Oxford Cambridge and RSA

Wednesday 13 October 2021 – Afternoon

Level 3 Certificate Core Maths B (MEI)

H869/02 Statistical Problem Solving

Time allowed: 2 hours

You must have:
- the Insert (inside this document)
- the Statistical Tables (ST1) (inside this document)

You can use:
- a scientific or graphical calculator

* 8 9 8 4 1 6 4 9 8 4 *

* H 8 6 9 0 2 *

Please write clearly in black ink. **Do not write in the barcodes.**

Centre number | | | | | | Candidate number | | | | |

First name(s) ______________________

Last name ______________________

INSTRUCTIONS
- Use black ink. You can use an HB pencil, but only for graphs and diagrams.
- Write your answer to each question in the space provided. If you need extra space use the lined pages at the end of this booklet. The question numbers must be clearly shown.
- Answer **all** the questions.
- Where appropriate, your answer should be supported with working.
- Give your final answers to a degree of accuracy that is appropriate to the context.

INFORMATION
- The total mark for this paper is **60**.
- The marks for each question are shown in brackets **[ ]**.
- This document has **20** pages.

ADVICE
- Read each question carefully before you start your answer.

Answer **all** the questions.

Section A (30 marks)

1 Some of the people registered with a doctors' surgery form a group.

They prepare a questionnaire (for people aged over 15) to help them find out how the surgery can be more helpful to its patients. **Fig. 1.1** shows part of the questionnaire.

About yourself Please ring as appropriate	Sex Age	 16–25	M 26–35	 36–45	F 46–55	 56–65	 66+
How many times have you seen a doctor in the last year?							
On your last visit, did you see your own doctor?							
How long did you have to wait for an appointment?							
I would prefer the doctor I see to be	Male	Female	No preference				

Fig. 1.1

The group go to the surgery on a day when people are being given a flu jab and ask those waiting in the queue to fill in the questionnaire.

Fig. 1.2 shows the group's summary of data for these questions from the completed questionnaires.

Ages	16–25	26–35	36–45	46–55	56–65	66+
Males	1	2	6	8	12	10
Females	0	3	8	10	12	22

Appointments in the last year	Males	125
	Females	167

Own doctor or any doctor	Own doctor	Any doctor
Appointments	95	197
Average waiting time (days)	9.3	1.5

Preference	Male doctor	Female doctor	No preference
Male patient	15	4	20
Female patient	6	21	28

Fig. 1.2

(a) State which of these terms best describes the group's sample

Cluster, Opportunity, Quota, Self-selected, Simple random, Stratified. **[1]**

(b) **Fig. 1.2** gives a summary of the information from the sample. State, with accompanying calculations where necessary, whether the following statements are consistent with it.

(i) The number of people waiting to see their own doctor is over 6 times as great as those seeing any doctor. **[1]**

(ii) Most people don't mind whether the doctor they see is male or female. **[2]**

(iii) On average a female patient sees a doctor more often than a male patient. **[3]**

(c) State **two** ways in which the data collected in the sample may be unrepresentative of all the patients at the surgery. **[2]**

1(a)	**Ring one of:**
	Cluster, Opportunity, Quota, Self-selected, Simple random, Stratified.
1(b)(i)	
1(b)(ii)	
1(b)(iii)	
1(c)	**1st way**
	2nd way

BLANK PAGE

PLEASE DO NOT WRITE ON THIS PAGE

2 **Fig. 2.1** shows a Normal distribution.

