

Name: _____

GCSE Statistics

Scatter Diagrams

Total marks available: 51

Total marks achieved: _____

Instructions

- Use black ink or ball-point pen.
- Fill in the boxes at the top of this page with your name, Centre number and candidate number.
- Answer all questions.
- Answer the questions in the spaces provided
– There may be more space than you need.
- Scientific calculators may be used.
- You must show all your working out with your answer clearly identified
At the end of your solution.

Information

- The marks for each question are shown in brackets
– use this as a guide as to how much time to spend on each question.

Advice

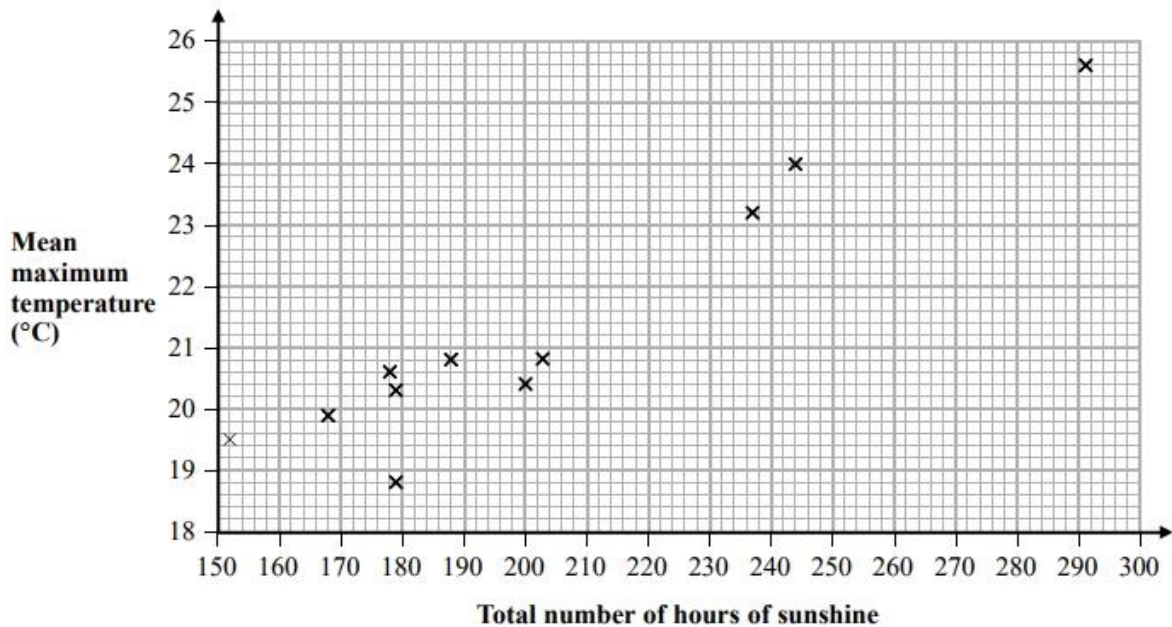
- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.

Q1.

The maximum temperature ($^{\circ}\text{C}$) and the number of hours of sunshine were recorded in Sheffield each day in July for 11 successive years.

For each July, the mean maximum temperature and the total number of hours of sunshine were calculated.

The scatter diagram shows this information.



(Source: Metoffice)

In one of these years there was a total of 244 hours of sunshine in July.

(a) For this year, write down the mean maximum temperature.

..... $^{\circ}\text{C}$
(1)

(b) For the year with the lowest total number of hours of sunshine in July, write down the mean maximum temperature.

..... $^{\circ}\text{C}$
(1)

(c) Draw a line of best fit on the scatter diagram.

(1)

(d) Describe and interpret the type of correlation shown by the scatter diagram.

.....

(3)

For a different year in Sheffield, there was a total of 220 hours of sunshine in July.

(e) (i) Estimate the mean maximum temperature for July that year.

..... °C

(ii) Give a reason why your answer to part (e)(i) should be reliable.

.....

.....

