Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel Level 1/Level 2 GCSE (9–1)

Monday 19 June 2023

Afternoon (Time: 1 hour 30 minutes)

Paper reference **1ST0/2H**

Statistics

PAPER 2

Higher Tier

You must have:
Ruler graduated in centimetres and millimetres, protractor, pair of compasses, pen, HB pencil, eraser, scientific calculator.

Total Marks

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 total mark for this paper is 80.
- 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.

Turn over ▶

P72896RA

N:1/1/1/1/

P 7 2 8 9 6 R A 0 1 2 4

Pearson

Higher Tier Formulae

You must not write on this page.

Anything you write on this page will gain NO credit.

$$\text{Skew} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

$$\text{Standard deviation} = \sqrt{\frac{1}{n}\sum(x - \bar{x})^2}$$

An alternative formula for standard deviation is

$$\text{standard deviation} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

Spearman's rank correlation coefficient

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

Rates of change (e.g. Crude birth rate = $\frac{\text{number of births} \times 1000}{\text{total population}}$)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

1 Grace asked a sample of 60 people in her town if they had ever visited France or Spain.

17 people visited both France and Spain
23 people visited Spain only
33 people visited France

(a) Draw a Venn diagram to represent this information.

(5)

Grace says

- more than half of the people in her sample have visited France
- therefore more than half of the people in her town have visited France

(b) Discuss the validity of each of Grace's comments.

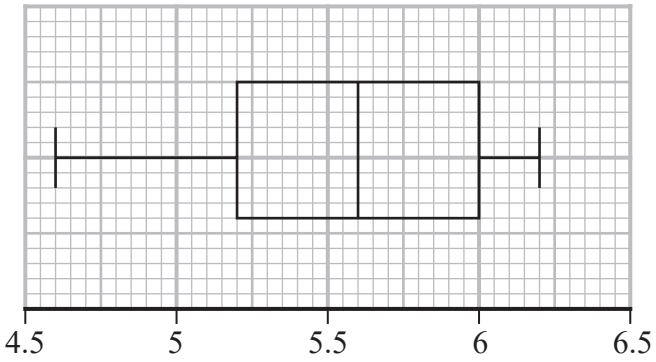
(3)

(Total for Question 1 is 8 marks)

2 Logan is investigating the heights of male adult giraffes and the heights of female adult giraffes.

He records the height, in metres, of each of a sample of male adult giraffes and the height, in metres, of each of a sample of female adult giraffes.

He draws the box plot below for the recorded heights of the male adult giraffes.



Male adult giraffe height (metres)

The table gives information about the recorded heights of the female adult giraffes.

Summary statistic	Mean	Median	Minimum	Maximum	Lower quartile	Upper quartile
Height (metres)	4.8	4.9	3.9	5.9	4.2	5.4

Logan makes the following two conclusions.

1. Male adult giraffes are generally taller than female adult giraffes.
2. The heights of the female adult giraffes are more consistent than the heights of the male adult giraffes.

Assess Logan's two conclusions.
You should show clearly the values of any statistics you use in your answer.

.....

.....

.....

.....

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

.....

.....

.....

