OCR
Oxford Cambridge and RSA

Thursday 23 May 2024 – Afternoon

Level 3 Certificate Core Maths B (MEI)

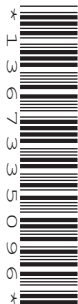**H869/02** Statistical Problem Solving

Time allowed: 2 hours

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

You can use:
- a scientific or graphical calculator

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 **24** pages.

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

Turn over

Section A

1 A town's council announce that they are planning to install speed bumps on Oak Road which runs in front of Oak School.

Not everyone supports this proposal and the council receive several letters of objection.

(a) Give **one** reason in favour of the speed bumps. **[1]**

1(a)	**Reason in favour**

Before going ahead with the speed bumps, the council post the questionnaire in **Fig. 1.1** on their website and invite the town's residents to respond.

Fig. 1.1

1	How often do you drive along Oak Road? (A) Every school day (B) Several times a month (C) Seldom or never
2	How many children do you have at Oak School? (A) None (B) 1 (C) 2 or more
3	Do you support the plan for speed bumps on Oak Road? (A) Yes (B) No view (C) No
4	What is your age? (A) under 21 (B) 21–45 (C) Over 45

Fig. 1.2 shows the answers from the first 8 responses the council receive. The responses are labelled L to S.

Fig. 1.2

Question	Response							
	L	**M**	**N**	**O**	**P**	**Q**	**R**	**S**
1	A	A	B	A	A	A	B	C
2	A	C	A	A	C	C	A	B
3	C	A	C	C	A	A	C	A
4	C	B	C	A	B	C	A	B

A member of the council looks at these 8 early responses and reports to the others on the outcomes so far.

(b) (i) How would you best describe the sample?
Circle **one** of the options in the answer space. **[1]**

1(b)(i)	**Opportunity**	**Self-selected**	**Cluster**
	Simple random	**Quota**	**Stratified**

(ii) What do the responses to question 1 in **Fig. 1.2** suggest? **[1]**

(iii) Look at the responses to questions 2 and 3 in **Fig. 1.2** together. What do they suggest? **[2]**

1(b)(ii)	
1(b)(iii)	

Turn over

When all the responses have come in, the council publish the results as **Fig. 1.3**.

Fig. 1.3

	Responses		
Question	**A**	**B**	**C**
1	95	34	4
2	52	26	55
3	70	3	60
4	9	72	31

(c) **(i)** Which question did some respondents definitely **not** answer? How do you know this? Circle **one** of the options in the answer space. **[2]**

(ii) Suggest a reason for some people **not** answering that particular question. **[1]**

(iii) Given your answer to part **(b)(iii)**, explain briefly why **Fig. 1.3** is **not** an informative way to show the results. **[1]**

1(c)(i)	Circle the question some respondents did **not** answer.
	1 2 3 4
	How you know this
1(c)(ii)	
1(c)(iii)	

BLANK PAGE

PLEASE DO NOT WRITE ON THIS PAGE

Turn over for the next question

2 Parkruns are held every Saturday morning in many places around the country. Each Parkrun is 5 km long and is free to enter.

A group of 8 people want to set up a Parkrun in their town. They have a suitable course but need to get agreement from the national organisers.

To help with this, the group run the course. Their times are recorded and given in cells B1 to B8 of the spreadsheet shown below.

	A	B
1	1st	17.25
2	2nd	18.75
3	3rd	20.00
4	4th=	22.15
5	4th=	22.15
6	6th	27.22
7	7th	30.55
8	8th	33.95
9		
10	Mean	24.00
11	SD	5.96

The times are given in minutes, as decimals, to the nearest 0.01 minutes. They are **not** given in minutes and seconds.

Thus in the entry 17.25, the .25 means $\frac{25}{100}$ of a minute or

$\frac{25}{100} \times 60 = 15$ seconds.

So the time of 17.25 means 17 minutes 15 seconds.

The figures in cells B10 and B11 are the mean and standard deviation of the figures in cells B1 to B8.

