OCR
Oxford Cambridge and RSA

Friday 23 May 2025 – 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 page 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.

Section A

1 Emma is investigating the reactions of people of different ages to climate change issues.

Fig. 1.1 shows a questionnaire she designs.

She gives the questionnaire to 90 people. There are 30 people from each of these three age groups: 16 to 24, 25 to 64, 65 and over.

Fig. 1.1

Questionnaire

To what extent do you agree with these statements about climate change?

Circle your answer to each question.

0 = Disagree **1 = Neither agree nor disagree** **2 = Agree**

A All electricity should be generated from renewable sources.
0 1 2

B Diesel and petrol cars should be banned as soon as possible.
0 1 2

C The numbers of all flights should be cut down by at least half.
0 1 2

D Farmers should produce less meat.
0 1 2

Circle your age group: 16 to 24 25 to 64 65 and over

Thank you for participating.

Most people return the questionnaire to Emma, but some do **not**.

In all the returned questionnaires, all four questions have been answered.

For each returned questionnaire, Emma adds the person's answers for questions A to D together to find a score between 0 and 8 for each person.

Fig. 1.2 shows the number of people with each score, split by their age group.

(a) Complete **Fig. 1.2** in the answer space by filling in the empty cells. **[1]**

(b) How many people did **not** return the questionnaire? **[1]**

(c) According to the data in **Fig. 1.2**, which age group seems the most concerned about climate change?

How do you know? **[1]**

1(a) Fig. 1.2

Age \ Score	0	1	2	3	4	5	6	7	8	Total
16 to 24	0	0	1	1	2	1	4	9	12	**30**
25 to 64	1	2	3	2	4	4	3	3	3	
65 and over	4	3	3	2	1	3	1	0	3	
Total	**5**	**5**	**7**	**5**	**7**	**8**	**8**			

1(b)

1(c)

(d) Emma is concerned about the different numbers of responses from each age group.
She describes the questionnaires that were **not** returned as "missing data".

She considers two possible options for dealing with the missing data.

Option 1: Ignore the missing data and work only with the scores from the responses she has.

Option 2: Give all the missing responses a score of zero.

(i) Show that, using **Option 1**, the mean score for the whole sample is 5.04. **[3]**

(ii) Work out the mean score for the whole sample using **Option 2**. **[1]**

(iii) State which of the two options you think is more appropriate.

Give a reason for your answer. **[1]**

1(d)(i)	

1(d)(ii)	
1(d)(iii)	

2 A forest containing ash trees was planted around the year 1900. A hundred years later, 240 of the ash trees had matured and grown to full height. They supported a great quantity and variety of wildlife.

The heights of these 240 mature ash trees can be modelled by a Normal distribution with mean 32 metres and standard deviation 2.5 m.

(a) Show that over half of these ash trees were over 100 feet tall.

Use the conversion 1 foot is 0.3048 metres. **[2]**

(b) Calculate an estimate of the number of these ash trees over 35 metres tall. **[3]**

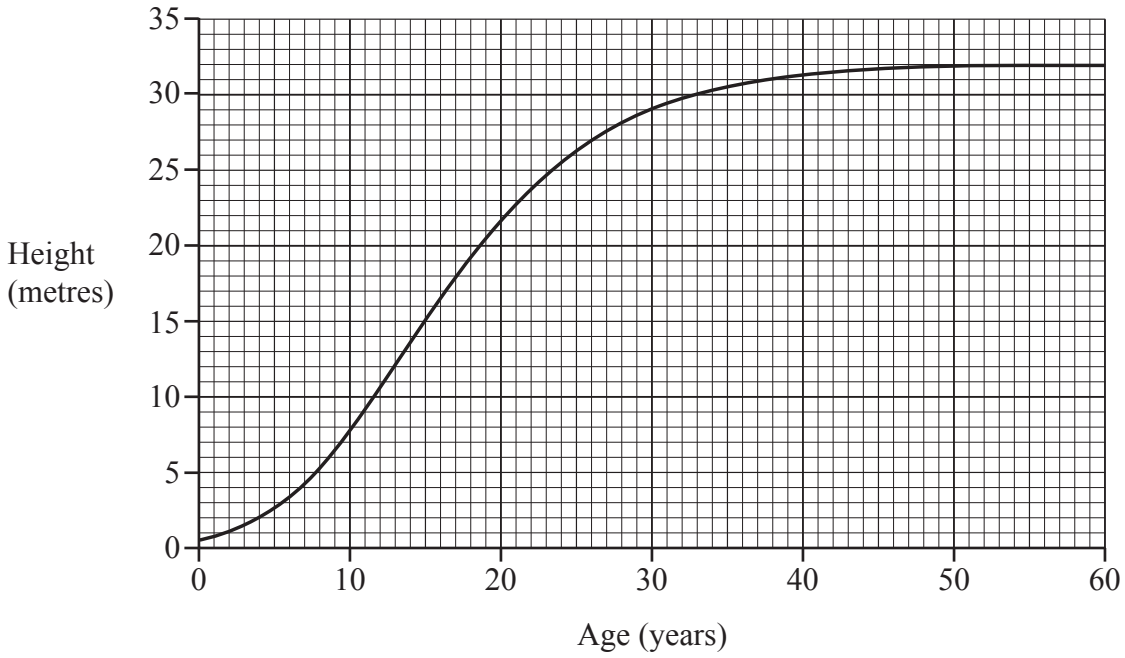
2(a)	
2(b)	

