OCR
Oxford Cambridge and RSA

Level 3 Certificate
Quantitative Problem Solving (MEI)
H867/02 Statistical Problem Solving

Wednesday 24 May 2017 – Morning
Time allowed: 2 hours

You must have:
• the Insert (inserted)
• the Statistical Tables (ST1) (inserted)

You may use:
• a scientific or graphical calculator

First name	
Last name	

Centre number						Candidate number				

INSTRUCTIONS
• The Insert will be found inside this document.
• Use black ink. You may use an HB pencil for graphs and diagrams.
• Complete the boxes above with your name, centre number and candidate number.
• Answer **all** the questions.
• Write your answer to each question in the space provided. If additional space is required, you should use the lined page(s) at the end of this booklet. The question number(s) must be clearly shown.
• Do **not** write in the barcodes.
• You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.

INFORMATION
• The total mark for this paper is **60**.
• The marks for each question are shown in brackets **[ ]**.
• This document consists of **20** pages.
• Final answers should be given to a degree of accuracy appropriate to the context.

OCR is an exempt Charity

Turn over

www.xtrapapers.com

2

Answer **all** the questions.

Section A (30 marks)

1 A town council is trying to save money by closing its tourist information centre. It costs £45 000 a year to run it.

David is trying to estimate the cost to businesses in the town of not having it.

He identifies the following businesses that might suffer as a result.

- 35 small shops
- 10 large shops
- 16 guest houses
- 6 hotels
- 33 restaurants.

He selects a sample of 20 businesses and asks them the value of the trade, to the nearest £1000, that they think they would lose in a year if the tourist information centre was closed.

(i) (*A*) How many of each type of business should he select for his sample to be representative? **[3]**

(*B*) State which of the following terms best describes your sampling method.

Opportunity, simple random, stratified, quota, cluster, self-selected **[1]**

(ii) All of the businesses in the sample reply to David's question. This table gives a summary of their replies.

Amount (£1000)	Frequency	Amount (£1000)	Frequency
0	3	6	1
1	2	7	0
2	5	8	2
3	1	9	0
4	0	10	2
5	4	> 10	0

Use these data to estimate the total cost to businesses of closing the tourist information centre. **[3]**

(iii) It is suggested that the answer to part **(ii)** may be an overestimate. Give one reason why this might be the case. **[1]**

© OCR 2017

1 (i)(*A*)

1 (i)(*B*)

1 (ii)

1 (iii)

4

2   Simon is worried about his blood pressure. He buys an instrument to measure it; it gives two figures, the systolic and diastolic pressures. These are referred to as S and D in this question.

The table below gives Simon's figures for ten days chosen at random. There are also columns for working out Spearman's rank correlation coefficient.

(i) Complete the table and calculate Spearman's rank correlation coefficient. **[4]**

(ii) Carry out a hypothesis test at the 5% significance level of whether there is positive association between systolic and diastolic pressures. State your null and alternative hypotheses, and your conclusion. **[4]**

2 (i)

Systolic	Diastolic	S rank	D rank	d	d^2
158	95	6	2	4	16
178	88	2	4½	−2.5	6.25
174	97	3	1	2	4
162	85	5	7	−2	4
170	88	4	4½	−0.5	0.25
156	84	8			
128	87	10	6	4	16
179	93	1	3	−2	4
157	83	7			
154	79	9			
				Σ	

2 (ii)

Simon goes to see his doctor who prescribes some medicine. Simon then takes his blood pressure every day. He records his **systolic** readings for 100 days and uses a spreadsheet to draw this graph.
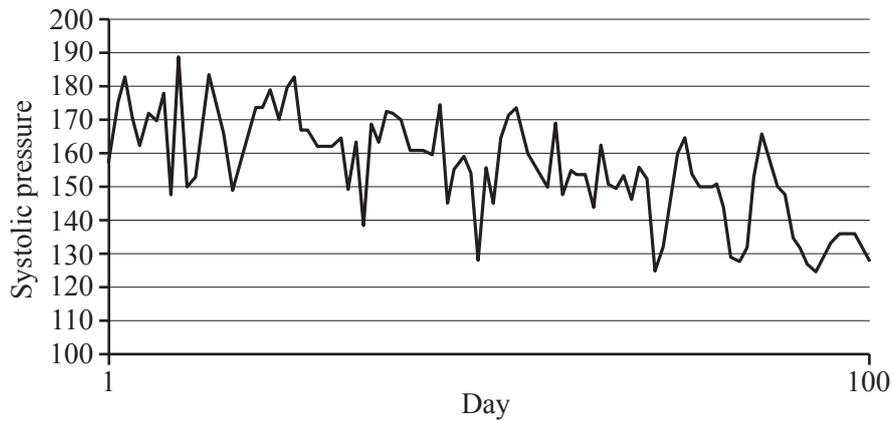
(iii) Write down two things that this graph tells Simon about his blood pressure. **[2]**