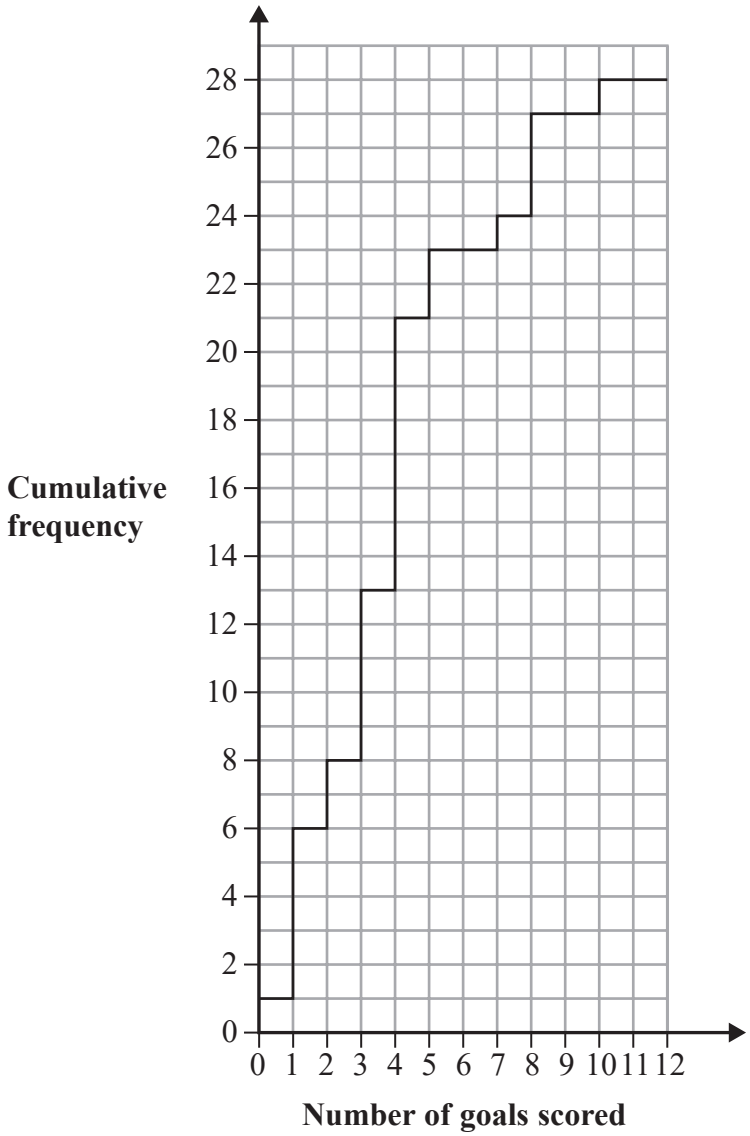
.....

.....

.....

(Total for Question 2 is 5 marks)

3 The cumulative frequency step polygon shows information about the number of goals scored in each of 28 matches played by the German women's national football team.



(Source: *www.worldfootball.net/teams/deutschland-frauen-team/*)

(a) Give a reason why a cumulative frequency step polygon is used to represent this information rather than a cumulative frequency curve.

.....

.....

(1)

(b) Find the mode of the number of goals scored.

.....

(1)

(c) Find the number of these matches where

(i) exactly 6 goals were scored,

.....

(1)

(ii) more than 6 goals were scored.

.....

(2)

In 24 matches fewer than n goals were scored.

(d) Find the value of n

.....

(1)

Klara tries to calculate the interquartile range of the number of goals scored.

She gets an answer of 14

(e) Explain how you know that her answer is incorrect.

.....

.....

(1)

(Total for Question 3 is 7 marks)

4 Maya works in a factory that produces hammers.
Today she is going to take a sample of the hammers produced and check them for defects.

(a) Which one of these statements best describes the population for Maya's sample?

This question must be answered with a cross in a box ☒. If you change your mind about an answer, put a line through the box ☒ and then mark your new answer with a cross ☒.

- ☒ 50 employees at the factory
- ☒ 100 hammers produced today
- ☒ all of the employees at the factory
- ☒ all of the hammers produced today

(1)

Maya wants to avoid bias in her sample.
She obtains a list of the unique identification numbers for each hammer.
Maya is deciding between using quota sampling and systematic sampling.

(b) For each sampling method, assess whether or not it would be an appropriate method to use for Maya's sample.

(i) quota sampling

.....

.....

.....

.....

(2)

(ii) systematic sampling

.....

.....

.....

.....

(2)

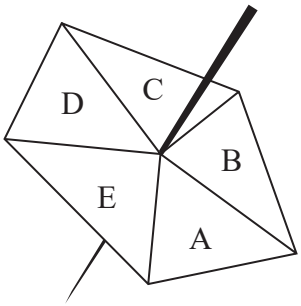
(Total for Question 4 is 5 marks)

DO NOT WRITE IN THIS AREA

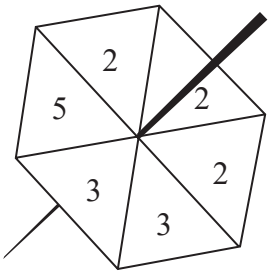
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

5 Roslyn makes two spinners.



Spinner X



Spinner Y

To test if each spinner is fair, she spins each spinner 60 times and records the side it lands on.