(2)

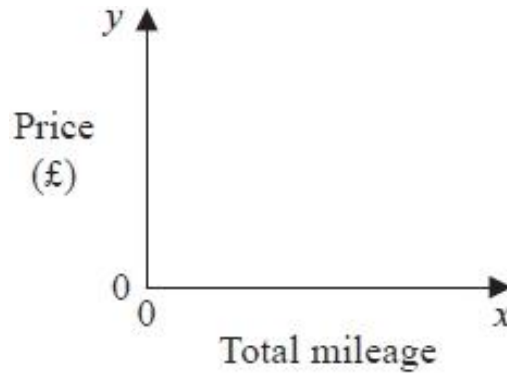
(Total for question = 8 marks)

Q2.

Mark is investigating how the total mileage, x miles, of a car affects its price, $\pounds y$

He has collected information about two models of car, model A and model B. He found the total mileage and the price of each of ten cars for each model. He used his information to draw two scatter diagrams, one for each model.

Here is a sketch of the axes he used for each scatter diagram.



On each scatter diagram, Mark drew a line of best fit. For each line he calculated its gradient and found its intercept on the y -axis.

Here are his results.

Model	Gradient of line of best fit	Intercept of line of best fit on the y -axis
A	-0.135	13 500
B	-0.105	20 500

Interpret and compare these results in context.

.....

.....

.....

.....

.....

.....

.....

.....

(Total for question = 5 marks)

Q3.

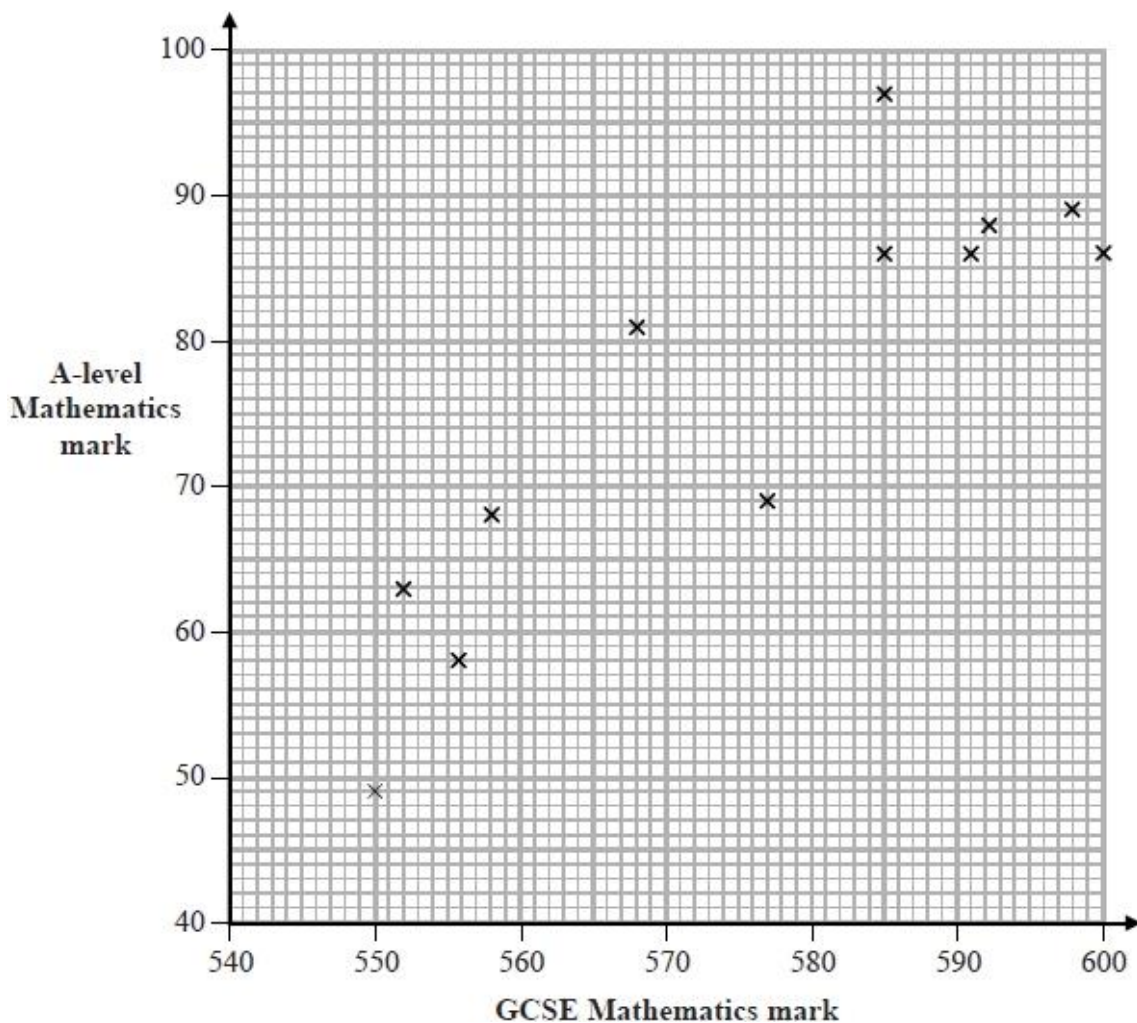
Julie was investigating the relationship between the marks gained by students in their GCSE Mathematics exam and the marks gained by the same students in an A-level Mathematics exam.