(iv) Simon shows his doctor the graph. State, with brief reasons, what the doctor can conclude from the graph about Simon's **diastolic** pressure

    (A) on any particular day,

    (B) over the whole period? **[2]**

2 (iii)	

2 (iv)(*A*)	

2 (iv)(*B*)	

Turn over

8

3   There is concern about the use of a certain type of rat poison. Some of the dead rats are eaten by birds which are then also poisoned.

A team of scientists are investigating whether this is happening in a particular area and whether it is affecting all types of birds to the same extent. The study covers three families of birds: owls, hawks and crows.

The scientists collect dead birds and analyse their livers. They classify the level of rat poison in them.

- Low (or none). The rat poison was unlikely to be the cause of death.
- Medium. The rat poison was a possible cause of death.
- High. The rat poison was almost certainly the cause of death.

At the end of one year their findings are summarised in Table 3.1.

Observed frequency, f_o	Low	Medium	High	Total
Owls	6	16	26	**48**
Hawks	10	26	12	**48**
Crows	36	12	16	**64**
Total	**52**	**54**	**54**	**160**

Table 3.1

The scientists carry out a χ^2 test using these data.

(i) Complete Table 3.2 for the expected frequencies, f_e.   **[2]**

(ii) Table 3.3 shows the contributions of the individual cells to the test statistic X^2.

Complete the empty cells and work out the value of X^2.   **[3]**

(iii) State the result of the χ^2 test at the 1% significance level.   **[3]**

3 (i)

Expected frequency, f_e	Low	Medium	High	Total
Owls	15.6	16.2	16.2	**48**
Hawks				
Crows			21.6	
Total	**52**	**54**	**54**	

Table 3.2

9

3 (ii)

	Low	Medium	High
Owls	$\dfrac{(6-15.6)^2}{15.6}=5.9077$	0.0025	5.9284
Hawks			
Crows			

Table 3.3

3 (iii)

(iv) Answer these two questions in terms of the situation being investigated.

    (A) What should the scientists conclude from the hypothesis test?

    (B) What else might the data suggest to them? **[2]**

3 (iv)(A)

3 (iv)(B)

Turn over

10

BLANK PAGE

PLEASE DO NOT WRITE ON THIS PAGE

11

Section B (30 marks)

The questions in this section are based on the pre-release data set. A hard copy of this is provided with this examination paper.

4   **(a)**   Calculate the total GDP for the Cayman Islands in millions of US$.

Give your answer to the nearest million US$.   **[3]**

   (b)   During the year in which the data were collected the death rate in Iceland was 6.20 per 1000.

Estimate the change in Iceland's population that year, giving your answer to a reasonable level of accuracy.

State one assumption that you have made.   **[4]**

4 (a)	
4 (b)	

Turn over

5 The graph in Fig. 5.1 is a Normal probability plot for the life expectancy in the 43 mainland countries in Sub-Saharan Africa for which data is available.

(i) Draw rings round the points on the graph corresponding to Chad and South Africa. **[1]**

5(i)

