



Scatter diagrams, regression and correlation coefficients

About 85 minutes Calculator allowed 75 marks

Date _____

A Do Now

The first two are what this chapter needs. The rest bring back earlier chapters so nothing is forgotten.

A1 Needed today [2 marks]
Find the gradient of the line through $(5, 2)$ and $(-3, 0)$.

Answer

A2 Needed today [2 marks]
An overall mark is worked out from Paper 1, worth 35%, and Paper 2, worth 65%. Liam scored 93 in Paper 1 and 57 in Paper 2. Work out Liam's overall mark.

Answer

A3 From chapter 2 [2 marks]
There are 1800 fish in a lake. Sample A is 20 of them, and 9 are over 20 cm long. Sample B is 120 of them, and 32 are over 20 cm long. Use the more reliable sample to estimate how many of the 1800 are over 20 cm long, to the nearest whole number.

Answer

A4 From chapter 2 [2 marks]
Isla numbers 260 people from 1 to 260. The calculator's random number key gives 0.893. Isla multiplies it by 260 and rounds up to choose a person. Which person is chosen?

Answer

Maya wants to find out how many students in a school own a pet. Maya plans a stratified sample by whether each student owns a pet. Is this a sensible variable to stratify by? Give a reason.

B

Examples, then your turn

Your teacher works through each example with you. Then try the one beside it on your own. Spearman's rank correlation coefficient is $r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$.

EXAMPLE 1

For nine months, the scatter diagram shows how many adverts a shop placed in the local paper and how many customers it had. Describe the correlation, then say what it means for the shop.

Adverts in the month	Customers (hundreds)
2.5	6
3.5	9
4.5	8
5.5	12
6.5	14
7.5	16
8.5	19
9.5	21
10.5	23

Hint

Name the type of correlation, then say what it means for adverts and customers.

B1

The scatter diagram shows the height above sea level and the temperature at eight weather stations at noon. Describe the correlation, and say what it shows about height and temperature.

Height above sea level (m)	Temperature (°C)
100	19
200	18
300	16
400	15
500	14
600	12
700	11
800	10

EXAMPLE 2

The cooking time, T minutes, for a chicken of mass m kg is $T = 20 + 40m$. Interpret the 20 and the 40 in this context.

Hint

The 40 is what is added for each extra kg. The 20 is what is there when $m = 0$.

B2

The monthly phone bill, $\pounds C$, for t minutes of calls is $C = 8 + 0.05t$. Interpret the 8 and the 0.05 in this context.

EXAMPLE 3

For laptops from 1 to 8 years old, the value, £ y , is given by $y = 900 - 75x$, where x is the age in years. Estimate the value of a 5-year-old laptop and of a 14-year-old laptop. Comment on the reliability of each.

Hint Substitute each age. Then check whether the age is inside the range of the data.

B3 For sunflowers from 2 to 10 weeks old, the height, h cm, is given by $h = 12 + 9w$, where w is the age in weeks. Estimate the height of a sunflower at 6 weeks and at 20 weeks. Which estimate is more reliable? Give a reason. [3 marks]**EXAMPLE 4**

The table shows the population and the number of libraries in six towns. Work out Spearman's rank correlation coefficient, correct to 3 decimal places.

Town	P	Q	R	S	T	U
Population (thousands)	120	85	240	60	150	95
Libraries	14	9	22	8	12	11

Hint Rank the populations, largest as 1. Rank the libraries the same way.

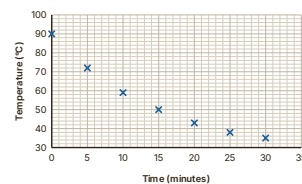
B4 Eight used cars, A to H. Ages in years: 1, 3, 5, 2, 7, 4, 6, 8. Prices in £: 9800, 6100, 5200, 8800, 3900, 7900, 3300, 29800. Find Spearman's rank correlation coefficient, to 2 decimal places. [4 marks]

Answer

EXAMPLE 5

Two judges each rank eight photographs. Should the PMCC or Spearman's rank correlation coefficient be used to measure how well they agree? Give a reason.