The tables below show her results.

Spinner X	
Outcome	**Frequency**
A	11
B	12
C	10
D	13
E	14

Spinner Y	
Outcome	**Frequency**
2	19
3	20
5	21

Roslyn believes one of the spinners is biased.

Discuss Roslyn's belief.
You should comment on the outcomes for both spinners in your discussion.

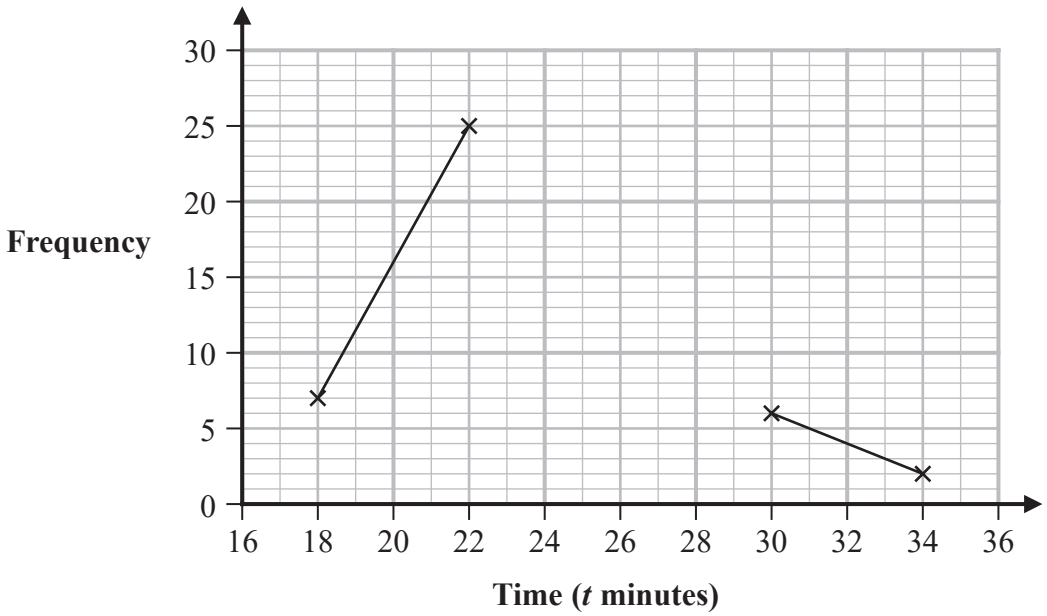
(Total for Question 5 is 4 marks)

6 The table shows information about the times, in minutes, taken for 52 people to complete a 5 km race.

Time (t minutes)	Frequency
$16 < t \leqslant 20$	7
$20 < t \leqslant 24$	25
$24 < t \leqslant 28$	12
$28 < t \leqslant 32$	6
$32 < t \leqslant 36$	2

(Source: *www.parkrun.org.uk/southwark/results*)

The incomplete frequency polygon below is drawn to represent the information in the table.



(a) Work out the number of people who completed the race in 24 minutes or less.

.....

(1)

(b) Complete the frequency polygon.

(2)

One of the people who completed the race is selected at random.

(c) Find the probability that this person completed the race in more than 20 minutes.

.....

(1)

(d) Explain how the shape of this distribution of times can be used to compare the mean time with the median time.

.....

.....

.....

(2)

The 1st decile for these results is 19 minutes and the 9th decile for these results is 30 minutes.

(e) Work out an estimate for the 1st decile to 9th decile interdecile range.

..... minutes

(1)

(f) Find the class interval that contains the 65th percentile.

.....

(1)

(Total for Question 6 is 8 marks)

7 The table shows information about the death rate in England and Wales in 2000

Year	Total population	Number of deaths
2000	52 140 181	537 877

(Source: *www.ons.gov.uk*)

(a) Using the formula below, calculate the crude death rate for England and Wales in 2000

$$\text{crude death rate} = \frac{\text{number of deaths} \times 1000}{\text{total population}}$$

.....