Fig. 5.1
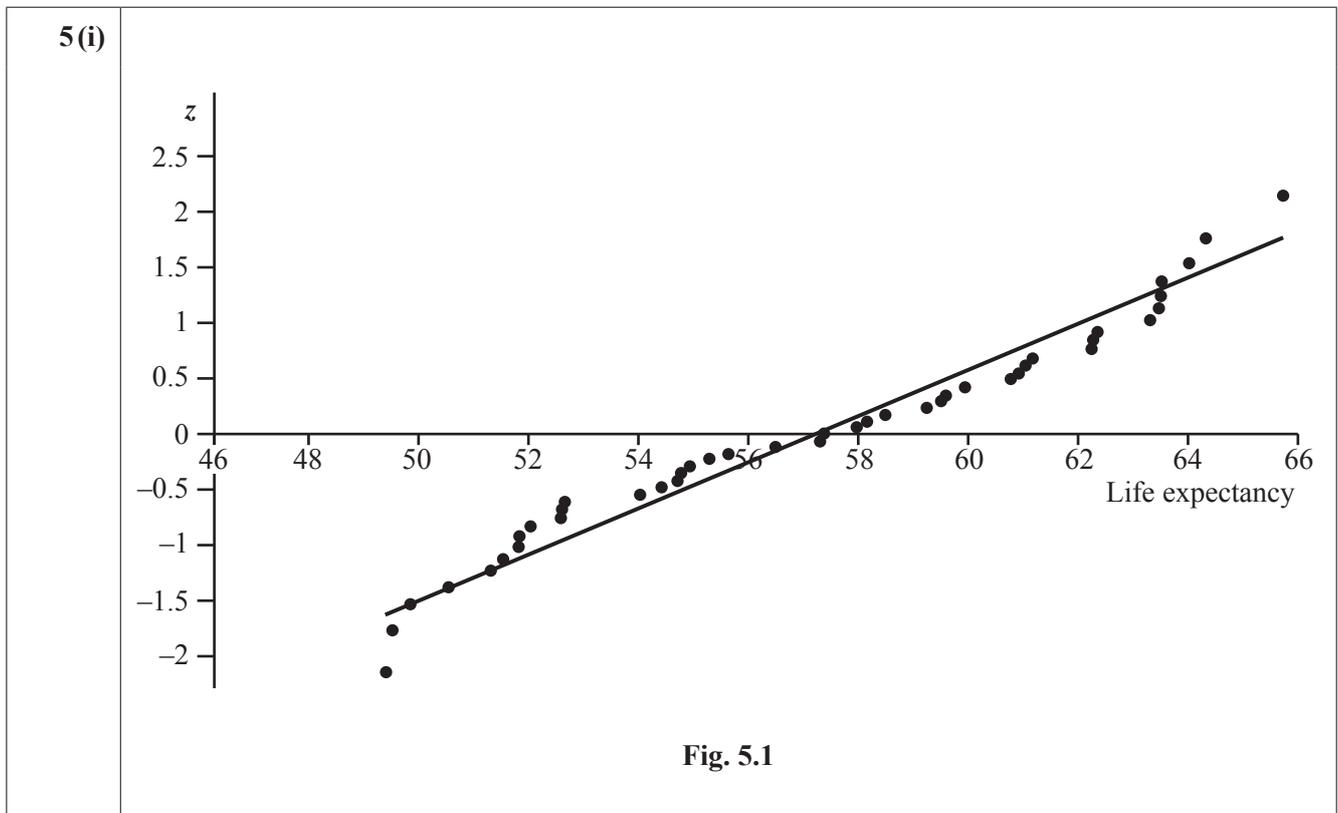
(ii) State which features of the graph indicate the following points about the distribution.

(A) The Normal distribution is a possible model.

(B) The Normal distribution is not a perfect model.

(C) The mean of the life expectancies in these countries is about 57.3 years. **[3]**

(iii) The standard deviation of these data is 4.8 years. Show that none of the values of life expectancy is more than 2 standard deviations from the mean.

State, with justification, how many values you would expect to be more than 2 standard deviations from the mean if the data followed a perfect Normal distribution. **[3]**

© OCR 2017

5 (ii)(*A*)	
5 (ii)(*B*)	
5 (ii)(*C*)	
5 (iii)	