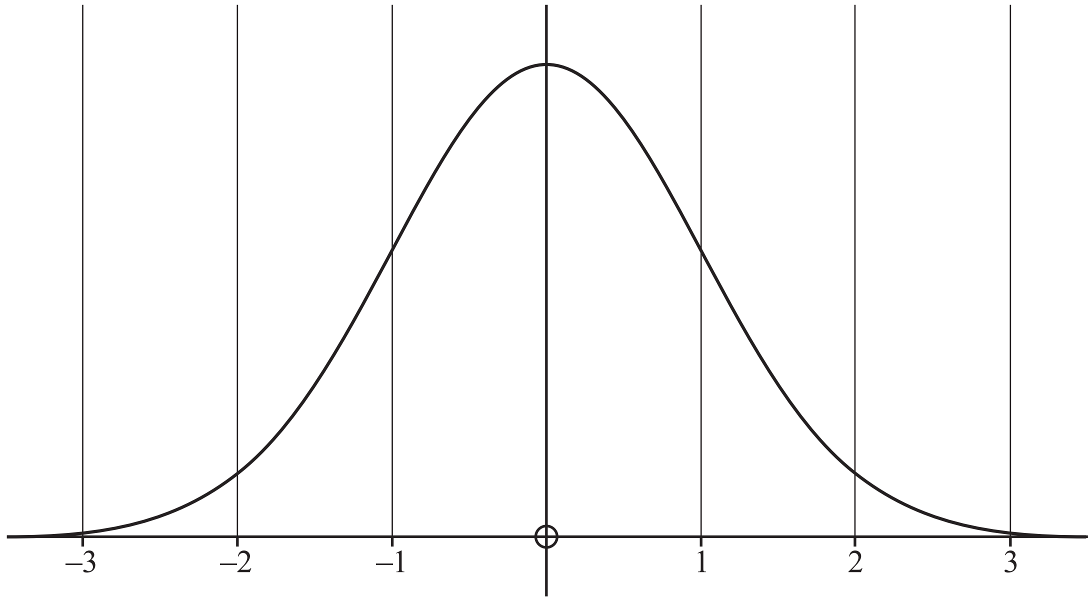

Fig. 2.1

(a) Write down the mean and standard deviation of this distribution. **[2]**

(b) Jane says
"For this Normal distribution, 16% of observations are less than –1.0."

Show that Jane is correct. **[2]**

(c) Find the percentage of observations that are less than 1.5 for this distribution. **[1]**

2(a)	**Mean**
	Standard deviation
2(b)	
2(c)	

2 continued

A lake in the north of Canada is permanently covered in ice.

A record has been kept for a very long time of the thickness of the ice on June 30th.

It was found that until recently the thickness was Normally distributed with mean 110 cm and standard deviation 2.5 cm.

(d) In what proportion of years would you expect the ice to have been less than 107.5 cm thick? **[1]**

2(d)	

The 30 most recent measurements are given in **Fig. 2.2**.

Thickness (cm)	104	105	106	107	108	109	110	111	112	113	114	115
Frequency	1	3	9	8	4	2	0	2	0	0	0	1

Fig. 2.2

(e) The data in **Fig. 2.2** have been rounded.

Write down the values of the thickness that would be given as 107 cm. **[1]**

(f) Using the mean of 110 cm and the standard deviation of 2.5 cm, state the percentage of these observations that are below the mean by more than 1 standard deviation. **[1]**

(g) Ben says "The data in **Fig. 2.2** prove that global warming is taking place". Jane disagrees.

Give **two** statements, one supporting each point of view. **[2]**

(h) It is suggested that the figure of 115 must be a mistake and should be ignored.

What effect would that have on the mean of the data in **Fig. 2.2**? **[1]**

2 (e)	
2(f)	
2(g)	**In favour of Ben's statement**
	Against Ben's statement
2(h)	

3 Here are two articles from a local newspaper.

> **2016** **Bringing life back to our rivers**
>
> After well over a century when no fish could possibly live in them, many of our rivers have now been cleaned up. In an ambitious programme, scientists have buried salmon eggs at secret locations in the gravel of three rivers. They hope that some of the eggs will hatch and the young fish will swim down into the Atlantic, to return as adult fish to breed in the same rivers.

> **2021** **The salmon return**
>
> Five years ago we reported on a brave experiment to return salmon to our rivers. It has worked! Adult salmon are now often seen in all three rivers.
> But no fishing allowed!

To compare the three rivers the scientists carried out annual observations in all three rivers, counting the numbers of one-year olds, two-year olds and returned adults in a small stretch of each river. Their results are given in **Fig. 3.1**.

Observed frequency, f_o	**One-year olds (2016)**	**Two-year olds (2017)**	**Adults (2021)**	**Total**
River A	190	94	16	**300**
River B	165	85	14	**264**
River C	195	36	5	**236**
Total	**550**	**215**	**35**	**800**

Fig. 3.1

The scientists use the data in **Fig. 3.1** to carry out a chi-squared test.