(a) Write the time given in cell B8 in minutes and seconds. **[1]**

(b) Complete the blank spaces in the formula =STDEV(.... : ....) used to find the number in cell B11. **[1]**

2(a)	

2(b)	$= \text{STDEV}(\, ... \;:\; .... \,)$

(c) The leader of the group says "Our data show that when our Parkrun happens and a lot of people do it, their times will be Normally distributed with mean 24 minutes and standard deviation 6 minutes."

Give **two** criticisms of the leader's statement. **[2]**

2(c)	**Criticism 1**
	Criticism 2

The Parkrun is given permission to go ahead. On the first day 400 people participate.

The times of the 400 people are recorded carefully and it is found that they can be modelled by a Normal distribution with mean 30 minutes and standard deviation 6 minutes.

(d) How many of the 400 participants would be expected to take:

   (i) less than 18 minutes

   (ii) 36 minutes or more? **[4]**

2(d)(i)	**Less than 18 minutes**
2(d)(ii)	**36 minutes or more**

On another Saturday, 400 people take part in the Parkrun. The distribution of their times is shown in this histogram.

Time (minutes)
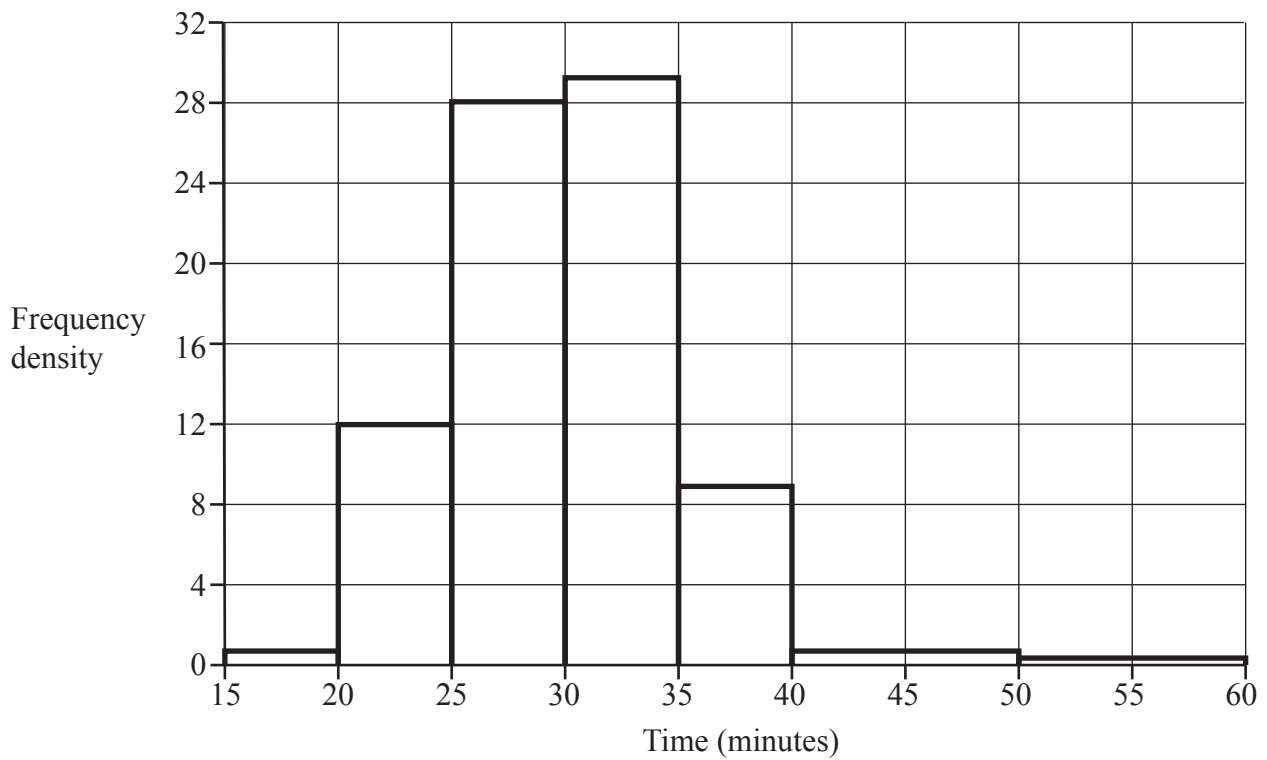
(e) Write down **one** feature of the histogram that shows the distribution of these runners' times is **not** Normal.

