

Correlation and Regression – GCSE Statistics**1. No.1****1**

Two variables have no correlation.

Which of these could be the value of the Spearman's Rank Correlation Coefficient between the two variables?

Circle your answer.

[1 mark]

1

– 1

0.05

0.5

2. No.12

12

Large car parks often have signs indicating the number of spaces available.

Councils can collect these data to judge whether they have enough car parks.

Negan, a council worker, records data hourly from 8 am to 5 pm about the spaces available in the main city centre car park.

The car park has 800 spaces.

Number of hours after 8 am	0	1	2	3	4
Spaces available	632	176	34	8	0

Number of hours after 8 am	5	6	7	8	9
Spaces available	0	0	25	106	447

- 12 (a) Estimate the **percentage** of this car park's spaces available at 9.30 am.

1 hour 30 minutes after 8 am

You **must** show your working.

$$\% = \frac{n}{800} \times 100\% \quad , \quad \text{where } 34 < n < 176$$

[2 marks]

$$\text{Let } n = 100$$

$$\% = \frac{100}{800} \times 100\% = 12.5\%$$

Answer 12.5 %

- 12 (b) The Product Moment Correlation Coefficient (PMCC) is calculated for the 5 data pairs in the top table and separately for the 5 data pairs in the bottom table.

By looking at patterns in the data, use **one** of the values listed below to complete each statement.

[2 marks]

1.25	0.831	0	-0.016	-0.845
------	-------	---	--------	--------

The PMCC for the top table is

-0.845

The PMCC for the bottom table is

0.831

12 (c) Interpret your answers to **part (b)** in context.

[2 marks]

The PMCC for the top table shows a strong negative correlation between number of hours after 8 am and the spaces available.

The PMCC for the bottom table shows a strong positive correlation between number of hours (5 hours) after 8 am and the spaces available.

12 (d) Lucille looks at the data and says,

"This car park is full for most of the day which shows there are not enough car parks."

Criticise both parts of her statement.

[2 marks]

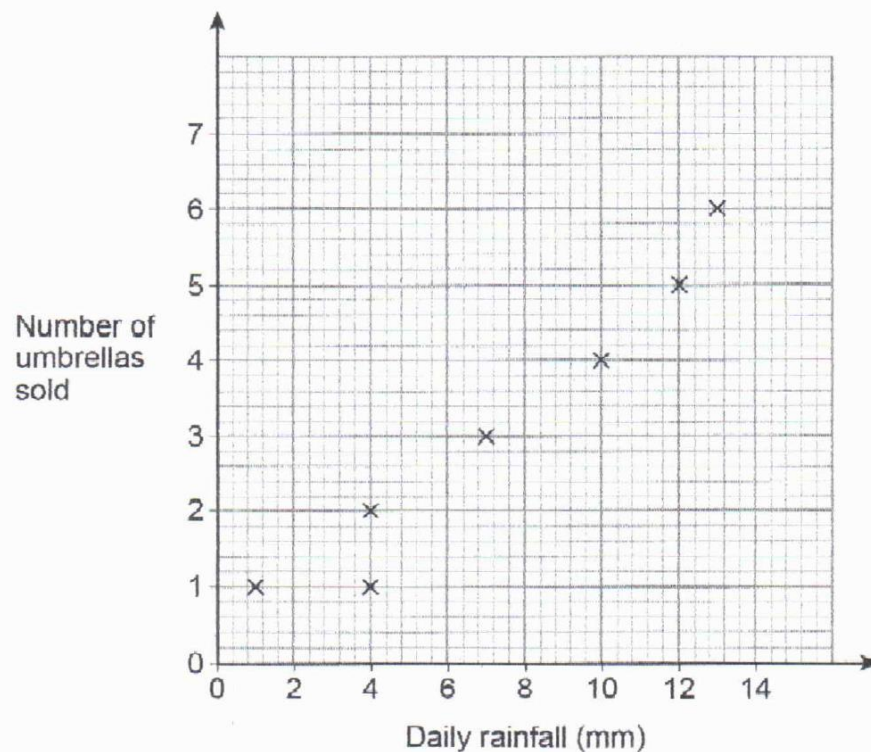
"This car park is full for most of the day..." The car park is only full for about 2 to 4 hours during the day

"...which shows there are not enough car parks." The data sample was recorded for only one day which is not representative.

3. No.12

12

Caro has plotted this scatter diagram to show the numbers of umbrellas she sold against the amount of rain that fell each day.



- 12 (a) Caro can see from the scatter diagram that the greater the rainfall, the greater the number of umbrellas sold.

Tess says that if Caro sells more umbrellas, there will be greater rainfall.

Give a reason why Tess is **not** correct.

[1 mark]

The amount of rain is not affected by the number of umbrellas sold.

- 12 (b) The weather forecast for tomorrow suggests that there will be 20 mm of rain throughout the day.

Caro says she can use the scatter diagram to predict how many umbrellas she would sell if the rainfall was 20 mm.

Is she correct?

Tick (✓) a box.

[1 mark]

Yes

☐

No

☒

Give a reason for your answer.

The highest value from the graph is 14 mm. 20 mm is greater than 14 mm so we will have to extrapolate which may lead to inaccuracy.