(a) State which one of these is a correct statement of the null hypothesis.

A The proportions of one-year old, two-year old and adult salmon are the same for each river.

B The three rivers have the same numbers of salmon.

C There are more predators in some rivers than others.

D The level of pollution is the same in all the rivers. **[1]**

(b) Complete **Fig. 3.2** in the answer space, giving the Expected values in the cells. **[2]**

(c)
- Show how the number 1.5 in **Fig. 3.3** was calculated.
- Calculate the value of X^2.
- Carry out the test at the 5% significance level and state the conclusion. **[6]**

(d) Based on the data, make a realistic conjecture about the situation. **[1]**

3 (a)

3 (b)

Expected frequency, f_e	**One-year olds**	**Two-year olds**	**Adults**	**Total**
River A	206.25	80.625	13.125	**300**
River B	181.5	70.95	11.55	**264**
River C	162.25			**236**
Total	**550**			**800**

Fig. 3.2

3 (c)

$$\sum \frac{(f_o - f_e)^2}{f_e} = 1.2803\ldots + 2.2187\ldots + 0.6297\ldots$$
$$+ 1.5 + 2.7822\ldots + 0.5196\ldots$$
$$+ 6.6105\ldots + 11.8585\ldots + 2.7463\ldots$$

Fig. 3.3

3(d)

BLANK PAGE

PLEASE DO NOT WRITE ON THIS PAGE

Answer **all** the questions.

Section B (30 marks)

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

4 **(a)** Write down the population of Malawi and its total road length. **[1]**

(b) **(i)** Imagine that everyone in Malawi stands in a line on a road and holds hands with the people on either side, as shown in **Fig. 4.1**.

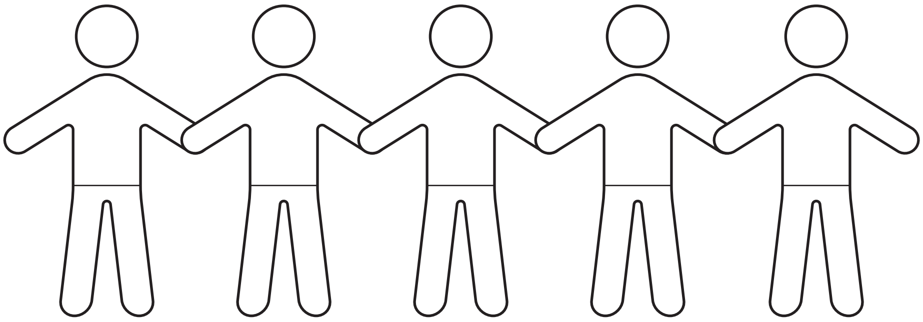

Fig. 4.1

Show that there are enough people to cover all the roads in the country. **[3]**

(ii) Investigate whether the same is true for Canada. **[1]**

4(a)	**Population**
	Total road length
4(b)(i)	
4(b)(ii)	

5 The charts in **Fig. 5.1** and **Fig. 5.2** illustrate the population percentages for different age groups in two imaginary countries.