(a) Suggest a hypothesis Julie could use.

.....
.....

(1)

Julie drew a scatter diagram using the marks gained in GCSE Mathematics and the marks gained in A-level Mathematics by each of 12 students.



(b) For this scatter diagram, explain why the GCSE Mathematics mark is the explanatory variable.

.....
.....
.....

(1)

(c) Explain, giving a statistical reason, whether or not the scatter diagram supports your hypothesis in part (a).

.....
.....
.....

(2)

Using statistical software, Julie obtained the following information for her 12 students.

Mean GCSE Mathematics mark	578
Mean A-level Mathematics mark	78
Gradient of line of best fit	0.7

(d) Using this information, draw a line of best fit on the scatter diagram.

(2)

(e) Interpret the gradient of the line of best fit.

.....
.....

(1)

Julie wants to use the line of best fit to predict the A-level Mathematics mark for a 13th student. For GCSE Mathematics this student gained a mark of 540

(f) Explain whether or not it would be appropriate to use the line of best fit to make her prediction.

.....
.....

(1)

(Total for question = 8 marks)

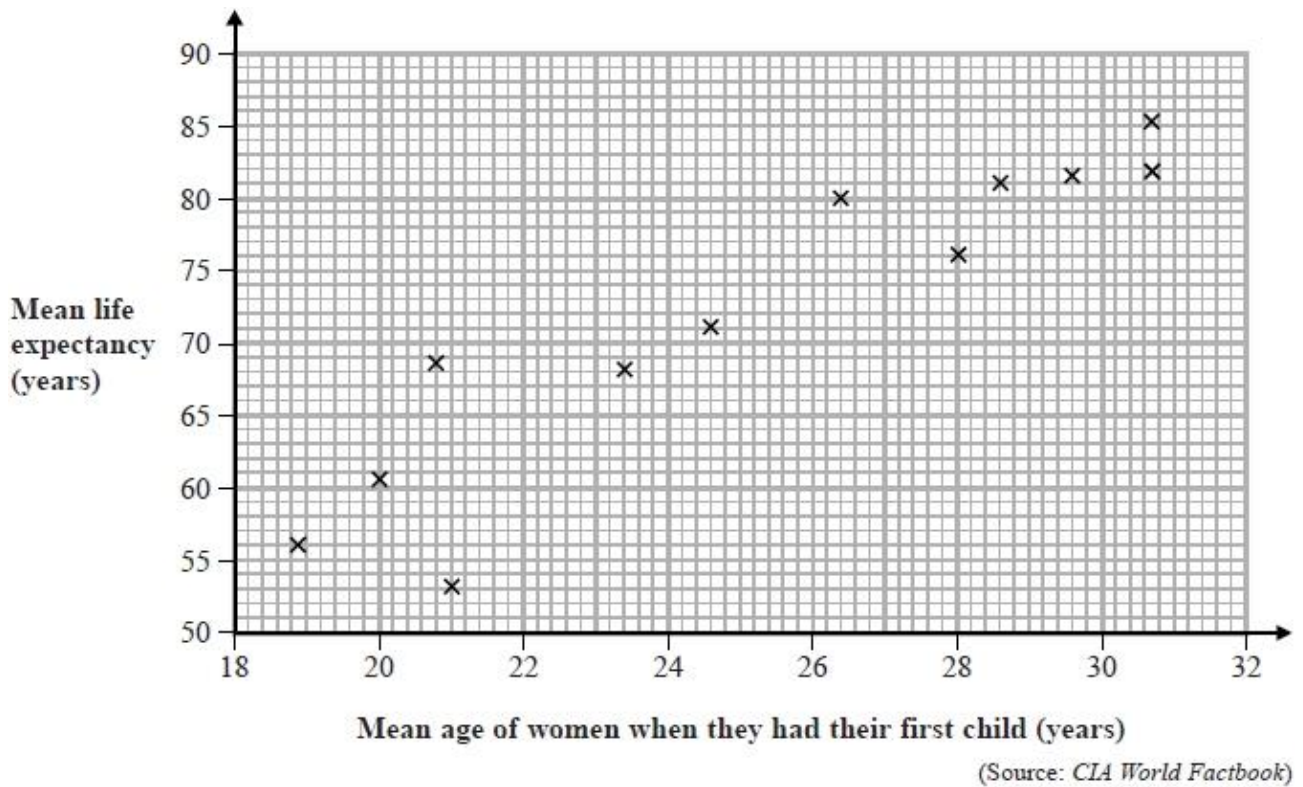
Q4.

