AQA - Correlation and Regression – GCSE Statistics – 2021

1. June/2021/Paper_1F/No.3

Circle the value that indicates a strong correlation.

[1 mark]

1.2 – 0.86 0.51 0

2. June/2021/Paper_1F/No.9(b)

(b) Dr Cho wants to investigate any relationship that might exist between the length of an appointment and the age of the patient.

She collects data from a random sample of 20 patients.

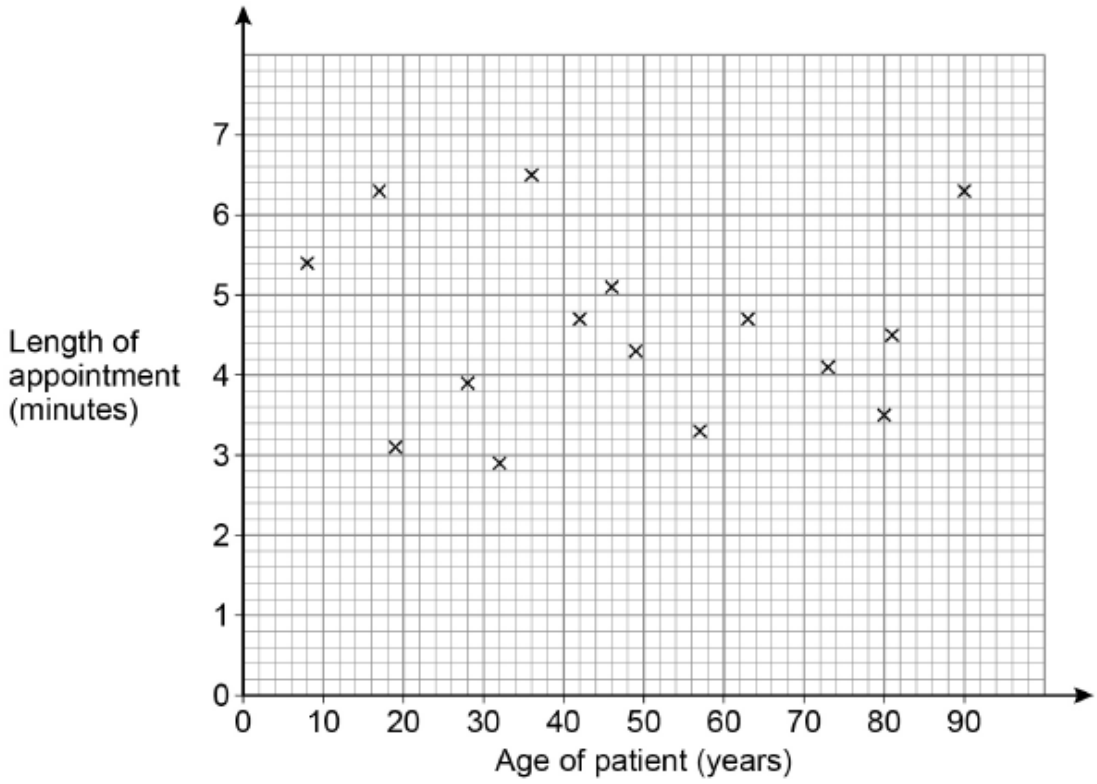
The scatter diagram shows 15 of the results.

The table shows the remaining 5 results.

Age of patient (years)	12	26	40	55	76
Length of appointment (minutes)	3.1	2.4	4.5	2.5	5.8

(b) (i) Use the data in the table to complete the scatter diagram.

[2 marks]



(b) (ii) Dr Cho says she can predict the length of an appointment if she knows the age of a patient.

Comment on her statement.

[1 mark]

3. June/2021/Paper_1H/No.8

Rachel and Mitch own a café.

They want to decide whether more staff are needed at certain times.

On one day, Rachel records how long customers wait to be served at specific times **after 9am**.