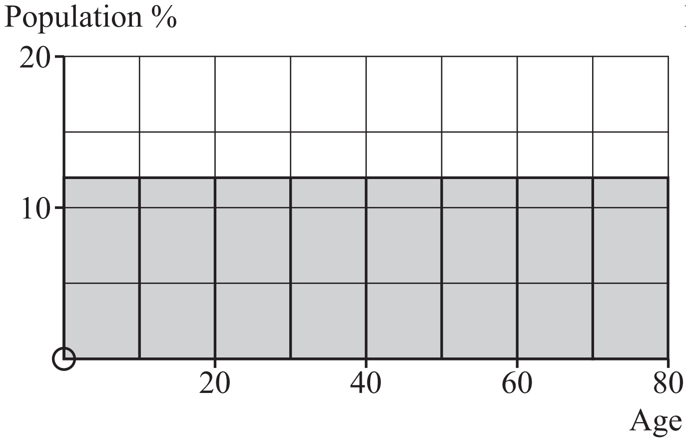

Fig. 5.1

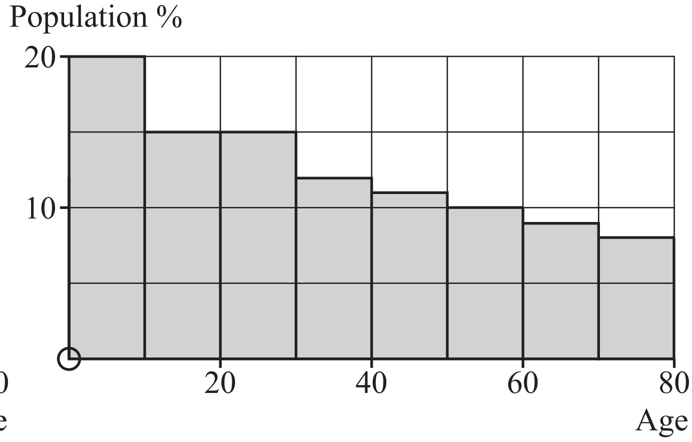

Fig. 5.2

(a) **(i)** Describe the trend of the population size in each country. Is it increasing, decreasing or stable? **[1]**

(ii) Estimate the median age for each country. **[2]**

(b) Show that in the UK the median age is almost exactly half of the life expectancy. **[1]**

5(a)(i)	**Trend Fig. 5.1**	**Ring one answer**	Increasing	Decreasing	Stable
	Fig. 5.2	**Ring one answer**	Increasing	Decreasing	Stable
5(a)(ii)	**Median Fig. 5.1**				
	Median Fig. 5.2				
5(b)					

After looking at the data, Jamal forms the following conjecture.

> In a country where the population has median age M years and life expectancy L years,
>
> - If $2M = L$, the population is stable, and in this case $\frac{2M}{L} = 1$,
> - If $2M < L$, the population is increasing, and in this case $\frac{2M}{L} < 1$,
> - If $2M > L$, the population is decreasing, and in this case $\frac{2M}{L} > 1$.

To investigate his conjecture, Jamal generates a random sample of size 8 from the pre-release data set. He then uses this sample to carry out a Spearman's Rank Correlation test at the 5% significance level for the variables $\frac{2M}{L}$ and the Annual Population Growth Rate.

(c) One of the countries in Jamal's sample is Japan.

Show that the value of $\frac{2M}{L}$ for Japan is 1.11 to **2** decimal places. **[2]**

(d) State the null hypothesis and the alternative hypothesis for Jamal's test. **[1]**

(e) Complete the table in **Fig. 5.3** and hence carry out the test, stating your conclusion. **[5]**

5(c)	
5(d)	**Null hypothesis**
	Alternative hypothesis

5(e)

Country	$\frac{2M}{L}$	$\frac{2M}{L}$ rank x	Growth rate	Growth rate rank y	$d = x - y$	d^2
Zambia	0.64	7	2.93	1	6	36
Canada	1.03	2	0.73	5	–3	9
Venezuela	0.74	6	1.24	3	3	9
Iraq	0.53	8	2.55	2	6	36
Japan	1.11		–0.21			
France	1.01	3	0.39	7	–4	16
Australia	0.94	4	1.03	4	0	0
Fiji	0.79	5	0.60	6	–1	1
					$\sum d^2 =$	

Fig. 5.3

To further his investigation, Jamal uses the spreadsheet to draw a scatter diagram of Annual Population Growth Rate against $\frac{2M}{L}$. This is shown below in **Fig. 5.4**.

The Cook Islands are in the South Pacific near New Zealand. Many people from the Cook Islands now live in New Zealand or in Australia.

(f) **(i)** Ring the point on the scatter diagram corresponding to the Cook Islands. **[1]**

(ii) Give an explanation for why this country should be an outlier. **[1]**

5(f)(i)