Audrey is investigating factors that affect life expectancy by using secondary data.

For a sample of 12 countries, Audrey recorded the mean age, in years, of women when they had their first child.

She also recorded the mean life expectancy, in years, for these 12 countries.

Audrey drew a scatter diagram for this information.



(a) Explain why a scatter diagram is appropriate for the type of data Audrey collected.

.....

(1)

For these 12 countries, the double mean point of the data is (25.2, 72).

(b) Using this information, draw a line of best fit on the scatter diagram.

(2)

Audrey includes the following two statements in the conclusion to her investigation.

- A) There is strong positive correlation between the mean age at which women in a country have their first child and the mean life expectancy in that country.
- B) Life expectancy can be increased if women have their first child later in life.

(c) Discuss whether or not Audrey's conclusions are appropriate.

You should consider her graph and how she carried out her investigation.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

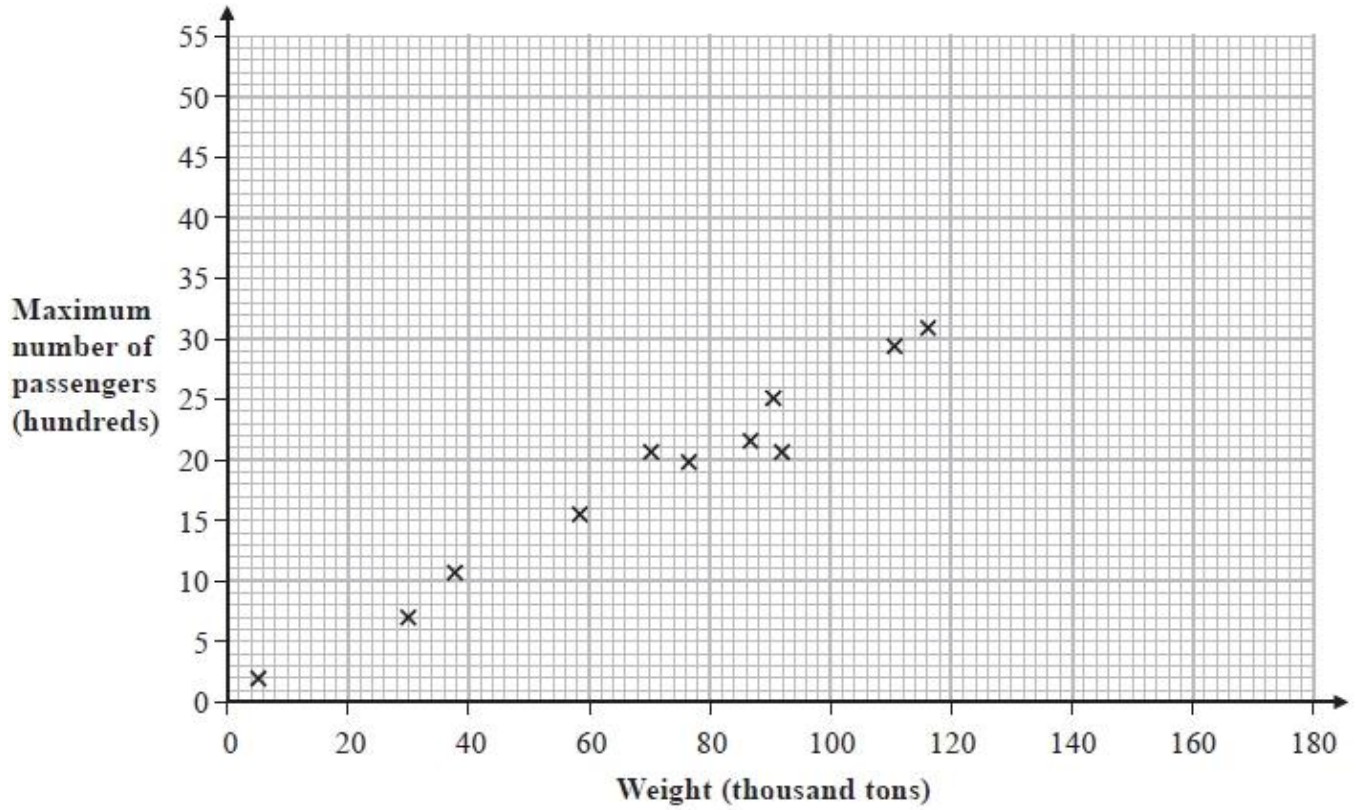
(5)

(Total for question = 8 marks)

Q5.

Naomi recorded the weight, in thousands of tons, and the maximum number of passengers, in hundreds, for 11 cruise ships.

Naomi drew the scatter diagram below for her results.



(Source: www.stat.ufl.edu)

One of the 11 cruise ships has a weight of 116 000 tons.

(a) For this ship, write down its maximum number of passengers.

..... hundred
(1)

(b) Draw a line of best fit on the scatter diagram.

(1)

(c) Describe and interpret the type of correlation shown by the scatter diagram.

.....

(3)

A new cruise ship is being built.
The ship will have a weight of 170 000 tons.

Naomi plans to use the line of best fit on her scatter diagram to predict the maximum number of passengers for the new cruise ship.

(d) Explain whether or not it is appropriate to use the line of best fit to make her prediction.

.....
.....

(2)

(Total for question = 7 marks)

Q6.