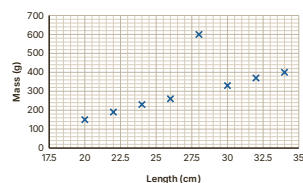
Hint Are the data measurements or ranks?

B5 The scatter diagram shows the temperature of a hot drink as it cools. Should the PMCC or Spearman's rank correlation coefficient be used to measure the correlation? Give a reason. [2 marks]

c Practice

Each question asks for something different. Show your working.

- C1** The scatter diagram shows the length and mass of eight fish. Write down the length and mass of the fish that does not fit the pattern. A scientist checks this fish and finds the mass was written down as 600 g instead of 300 g. What should she do with this point? [2 marks]



- C2** The table shows six pairs of values. Work out the coordinates of the double mean point. [2 marks]

x	3	5	6	8	9	11
y	20	26	25	33	34	42

Answer (.....,.....)

- C3** A café records the temperature, x °C, and the number of cold drinks sold, y , on 20 days. [2 marks]
The regression line is $y = a + 1.5x$. The mean temperature is 12 °C and the mean number of cold drinks is 40. Work out the value of a .

Answer

- C4** The table shows how two judges ranked six singers. Calculate r_s for the two judges, to 3 decimal places. [3 marks]

Singer	A	B	C	D	E	F
Judge 1	1	2	3	4	5	6
Judge 2	2	1	5	3	6	4

Answer

- C5** For 12 students, Spearman's rank correlation coefficient between their height and their score in a spelling test is 0.04. Interpret this value in context. [1 mark]

- C6** A farmer records the rainfall each month and the mass of potatoes his fields produce. [2 marks]
Which is the explanatory variable? Give a reason.

- C7** The number of fire engines sent to a fire and the cost of the damage show strong positive correlation. Lucy says, "Sending more fire engines causes more damage." Explain why Lucy is wrong. [2 marks]

- C8** The PMCC between the mass of some parcels and the cost of posting them is 0.86. Kat draws the scatter diagram again with the axes swapped, so the cost goes across and the mass goes up. Write down the PMCC for her new diagram. [1 mark]

Answer

D Exam-style questions

Answer every question in the space given. Show your working: most of the marks are for method.

- D1** The table shows the engine size and the fuel economy of eight cars.

Engine size (litres)	1.0	1.2	1.4	1.6	2.0	2.4	3.0	3.4
Fuel economy (mpg)	52	45	49	43	39	35	30	27

- (a) Work out the coordinates of the double mean point. [2 marks]

Answer (.....,.....)

- (b) The regression line is $y = 60 - 10x$. Show that it passes through the double mean point. [1 mark]

Answer

- (c) Interpret the gradient -10 in context. [1 mark]

- (d) Use the regression line to estimate the fuel economy of a car with a 2.6 litre engine. Comment on the reliability of your estimate. [2 marks]

Total for D1: 6 marks

- D2** For eight hospitals, Spearman's rank correlation coefficient between the number of nurses and the number of beds is 1. Raj says, "So the points on the scatter diagram must lie on a straight line." [2 marks]

Is Raj correct? Give a reason for your answer.

- D3** Mia's hypothesis is: "Gyms that charge more have fewer members." The table shows data for seven gyms.

Gym	A	B	C	D	E	F	G
Monthly fee (£)	18	25	32	22	40	29	35
Members (tens)	64	51	38	57	30	42	45

- (a) Calculate Spearman's rank correlation coefficient for these gyms. Round to 2 decimal places. [3 marks]

Answer

- (b) Does this value support Mia's hypothesis? Give a reason. [2 marks]

Total for D3: 5 marks

- D4** On days between 12 °C and 28 °C, a shop's daily sales of ice lollies, y , and the temperature, x °C, give the regression line $y = -20 + 6x$.

- (a) Estimate how many ice lollies the shop sells on a 20 °C day. [1 mark]

Answer

- (b) Ben uses the line for a day when the temperature is 2 °C and gets -8 ice lollies. [2 marks]
Give two reasons why this estimate is wrong.