(2)

In 2000, the crude birth rate in England and Wales was 11.6
In 2019, the crude birth rate in England and Wales was 10.8

A newspaper prints the following headline for an article.

> 36 000 more births in England and Wales in 2019 than in 2000

(b) Discuss whether or not this headline could be true given that the crude birth rate in 2019 is less than the crude birth rate in 2000

.....

.....

.....

.....

(2)

(Total for Question 7 is 4 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

8 A scientist is carrying out an experiment to test the effect of a vitamin on memory.
She plans to use matched pairs.

(a) Explain what is meant by matched pairs in an experiment.

.....

.....

.....

.....

(2)

Jamal is one of the participants in the experiment.

He takes four tests as part of the experiment.
Each test is given a different weighting.

The table below shows the weighting and the score for each test.

Test	Weighting	Score
A	3	43
B	4	25
C	5	19
D	8	24

(b) Calculate the weighted mean score for Jamal's four tests.

.....

(3)

(Total for Question 8 is 5 marks)

9 In a survey, children were asked to state their main source of news.
They were also asked how interested they were in the news.

The table shows the results of the survey for those stating Newspapers as their main source of news.

Level of interest	Number of children
Very	60
Quite	73
Not very	18
Not at all	1
Total	152

(Source: *www.ofcom.org.uk/research-and-data/data/statistics/stats20*)

For those who said Radio is their main source of news, 65 responded 'Very' as their level of interest.

Comparative pie charts are drawn to represent the responses for Newspapers and for Radio.

(a) Compare the area of the sector for 'Very' in the pie chart for Newspapers with the area of the sector for 'Very' in the pie chart for Radio.
Give a reason for your answer.

.....

.....

.....

.....

(2)

The radius for the pie chart for Newspapers is 3 cm.
The radius for the pie chart for Radio is 4.7 cm.

(b) Calculate the size of the angle for the sector for 'Very' in the pie chart for Radio.

.....°

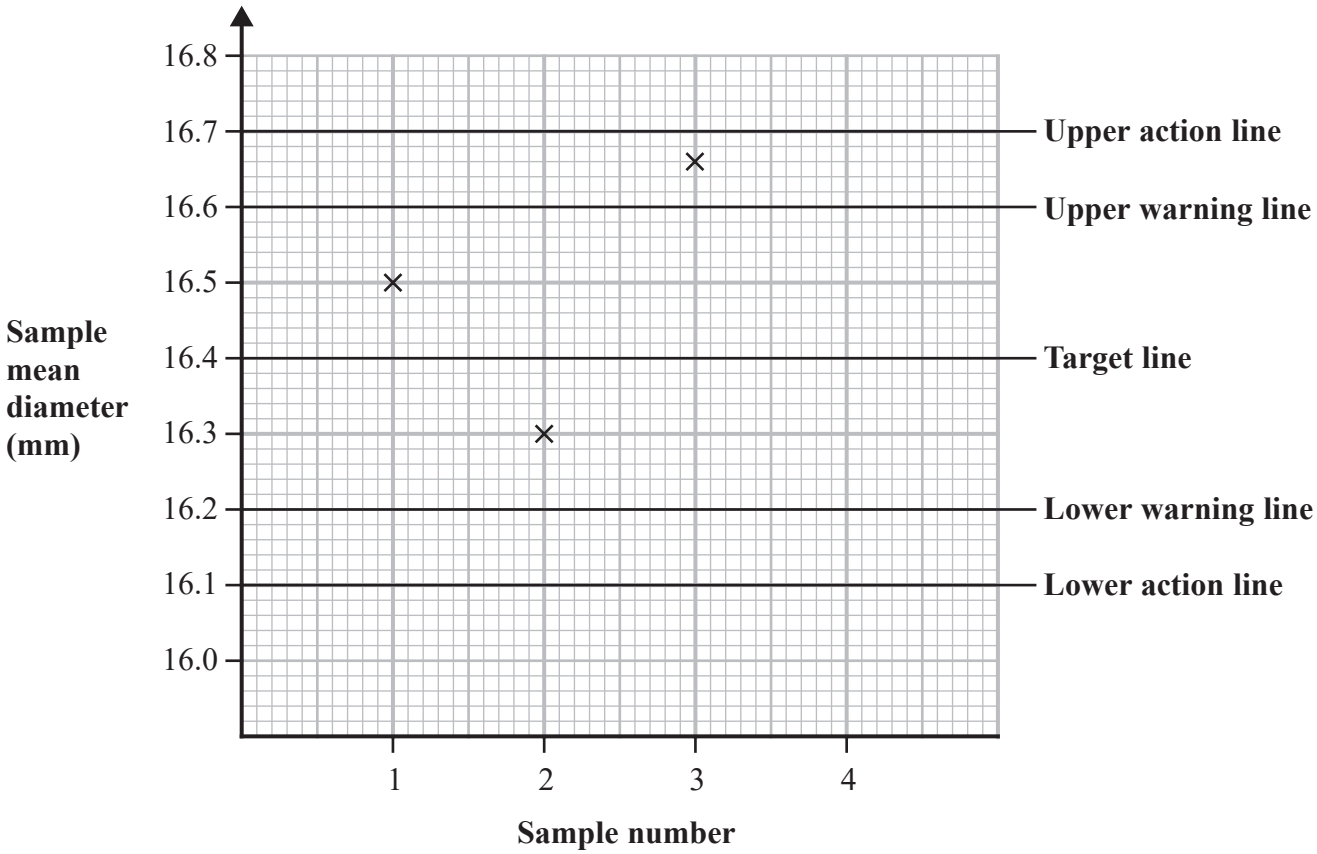
(3)