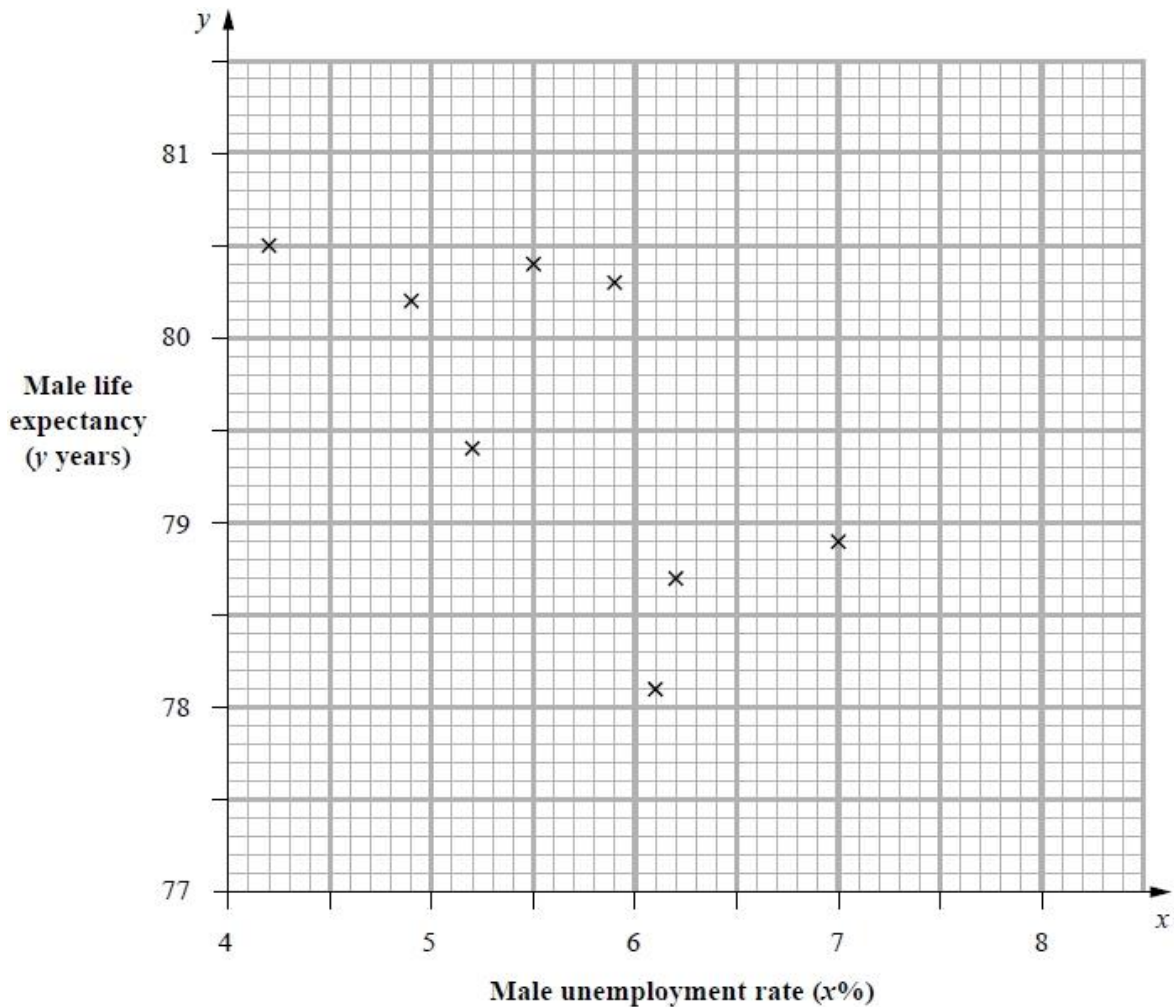
Rajesh is investigating to see if there is an association between the male unemployment rate and male life expectancy in the regions of England for 2014

His hypothesis is

"Where the male unemployment rate is high, the male life expectancy is low."

Rajesh finds the male unemployment rate ($x\%$) and the male life expectancy (y years) for each of eight of the nine regions of England for 2014

This information is shown on the scatter diagram below.



(Source: *ons.gov.uk*)

(a) Explain, giving a statistical reason, whether or not this scatter diagram supports Rajesh's hypothesis.

.....

.....

.....

The equation of the regression line for the data in the scatter diagram is $y = 83.5 - 0.7x$

(b) Draw this line on the scatter diagram.

(2)

The male unemployment rates for these eight regions have a mean of 5.6%

(c) Use this information to find the mean of the male life expectancies for these regions.

..... years

(1)

(d) Interpret the value of the gradient of this regression line.

.....
.....
.....
.....

(2)

Rajesh now finds that the male unemployment rate for the missing ninth region is 8.0%

(e) Give **two** reasons why Rajesh should not use the regression line to predict the male life expectancy in this region.

.....
.....
.....
.....
.....