Give an explanation of why this might be the case for participants in the Parkrun. **[2]**

2(e)	Feature of the histogram
	Explanation

BLANK PAGE

PLEASE DO NOT WRITE ON THIS PAGE

Turn over for the next question

3 The present standard treatment for a particular disease is not always effective. A pharmaceutical company is developing a new medicine. Their tests show that it is more effective in treating the disease and they are now investigating possible side effects.

One side effect of the standard treatment is sleep loss in many patients. The company investigate whether this is also the case for their new treatment.

A number of patients with the disease volunteer to take part in a test relating to sleep loss. Those selected are allocated to three groups:

- receiving no treatment

- taking the standard treatment

- trialling the new medicine.

They are asked whether they have suffered no sleep loss, mild sleep loss or severe sleep loss since having the disease. The results are shown in **Fig. 3.1**.

Fig. 3.1

Observed frequency, f_o	Sleep loss			
Treatment	No	Mild	Severe	Total
No	9	16	5	30
Standard	5	11	24	40
New	28	15	7	50
Total	42	42	36	120

A chi-squared test is carried out using these data, using a 5% significance level.

(a) Which of the following is a suitable Null Hypothesis for the test?
Circle **one** of the options in the answer space. **[1]**

A The proportions of people with No, Mild or Severe sleep loss are independent of the treatment.

B The new treatment causes less sleep loss than the standard.

C Side effects should be avoided.

D The proportions of people with sleep loss depend on the treatment.

3(a)	A	B	C	D

(b) Complete **Fig. 3.2** in the Answer space giving the Expected frequencies. **[2]**

(c) Complete the calculation of X^2 in **Fig. 3.3** in the Answer space. **[2]**

3(b) **Fig. 3.2**

Expected frequency, f_e	Sleep loss			
Treatment	**No**	**Mild**	**Severe**	**Total**
No	10.5	10.5	9.0	**30**
Standard				
New	17.5	17.5	15.0	**50**
Total	**42**	**42**	**36**	

3(c) **Fig. 3.3**

$$X^2 = 0.214... + 2.880... + 1.777...$$
$$+ 5.785... + 0.642... + 12.000...$$
$$+ 6.300... + ... + 4.266...$$

$$= .... \text{ (to 1 d.p.)}$$

Turn over

(d) Complete the test and show that it is significant. **[4]**

(e) State whether there is enough evidence that the new treatment should be refused a licence for public use. Give your reason. **[2]**

3(d)	
3(e)	

BLANK PAGE

PLEASE DO NOT WRITE ON THIS PAGE

Turn over for the next question

Section B

Use the **pre-release data** to answer the questions. A hard copy of the **pre-release data** is provided.

4   (a)   Using figures from the pre-release data, show that the population of Cuba is decreasing by over 30 000 people per year.   **[2]**

4(a)	

Simon and Tara need to estimate the annual change in the number of people in the world.

They must decide on the best procedure to use given the data available to them. They have a paper copy of the pre-release data (the same as the one given out with this paper), but do not have an electronic version.

They suggest different methods, as shown in the table below.

Simon's method	**Tara's method**
Find the total population for all countries and the average growth rate per country. Then multiply the total population by the average growth rate and divide by 100.	Add up the values of $\text{Population} \times \dfrac{\text{Growth rate}}{100}$ for all countries

They decide to try the two methods on the five countries of North America because they are a small sample and so do not involve a lot of work.

