gt = 1550.2$$

- (a) Find S_{tt} and S_{gt} .

(3)

- (b) Calculate, to 3 significant figures, the product moment correlation coefficient between t and g .

(2)

The teacher also recorded the number of hours of revision, v , these 10 students completed during the week before their mock exam. The correlation coefficient between t and v was -0.753 .

- (c) Describe, giving a reason, the nature of the correlation you would expect to find between v and g .

(2)

(Total 7 marks)

3. A random sample of 50 salmon was caught by a scientist. He recorded the length l cm and weight w kg of each salmon.

The following summary statistics were calculated from these data.

$$\sum l = 4027 \quad \sum l^2 = 327\,754.5 \quad \sum w = 357.1 \quad \sum lw = 29\,330.5 \quad S_{ww} = 289.6$$

- (a) Find S_{ll} and S_{lw} .

(3)

- (b) Calculate, to 3 significant figures, the product moment correlation coefficient between l and w .

(2)

- (c) Give an interpretation of your coefficient.

(1)

(Total 6 marks)

4. An estate agent recorded the price per square metre, p £/m², for 7 two-bedroom houses.

He then coded the data using the coding $q = \frac{p-a}{b}$, where a and b are positive constants.

His results are shown in the table below.

p	1840	1848	1830	1824	1819	1834	1850
q	4.0	4.8	3.0	2.4	1.9	3.4	5.0

(a) Find the value of a and the value of b .

(2)

The estate agent also recorded the distance, d km, of each house from the nearest train station. The results are summarised below.

$$S_{dd} = 1.02 \quad S_{qq} = 8.22 \quad S_{dq} = -2.17$$

(b) Calculate the product moment correlation coefficient between d and q .

(2)

(c) Write down the value of the product moment correlation coefficient between d and p .

(1)

The estate agent records the price and size of 2 additional two-bedroom houses, H and J .

House	Price (£)	Size (m ²)
H	156 400	85
J	172 900	95

(d) Suggest which house is most likely to be closer to a train station. Justify your answer.

(3)

(Total 8 marks)

5. A bank reviews its customer records at the end of each month to find out how many customers have become unemployed, u , and how many have had their house repossessed, h , during that month. The bank codes the data using variables $x = \frac{u-100}{3}$ and $y = \frac{h-20}{7}$.

The results for the 12 months of 2009 are summarised below.

$$\sum x = 477 \quad S_{xx} = 5606.25 \quad \sum y = 480 \quad S_{yy} = 4244 \quad \sum xy = 23\,070$$

- (a) Calculate the value of the product moment correlation coefficient for x and y . **(3)**

- (b) Write down the product moment correlation coefficient for u and h . **(1)**

The bank claims that an increase in unemployment among its customers is associated with an increase in house repossessions.

- (c) State, with a reason, whether or not the bank's claim is supported by these data. **(2)**

(Total 6 marks)

6. A researcher believes that parents with a short family name tended to give their children a long first name. A random sample of 10 children was selected and the number of letters in their family name, x , and the number of letters in their first name, y , were recorded.

The data are summarised as:

$$\sum x = 60, \quad \sum y = 61, \quad \sum y^2 = 393, \quad \sum xy = 382, \quad S_{xx} = 28$$

- (a) Find S_{yy} and S_{xy} (3)
- (b) Calculate the product moment correlation coefficient, r , between x and y . (2)
- (c) State, giving a reason, whether or not these data support the researcher's belief. (2)

The researcher decides to add a child with family name "Turner" to the sample.

- (d) Using the definition $S_{xx} = \sum (x - \bar{x})^2$, state the new value of S_{xx} giving a reason for your answer. (2)

Given that the addition of the child with family name "Turner" to the sample leads to an increase in S_{yy}

- (e) use the definition $S_{xy} = \sum (x - \bar{x})(y - \bar{y})$ to determine whether or not the value of r will increase, decrease or stay the same. Give a reason for your answer. (2)

(Total 11 marks)

TOTAL FOR PAPER: 45 MARKS