Turn over

14

The frequency chart shows the life expectancy for the 38 mainland European countries.
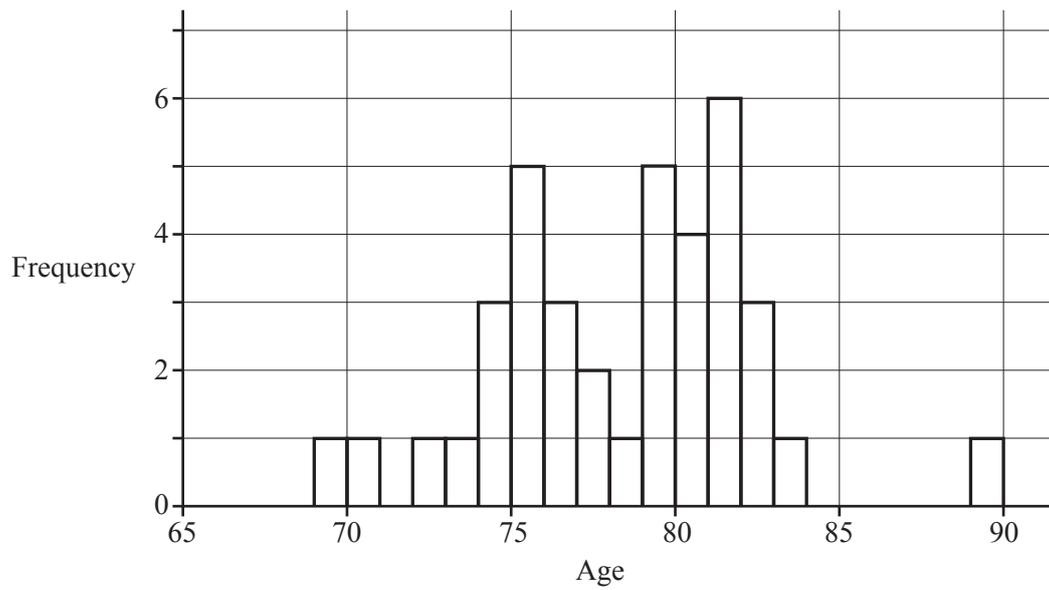
(iv) (*A*) State which one of the following terms best describes this distribution.

   Positively skewed   Bimodal   Normal

   (*B*) Which country in Europe has the highest life expectancy?   **[2]**

 (v) Write two statements comparing this distribution with that for the mainland Sub-Saharan African countries.   **[2]**

(vi) In the pre-release data set, European countries are classified as "Europe (Western)" or "Europe (Eastern)".

   Look through the figures for these two groups.

   Without doing any calculations, comment on how they influence the graph above.   **[1]**

5 (iv)(*A*)	
5 (iv)(*B*)	

5(v)

5(vi)

Turn over

16

6 Aisha is investigating access to health care for ordinary people in countries of different wealth. To model the situation she devises a variable which she calls M; it is given by the formula $M = \frac{y}{x}$ where

- x is the GDP per capita, measured in 10 000 US\$,
- y is the number of medical doctors per 1000 population.

Aisha starts with a pilot study based on the 13 members of OPEC, the Organisation of Petroleum Exporting Countries. They are given in the spreadsheet below.

(i) (*A*) Fill in the missing entries in column D.

(*B*) What formula could Aisha have used in cell D7? **[2]**

6 (i)(A)

	A	B	C	D
1	Country	x, GDP per capita (in 10 000 US\$)	y, Medical doctors (per 1000 people)	$M = \frac{y}{x}$
2	Algeria	0.75	1.21	1.61
3	Angola	0.63	0.17	
4	Ecuador	1.06	1.69	
5	Indonesia	0.52	0.20	
6	Iran	1.28	0.89	0.70
7	Iraq	0.71	0.61	0.86
8	Kuwait	4.21	1.79	0.43
9	Libya	1.13	1.90	1.68
10	Nigeria	0.28	0.40	1.43
11	Qatar	10.21	2.76	0.27
12	Saudi Arabia	3.13	0.94	0.30
13	UAR	2.99	1.93	0.65
14	Venezuela	1.36	1.94	1.43

6 (i)(B)

(ii) Give a brief statement for each of the three figures for Nigeria in the spreadsheet interpreting it in comparison with those for the other OPEC countries. **[3]**

(iii) Aisha thinks that the quantity M will be a useful measure for her study.

Give one reason why this might be the case. **[1]**

6 (ii)	**Column B**
	Column C
	Column D
6 (iii)	

Aisha decides to move from the pilot study to one covering all the countries of the world. She obtains data for the number of doctors per 1000 people for a total of 194 countries. She then uses a spreadsheet to draw the scatter diagram below.

(iv) On the scatter diagram draw, by eye, a line of best fit passing through the origin.
Write the corresponding value of M beside it.