4. No.11

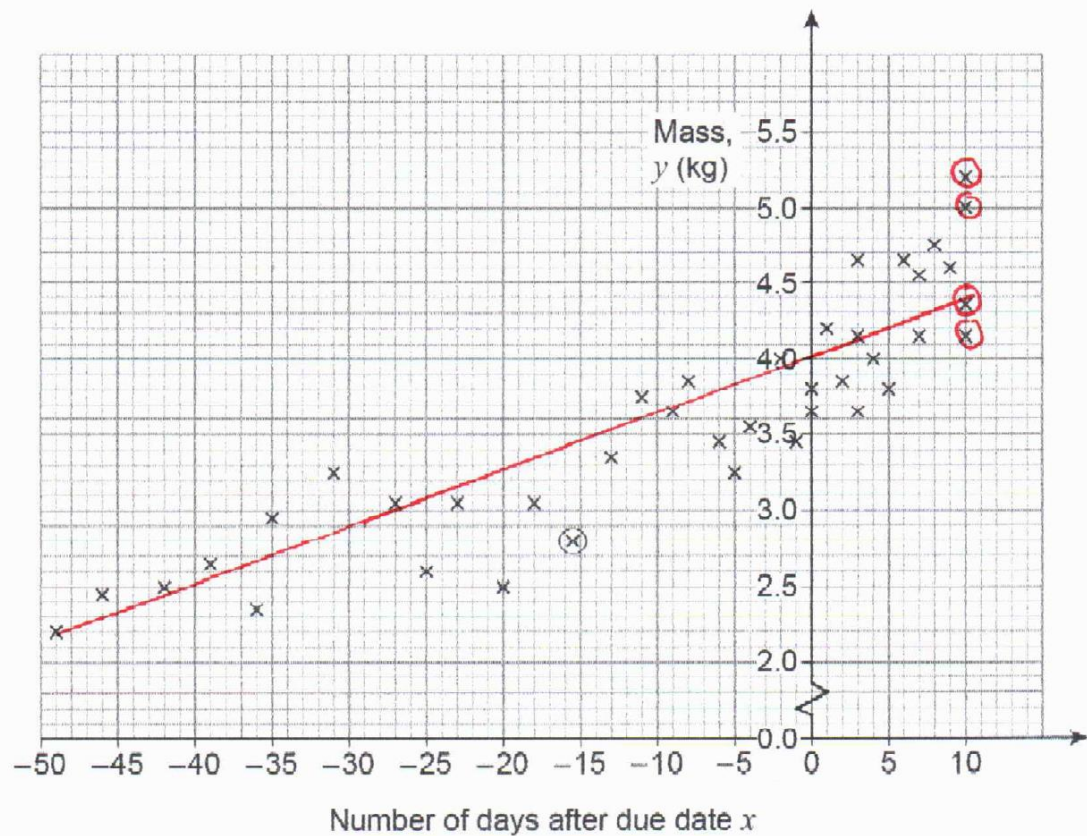
11

The due date of a baby is the date on which it is expected to be born.

The scatter diagram shows the mass of 40 new-born babies **born** on March 20th 2021 plotted against the number of days the babies were born after their due date in whole days.

For example,

Value of number of days after due date	Interpretation
-10	Baby born 10 days before due date
5	Baby born 5 days after due date



11 (a) How many of **these** babies had a due date of March 10th 2021?

Circle your answer.

1

2

3

4

Number of babies born 10 days after the due date.

[1 mark]

- 11 (b) The circled point is an incorrect plot.

Give a reason how you know this.

[1 mark]

The number of days after due date must be a whole number.

- 11 (c) The error is corrected.

The equation of the line of best fit for the data is $y = 4.01 + 0.04x$

- 11 (c) (i) Interpret the value 4.01 in the context of this scatter graph.

[1 mark]

4.01 is the expected mass in Kilograms of a baby born on its due date.

- 11 (c) (ii) Interpret the value 0.04 in the context of this scatter graph.

[1 mark]

0.04 represents gradient, so its the increase in the baby's mass in kilograms for every additional day.

- 11 (c) (iii) Draw the line of best fit on the graph.

[2 marks]

11 (d) It is later found that there were two more babies born on March 20th.

- Sam was born 15 days before his due date.
- Nim was born 15 days after her due date.

Discuss the valid use of the scatter graph to estimate the birth mass of each of these babies.

Where an estimate is possible, explain your method and give the value.

[3 marks]

Sam It is possible to estimate the mass since it's an interpolation. (-15 is within the data range).
Mass (kg) = 3.4 (from the line of best fit)

Nim The mass cannot be estimated since (15) is outside the data range.

5. No.13

13

A teacher marks some mock GCSE Statistics exam papers.

Each student sits two papers.

The teacher wants to see how the marks scored on the two papers are related.

She decides to calculate the value of Spearman's rank correlation coefficient (SRCC) using the formula

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

5 students sit the mock papers and the value of $r_s = 0.8$

- 13 (a) Interpret the value of SRCC in this context.

[1 mark]

There is a positive correlation between the marks scored on the two papers.

- 13 (b) A sixth student now sits the papers.
Both papers are ranked 6th.

- 13 (b) (i) How will including the additional pair of papers change the value of SRCC?

[1 mark]

The SRCC increases to become closer to 1.

13 (b) (ii) Calculate the new value of SRCC for all 6 students.

[4 marks]

For 5 students:

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2-1)} = 0.8$$

$$\Rightarrow 1 - \frac{6 \sum d^2}{5(5^2-1)} = 0.8$$

$$1 - 0.8 = \frac{6 \sum d^2}{120}$$

$$0.2 = \frac{6 \sum d^2}{20}$$

$$\Rightarrow \sum d^2 = 20 \times 0.2 = 4$$

For all 6 students:

$$r_s = 1 - \frac{6 \times 4}{6(6^2-1)}$$

$$= 1 - 0.1143$$

$$= 0.8857$$

$$\therefore r_s = 0.886 \text{ (3 s.f.)}$$

$$\text{or } r_s = 0.89 \text{ (2 s.f.) or } = 0.9 \text{ (1 s.f.)}$$

Answer 0.89