(Total for Question 9 is 5 marks)

10 A company produces bolts with a diameter 16.4 mm.

Tamika takes random samples of 4 bolts at regular intervals from the production line and the mean diameter of each sample is calculated.

Her incomplete quality assurance report is shown below.



Target sample mean 16.4 mm

Target sample standard deviation ..... mm

Sample number	Bolt 1	Bolt 2	Bolt 3	Bolt 4	Sample mean	Action
1	16.7	16.6	16.3	16.4	16.5	Take no action
2	16.4	16.5	16.1	16.2	16.3	Take no action
3	16.6	16.6	16.7	16.8	16.675	....
4	16.8	16.3	16.8	16.7	.....	....

Complete the missing information in Tamika's quality assurance report.

(Total for Question 10 is 5 marks)

11 The heights of a group of seven-year-old boys have a mean of 121.7 cm and a standard deviation of 5.3 cm.

(Source: *www.who.int/growthref/who2007_height_for_age/en/*)

Daniel is a seven-year-old boy with a standardised score for height of 0

(a) Write down Daniel's height.

..... cm

(1)

Syed and Timur are both seven-year-old boys.

Syed's standardised score for height is 1.4
Timur's standardised score for height is –1.6

Syed is taller than Timur.

(b) How much taller?

..... cm

(3)

Tarik takes a sample of 4 boys from the group of seven-year-old boys.
He wants to calculate the standardised score for the **sample mean** of their heights.
He uses 121.7 cm for the mean and 5.3 cm for the standard deviation.

(c) Discuss whether or not it is appropriate to use

(i) 121.7 cm as the mean in the calculation of the standardised score,

.....

.....

.....

.....

(2)

(ii) 5.3 cm as the standard deviation in the calculation of the standardised score.

.....

.....

.....

.....

(2)

(Total for Question 11 is 8 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

12 A bag contains red counters, blue counters and white counters only.

The table shows the proportion of each colour of counter in the bag.

Colour	Red	Blue	White
Proportion	p	$3p$	$\frac{1}{2}$

In an experiment, Salome selects one of the counters from the bag and records its colour. She then puts the counter back into the bag.

She repeats these steps 4 more times, so that she makes a total of 5 selections.

Salome says that the probability she selects exactly 1 red counter is more than the probability that she selects exactly 2 blue counters.