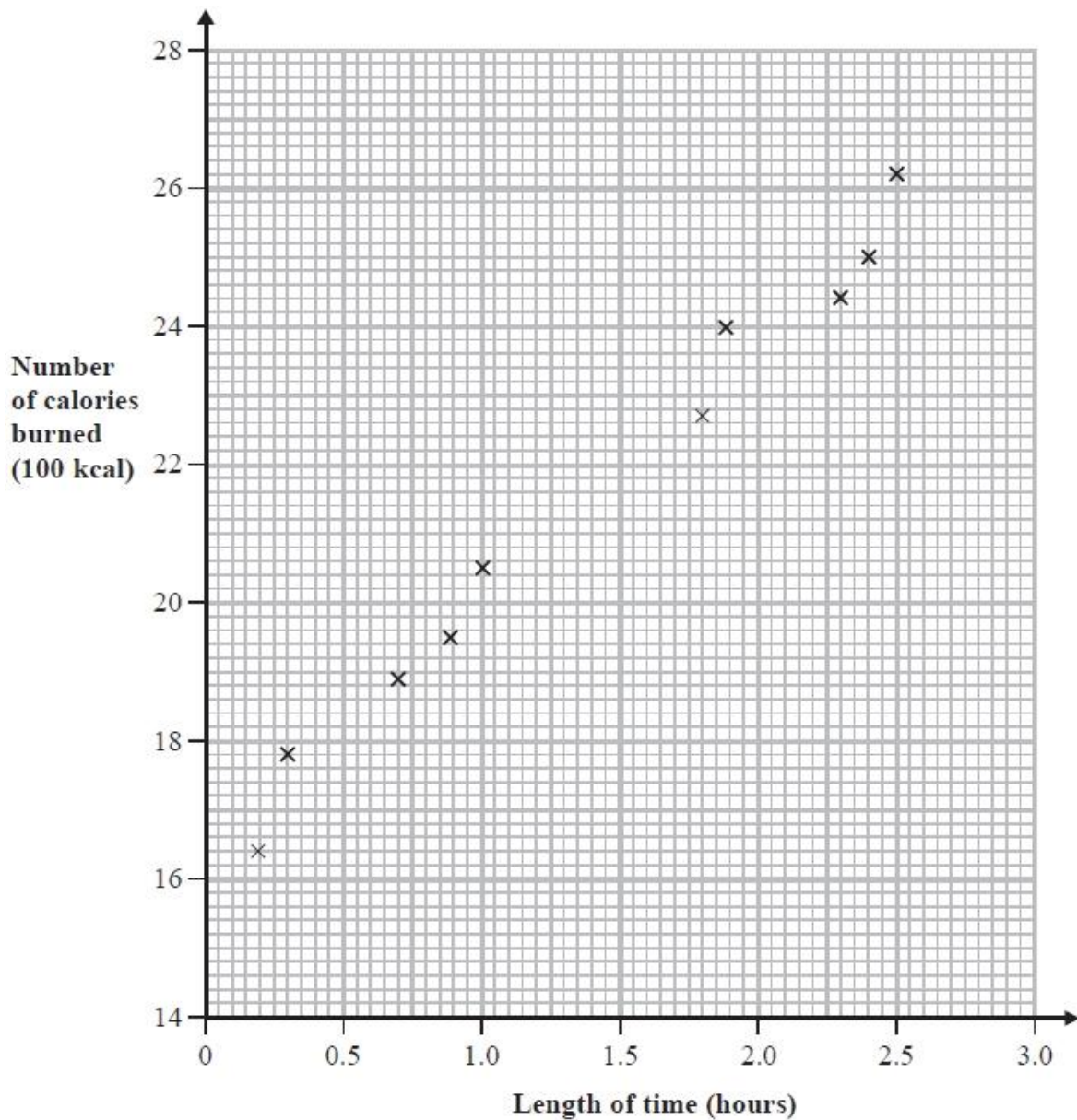
(2)

(Total for question = 9 marks)

Q7.

A sample of 10 people did an exercise for different lengths of time.

The scatter graph shows the number of calories burned, in 100 kcal, and the length of time, in hours, for each person.



The number of calories burned is the response variable.

(a) Explain what is meant by the term 'response variable'.

.....
.....

For this sample

the mean length of time is 1.4 hours,
the mean number of calories burned is 2150 kcal.

(b) On the scatter diagram

- (i) plot the mean point,
- (ii) draw a line of best fit.

(2)

(c) (i) Calculate the gradient of your line of best fit.

.....
(ii) Interpret the value of this gradient.
.....
.....

(3)

(Total for question = 6 marks)