They enter these countries into the table in **Fig. 4.1** in the Answer space, together with the information that they will need.

(b) Complete the **three** empty cells in **Fig. 4.1**. **[2]**

4(b)	Fig. 4.1			

Country	Population, p	Growth rate, $r\%$	$\dfrac{p \times r}{100}$
Bermuda	70 864	0.45	319
Canada	35 623 680		
Greenland	57 713	−0.03	−17
Saint Pierre and Miquelon	5 533	−1.08	−60
United States	326 625 791	0.81	2 645 669
Total	**362 383 581**	**0.88**	

(c) Find the annual increase in the population given by each of the methods, giving your answers to the nearest thousand people. **[2]**

(d) State, with a reason, which is better: Simon's method or Tara's method. **[1]**

4(c)	**Simon's method**
	Tara's method
4(d)	

Turn over

One of their colleagues then supplies them with an electronic version of the pre-release data. They use this to find the totals for all the countries for which the data are available. They are given below.

	Population, p	Growth rate, $r\%$	$\dfrac{p \times r}{100}$
Total for all countries	7 405 119 557	256.16	76 946 645

(e) Find the best estimate of the annual growth in the world's population to the nearest million people. **[1]**

You may not need to use all the figures given above.

4(e)	

BLANK PAGE

PLEASE DO NOT WRITE ON THIS PAGE

Turn over for the next question

5 **(a)** Write down the population and the total area of Namibia. **[1]**

(b) Using the pre-release data, calculate the value of C40/D40. Give your answer to **2** decimal places.

State what this value represents. **[2]**

(c) The value of SUM(C2:C57) is 1 222 204 378.

State what this value represents. **[1]**

(d) The value of SUM(D2:D57) is 30 312 642.

Calculate the value of SUM(C2:C57)/SUM(D2:D57).

State what this value represents. **[2]**

(e) A lot of Namibia is in the Namib desert.

Explain briefly how this influences the answers to parts **(b)** and **(d)**. **[1]**

5(a)	**Population**
	Land area
5(b)	**Value of C40/D40**
	What this value represents
5(c)	
5(d)	**Value of SUM(C2:C57)/SUM(D2:D57)**
	What this value represents
5(e)	

A student wants to investigate further.

She starts new columns, P and Q, on the spreadsheet of the pre-release data.

She enters

$$= C2/D2$$

into P2 and then copies it down to P237.

She then uses the spreadsheet's RANK command to obtain the ranks of the numbers in column P. She places them in column Q. The highest value of P is ranked 1.

(f) Row 180 had to be excluded.

Explain why. **[1]**

(g) Namibia is ranked 229.

How many countries have a lower value in column P than Namibia? **[1]**

5(f)	

5(g)	

6 (a) Show that, to the nearest whole number, the electricity consumption per person in Algeria is 1304 kWH per year. **[2]**

6(a)	

This letter is published in a newspaper in the UK.

> It is time we took our responsibility for the planet seriously.
>
> Throughout the world, the people who live in richer countries use more electricity and so cause more harmful emissions.
>
> The answer is simple. Tax rich people more.

Mary reads this and decides to check whether it is the case that countries with high GDP per capita also have high electricity consumption per capita. She works with a sample of 8 countries.

Mary selects 8 countries at random. They are given in **Fig. 6.1** in the Answer space, together with some relevant figures.

Mary completes **Fig. 6.1** and uses Spearman's Rank Correlation Coefficient to carry out a suitable hypothesis test at the 5% significance level.

(b) The Null Hypothesis for the test is
"H_0: There is no association between electricity consumption per capita and GDP per capita."

Write down the Alternative Hypothesis. **[1]**

6(b)	

(c) Complete **Fig. 6.1** and use it to calculate the value of r_s.