(a) Use binomial distributions to determine if Salome is correct.
You should show your working.

(6)

(b) By considering the conditions that make a binomial distribution a suitable model, assess the appropriateness of using a binomial distribution to model the number of red counters that Salome selects.

(3)

In a different experiment, Malena selects 1 of the counters from the bag and records its colour.

If the counter is red, she stops selecting. Otherwise she puts the counter back in the bag and selects again. She repeats this process until a red counter is selected.

Malena records the number of counters selected until she selects a red counter.

(c) Assess the use of a binomial distribution to model the number of counters selected by Malena.

(1)

(Total for Question 12 is 10 marks)

13 Indre is writing a report on the relationship between the income per person, in US dollars, and the life expectancy, in years, for different countries.

She finds data from the internet for 175 countries.

(a) Give a reason why it is important for Indre to acknowledge the source of the data she uses in her report.

.....

.....

(1)

Indre calculates Pearson's product moment correlation coefficient (PMCC) between income per person and life expectancy and she obtains a value of 0.72

(Source: *https://www.gapminder.org/data/*)

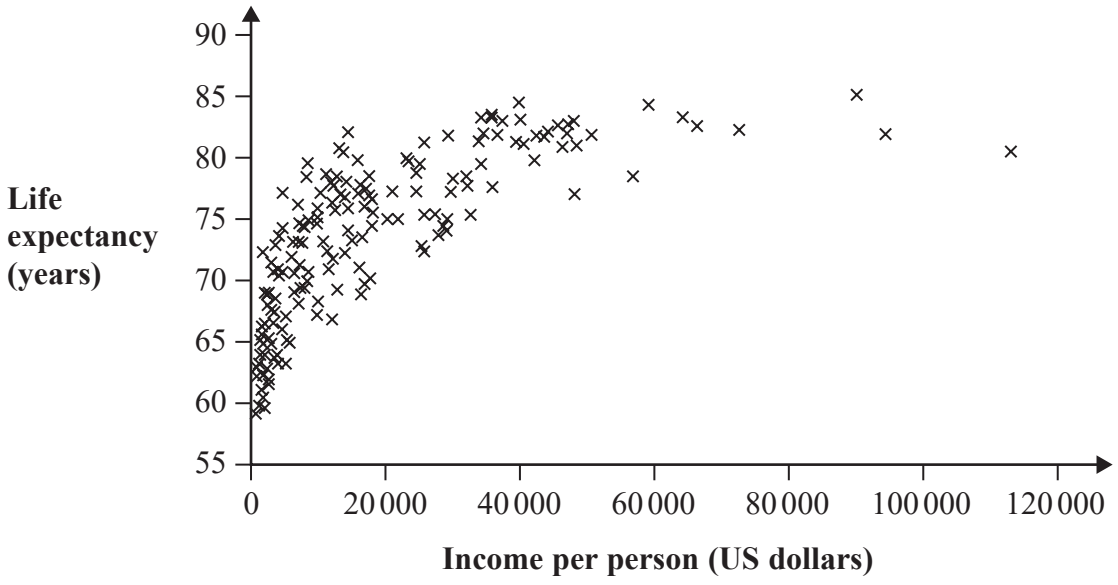
(b) Give an interpretation of this value in context.

.....

.....

(1)

Indre uses statistical software to draw a scatter diagram for the information collected.



She then uses the software to calculate the value of Spearman's rank correlation coefficient for this data.

(c) Describe how you would expect the value of Spearman's rank correlation coefficient for this data to compare with the value 0.72
Give a reason for your answer.

.....

.....

.....

.....

(2)

For the 175 countries, Indre calculates Pearson's product moment correlation coefficient between fertility rate and life expectancy and obtains a value of –0.80

(Source: *www.gapminder.org/data/*)

(d) Compare the strength of the correlation between fertility rate and life expectancy with the strength of the correlation between income per person and life expectancy.
Give a reason for your answer.

.....

.....

.....

.....

(2)

(Total for Question 13 is 6 marks)

TOTAL FOR PAPER IS 80 MARKS

BLANK PAGE

BLANK PAGE

BLANK PAGE