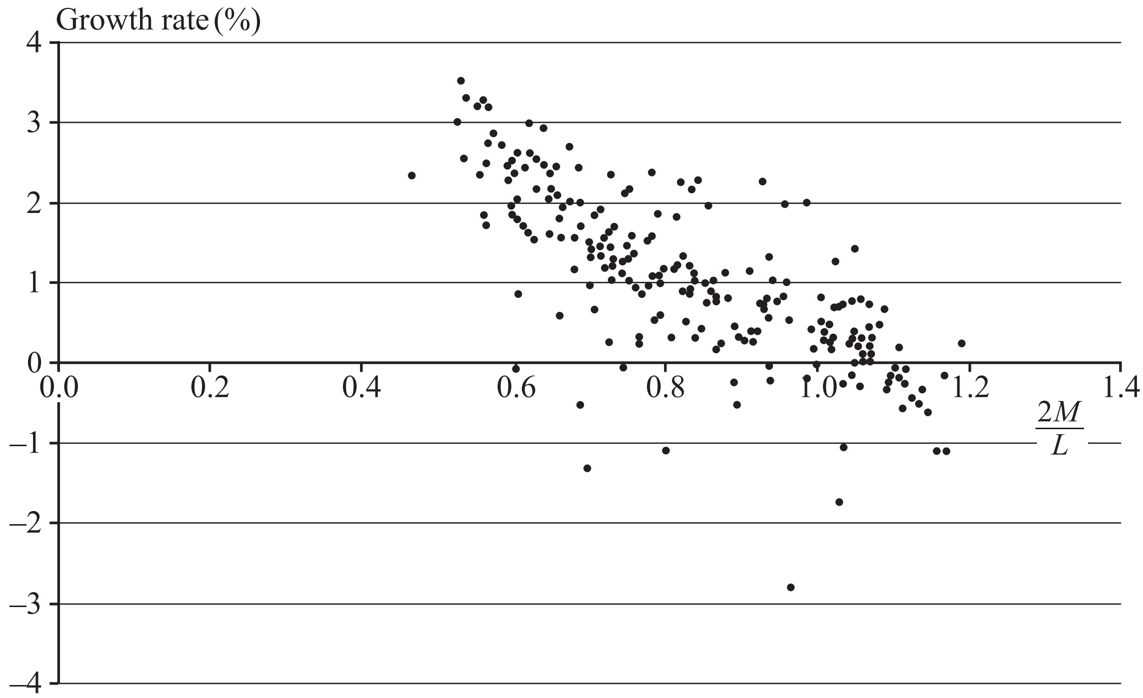

Fig. 5.4

5(f)(ii)

6 **(a)** Show that the population density of Algeria is 17.2 people per square kilometre. **[1]**

Anika is interested in how the population density varies around the world.

She uses the spreadsheet of pre-release data to give the figures for every country except the Holy See.

She starts by entering a suitable formula into cell P2 to calculate the population density. Then she copies it down the column.

(b) **(i)** Write down the formula Anika enters in cell P2. **[1]**

(ii) Suggest one reason why she omits the Holy See. **[1]**

Anika ranks the data and lists the 6 countries with the highest population density and the 6 with the lowest. They are given in **Fig. 6.1**. She has rounded the figures for population density.

Rank	**Country**	**Population density**	**Rank**	**Country**	**Population density**
1	Macau	21500	230	Australia	3.00
2	Monaco	15300	231	Western Sahara	2.27
3	Singapore	8450	232	Mongolia	1.96
4	Hong Kong	6490	233	Pitcairn	1.15
5	Gaza Strip	4990	234	Falklands	0.241
6	Gibraltar	4200	235	Greenland	0.0266

Fig. 6.1

(c) To what accuracy has Anika given the figures for population density? **[1]**

(d) Use your general knowledge to suggest one feature of the countries with high population density and one feature of some of those with low population density. **[2]**

6(a)	
6(b)(i)	
6(b)(ii)	
6(c)	
6(d)	**High population density**
	Low population density

6 continued

Anika thinks there may be a negative correlation between land area and population density. She uses the 12 countries given in **Fig. 6.1** as a sample to calculate the value of Spearman's Rank Correlation Coefficient, r_S.

Her answer is –0.608 and she thinks this is large enough for it to be worth further investigation using the whole population.

So she uses the spreadsheet to calculate the Product Moment Correlation Coefficient between Population Density and Land Area for every country except the Holy See, and also to show the data in a scatter diagram including a line of best fit.

The Product Moment Correlation Coefficient is –0.0646.

The scatter diagram is shown in **Fig. 6.2** below.