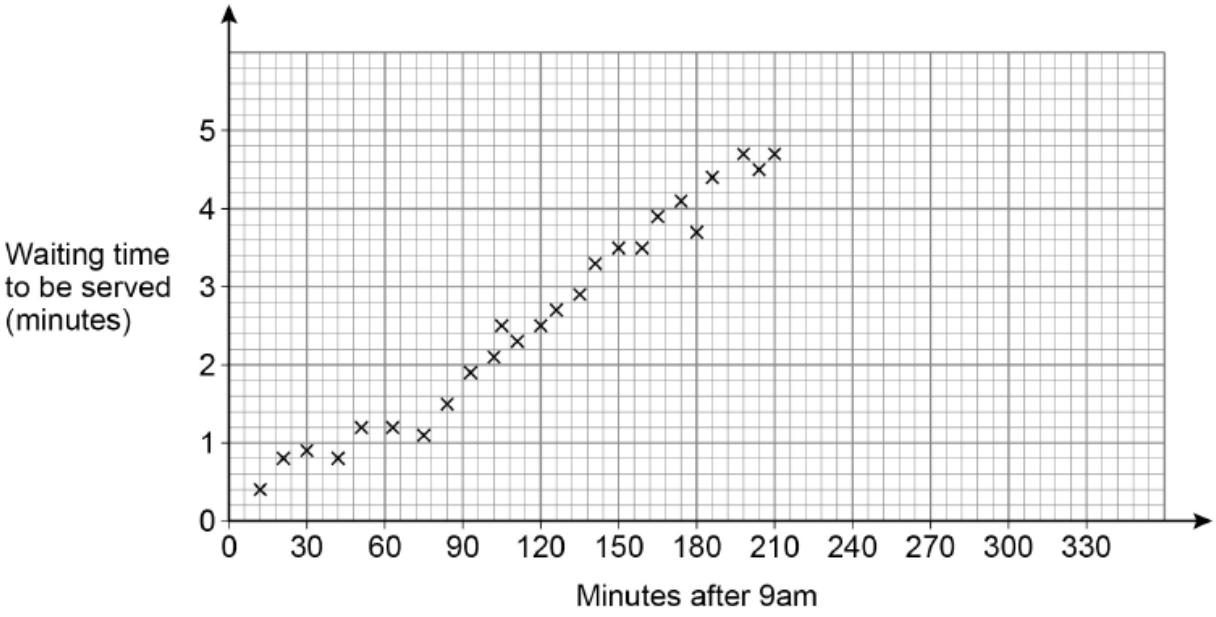
(a) Write down the independent and dependent variables.

[2 marks]

Independent variable _____

Dependent variable _____

(b) The scatter diagram shows the results for 25 customers Rachel sampled.



(b) (i) Which of these is a possible equation for a regression line for the data shown?

Circle your answer.

[1 mark]

$y = -0.06 + 0.689x$

$y = -0.06 + 0.023x$

$y = 4.8 - 0.689x$

$y = 4.8 - 0.023x$

(b) (ii) Here is some information about the results sampled for later in the day.

- After 12.30pm but before 1.30pm all five customers sampled waited between 4 and 5 minutes.
- From 1.30pm, the six results showed a strong negative correlation.

On the scatter diagram, show possible results for these additional 11 sampled customers.

[2 marks]

(b) (iii) Mitch decides to employ an extra person for a two-hour time period.

Which time period would you suggest?

[1 mark]

Answer _____ to _____

(c) Lucy is a statistician who visits the café.

She identifies a problem with Rachel's data collection strategy and offers a solution.

Describe the problem and the solution Lucy may have suggested.

[2 marks]

Problem _____

_____

_____

Solution _____

_____

_____

4. June/2021/Paper_1H/No.10

Fifteen people apply to take part in an 'IronMan' triathlon.

Before they can compete they must prove they are fit enough to attempt the course.

Angelina collects data on each competitor's,

- age
- blood pressure
- resting heart rate.

(a) Angelina thinks that finishing position in the triathlon (1st, 2nd, 3rd, etc) and resting heart rate will show positive correlation.

Write a hypothesis for her to investigate.

[1 mark]

(b) Once the race is run, Angelina notes the order in which the competitors finished, from 1st to 15th position.

She also ranks the resting heart rate data from lowest to highest.

She calculates that the value of $\sum d^2 = 60$, where d is the difference in the ranks of the finishing position and the resting heart rates.

Show that the value of Spearman's Rank Correlation Coefficient (SRCC) is 0.89 (to 2 decimal places).

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

[3 marks]

(c) Write a possible conclusion to your hypothesis in **part (a)**.

[1 mark]

5. June/2021/Paper_2H/No.11

In this question,

- SRCC is Spearman's rank correlation coefficient.
- PMCC is product moment correlation coefficient.

Tick (✓) the appropriate box for each statement.

[3 marks]

	True	False
If SRCC has a value of +1, PMCC must also have a value of +1		
If PMCC has a value of +1, SRCC must also have a value of +1		
If PMCC has a value of – 1, the data must form a straight line on a scatter diagram		
If SRCC has a value of – 1, the data must form a straight line on a scatter diagram		