Carry out the test and show the result is significant. **[5]**

6(c) **Fig. 6.1**

Country	Electricity consumption per capita (kWH pa)	Electricity rank, x	GDP per capita PPP ($)	GDP rank, y	$d = x - y$	d^2
Malawi	103	8	1 200	8	0	0
Canada	14 502	1	48 100	2	−1	1
Chile	3 820	3	24 600	4	−1	1
India	818	7	7 200	7	0	0
Dominica	1 183	5	12 000	5	0	0
Poland	3 672	4	29 300	3		
Sweden	12 590	2	51 300	1		
Fiji	899	6	9 900	6		
				Total		

Mary is encouraged by the significant result but thinks it should be treated with caution because the sample was so small.

Mary uses all the available data from the pre-release to draw a scatter graph of electricity consumption per capita and GDP per capita.

Fig. 6.2 shows the scatter graph after Mary has removed Iceland from the data set, which was an outlier.

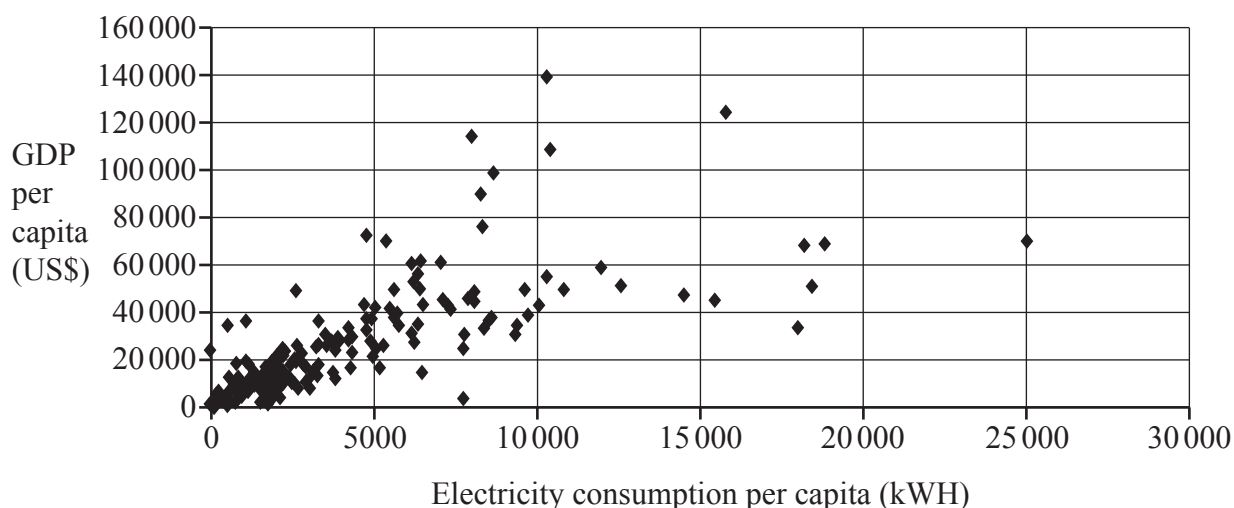**(d) (i)** Draw the line with equation $y = 6x$ on **Fig. 6.2** in the Answer space. **[1]**

(ii) State, with a reason, whether the line $y = 6x$ is a suitable Line of Best Fit for the scatter graph. **[1]**

(iii) Iceland has:

- electricity consumption per capita = 52 922 kWH

- GDP per capita = 52 100 US$

Would Iceland lie above, on or below the line $y = 6x$? **[1]**

Fig. 6.2

6(d)(i)	Answer on the graph above
6(d)(ii)	
6(d)(iii)	

(e) The letter to a newspaper earlier in the question included these two statements.

Statement 1

"*Throughout the world, the people who live in richer countries use more electricity.*"

Statement 2

"*The answer is simple. Tax rich people more.*"

State, giving brief explanations, whether these statements are justified.

Your explanations may be based on your general knowledge or on information from this question. **[2]**

6(e)	**Statement 1**
	Statement 2

END OF QUESTION PAPER

EXTRA ANSWER SPACE

If you need extra space use this lined page. You must write the question numbers clearly in the margin.