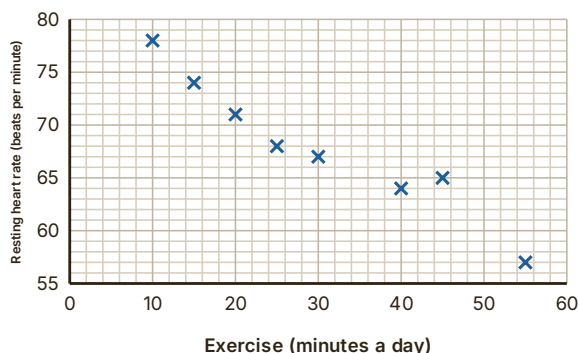
Total for D4: 3 marks

- D5** Leo wants to find out whether taller students have longer arm spans. His plan is to measure the heights of 20 students in Year 10 and the arm spans of 20 students in Year 11, then work out the PMCC. Assess Leo's plan. [3 marks]

- D6** For adults with heights from 150 cm to 190 cm, the mass, y kg, and the height, x cm, give the regression line $y = -40 + 0.6x$. Explain why the value -40 has no meaning in this context. [2 marks]

- D7** For 12 towns, Spearman's rank correlation coefficient between population and the number of shops is 0.95, and the PMCC is 0.71. What do these two values tell you about the scatter diagram? [2 marks]

- D8** The scatter diagram shows the minutes of exercise a day and the resting heart rate of eight adults. The double mean point is (30, 68).



- (a) Describe the correlation. Interpret it in context. [2 marks]

- (b) Draw a line of best fit that passes through (30, 68). Use it to estimate the resting heart rate of an adult who exercises for 35 minutes a day. [2 marks]

Answer

Total for D8: 4 marks

E Extension

No route is given. Spend five minutes on a problem before you look at a hint. The hints are at the end of the sheet. Use one at a time.

- E1** For 8 pairs of data, $\sum x = 96$ and $\sum y = 340$. The regression line $y = a + bx$ has gradient 2.5. Work out a , and estimate y when $x = 14$. [3 marks]

Answer $a = \dots\dots$ $y = \dots\dots$

- E2** Two judges rank 10 cakes. Judge B's ranks are Judge A's ranks with some pairs of neighbouring ranks swapped (for example ranks 4 and 5 swapped), and no rank is in more than one swap. Spearman's rank correlation coefficient is 0.964 to 3 decimal places. How many swaps are there? [4 marks]

Answer

- E3** Eight points have double mean point (6, 20) and a PMCC of 0.6. A ninth point, (6, 20), is added. Write down the new PMCC, and explain why it does not change. [3 marks]

.....

F How did it go?

Colour one circle on each line.

	Not yet	Nearly	Yes
I can describe correlation from a scatter diagram in context, and I know it does not show cause.	☐	☐	☐
I can use a regression line through the double mean point, interpret a and b , and judge a prediction.	☐	☐	☐
I can work out Spearman's rank from raw data and interpret it in context.	☐	☐	☐
I can interpret the PMCC and say when Spearman's rank is the better choice.	☐	☐	☐

Extension hints

Use one hint at a time. Cover the ones below it with a sheet of paper.

- E1**
- Hint 1** The regression line passes through the double mean point.
- Hint 2** $\bar{x} = 96 \div 8 = 12$ and $\bar{y} = 340 \div 8 = 42.5$.
- Hint 3** $42.5 = a + 2.5 \times 12$. Then substitute $x = 14$.
-
- E2**
- Hint 1** Swapping two neighbouring ranks gives two values of d . What are they?
- Hint 2** Each swap gives $d = 1$ and $d = -1$, so it adds 2 to $\sum d^2$.
- Hint 3** With k swaps, $r_s = 1 - \frac{6 \times 2k}{10 \times 99}$. Try $k = 1, 2, 3, \dots$
-
- E3**
- Hint 1** Work out the new double mean point.
- Hint 2** It is still $(6, 20)$. How far is the new point from it?
- Hint 3** The PMCC is worked out from each point's distances from the mean point. The new point's distances are 0.