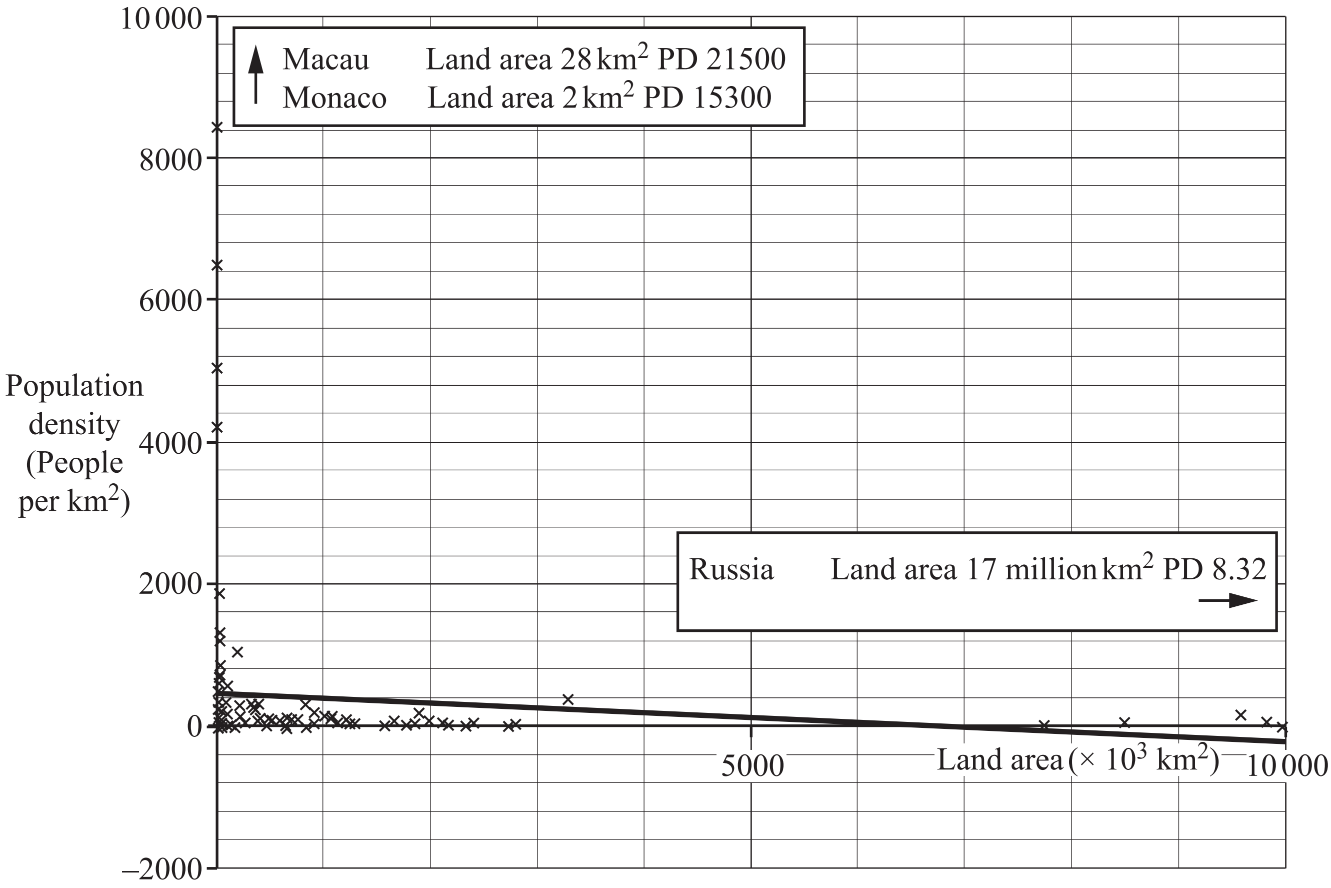

Fig. 6.2

(e) What does the value of the product moment correlation coefficient tell you about the data? **[1]**

(f) **Fig. 6.1** lists two groups of countries: those with population density ranked 1 to 6 and those ranked 230 to 235.

Where are these two groups of countries located in the scatter diagram in **Fig. 6.2**? **[2]**

(g) Anika says "The scatter diagram shows that a straight line of best fit is **not** suitable for these data."

Give **two** features of **Fig. 6.2** which support Anika's statement. **[2]**

6(e)	
6(f)	**Ranks 1 to 6**
	Ranks 230 to 235
6(g)	**Feature 1**
	Feature 2

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).