Circle the point on the scatter diagram corresponding to Saudi Arabia. **[3]**

(v) Aisha uses the spreadsheet to calculate the product moment correlation coefficient between x and y for all 194 countries. Its value is 0.58.

Make two comments indicating how this value, together with the scatter diagram, is relevant to Aisha's investigation. **[2]**

6 (v)	

END OF QUESTION PAPER
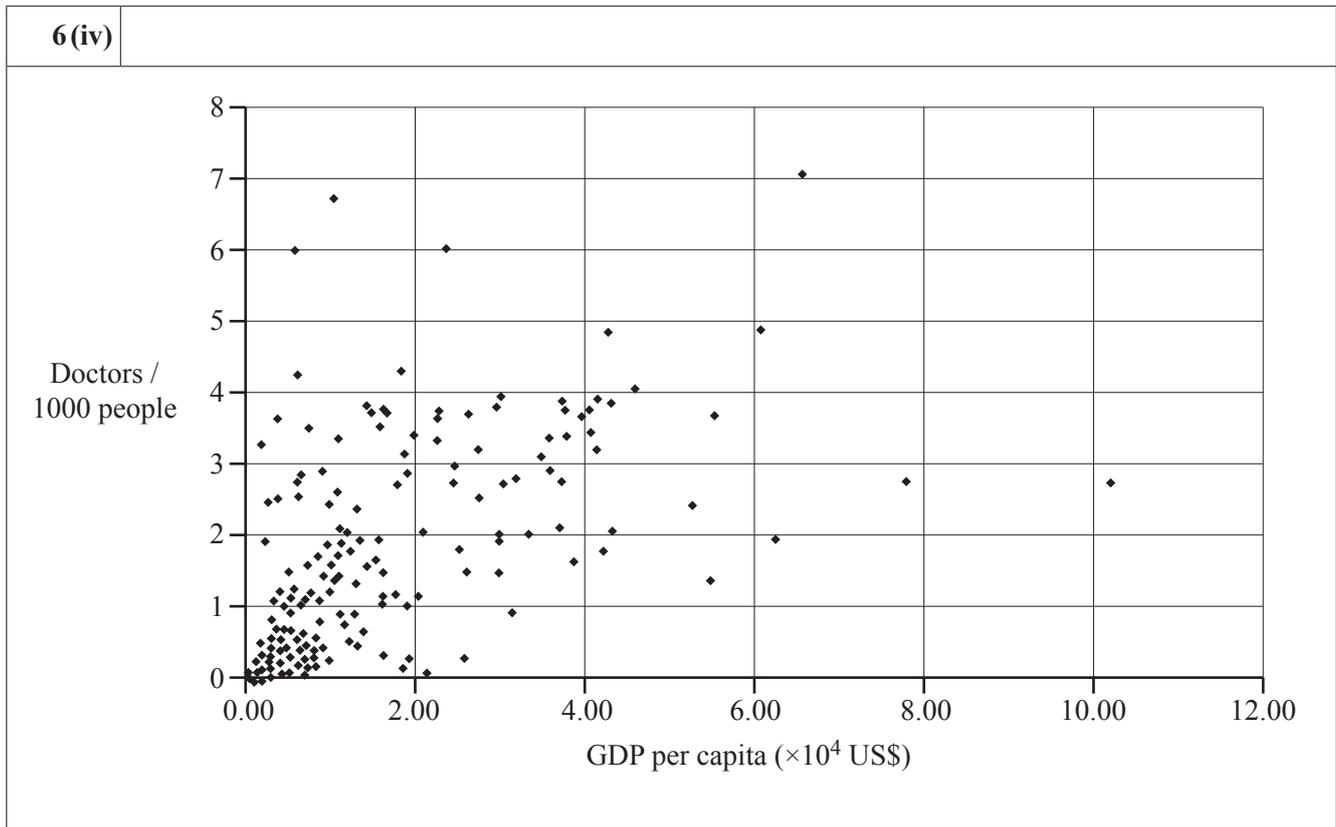
ADDITIONAL ANSWER SPACE

If additional space is required, you should use the following lined page(s). The question number(s) must be clearly shown in the margin(s).

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

OCR
Oxford Cambridge and RSA