(c) **Fig. 2.1** shows the height of a typical ash tree against its age.

Use **Fig. 2.1** to estimate the rate of growth of a 20-year-old ash tree.

You will need to draw a suitable line on the graph. **[3]**

2(c) **Fig. 2.1**



(d) In 2012, some of the mature ash trees in the forest showed signs of ash dieback, a disease that occurs in ash trees. After a number of years an affected tree will die and fall down. There is no known cure.

Every year a large number of new ash trees (saplings) start to grow in the forest. Until the arrival of ash dieback in 2012, all these saplings had been treated as weeds and were taken out.

However, a small proportion of saplings, between 1% and 5%, have a natural resistance to ash dieback. So once it had become clear that the mature trees were suffering from ash dieback, many of the saplings were allowed to grow.

(i) Suggest a suitable number of saplings that should be allowed to grow to mature into 240 resistant ash trees.

Explain your reasoning. **[2]**

2(d)(i)	

(ii) Comment briefly on changes to the forest over time, from 2012 onwards.

Your answer should include what is likely to happen in the future. **[1]**

2(d)(ii)	

3 Sam drives to work each **weekday**, Monday to Friday.

She is often delayed by having to stop at a level crossing over a railway line.

Sam wants to know whether the delays are related to the time she leaves home.

She tries three different start times which she records as Early, Normal and Late.

She records the delay each day as None, Short, Medium or Long.

Fig. 3.1 shows the number of days in each category.

(a) Fill in the empty cells in **Fig. 3.1**. **[1]**

(b) How many weeks did Sam collect the data for? **[1]**

3(a) **Fig. 3.1**

Observed frequencies

START	DELAY: None	Short	Medium	Long	Total
Early	10	10	9	6	
Normal	7	11	16	11	**45**
Late	13	15	8	4	
Total	**30**				

3(b)

(c) Sam looks at the data in **Fig. 3.1**. She thinks the numbers suggest that changing her start time would reduce her travel time.

Sam wants to check how strong the evidence is before asking if she can arrive at work at a different time.

She carries out a chi squared test at the 10% significance level.

(i) Which of these is a suitable null hypothesis, H_0, for the test?

Tick (✓) **one** box. **[1]**

3(c)(i)

The later the start time the greater the delay.	☐
The length of the delay depends on the start time.	☐
The length of the delay is independent of the start time.	☐
Trains come in groups.	☐

(ii) **Fig. 3.2** shows the Expected frequencies.

Fill in the empty cells in **Fig. 3.2**. **[2]**

(iii) Fill in the **six** empty boxes in the calculation of the value of X^2 in **Fig. 3.3** on the next page. **[2]**

(iv) Carry out the chi squared test **and** interpret the result in context. **[4]**

3(c)(ii) **Fig. 3.2**

Expected frequencies

		DELAY				
		None	**Short**	**Medium**	**Long**	**Total**
START	**Early**	8.75	10.5	9.625	6.125	
	Normal	11.25	13.5	12.375	7.875	**45**
	Late	10				
	Total	**30**				

3(c)(iii) **Fig. 3.3**

$$X^2 = \frac{1.25^2}{8.75} + \frac{(-0.5)^2}{10.5} + \frac{(-0.625)^2}{9.625} + \frac{(-0.125)^2}{6.125}$$
$$+ \frac{(-4.25)^2}{11.25} + \frac{(-2.5)^2}{13.5} + \frac{3.625^2}{12.375} + \frac{3.125^2}{7.875}$$
$$+ \frac{3^2}{10} + \frac{3^2}{12} + \square + \square$$

$$= 0.17857... + 0.02380... + 0.04058... + 0.00255...$$
$$+ 1.60555... + 0.46296... + 1.06186... + 1.24007...$$
$$+ \square + \square + \square + \square$$

$$= 8.37 \text{ (}\mathbf{2}\text{ decimal places)}$$

3(c)(iv)

BLANK PAGE

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)** Write down the **median age** and the **annual population growth rate** for Italy. **[1]**

Fig. 4.1 shows a chart of the population profile for Italy.

(b) Draw a vertical line on **Fig. 4.1** to show the median age. **[1]**

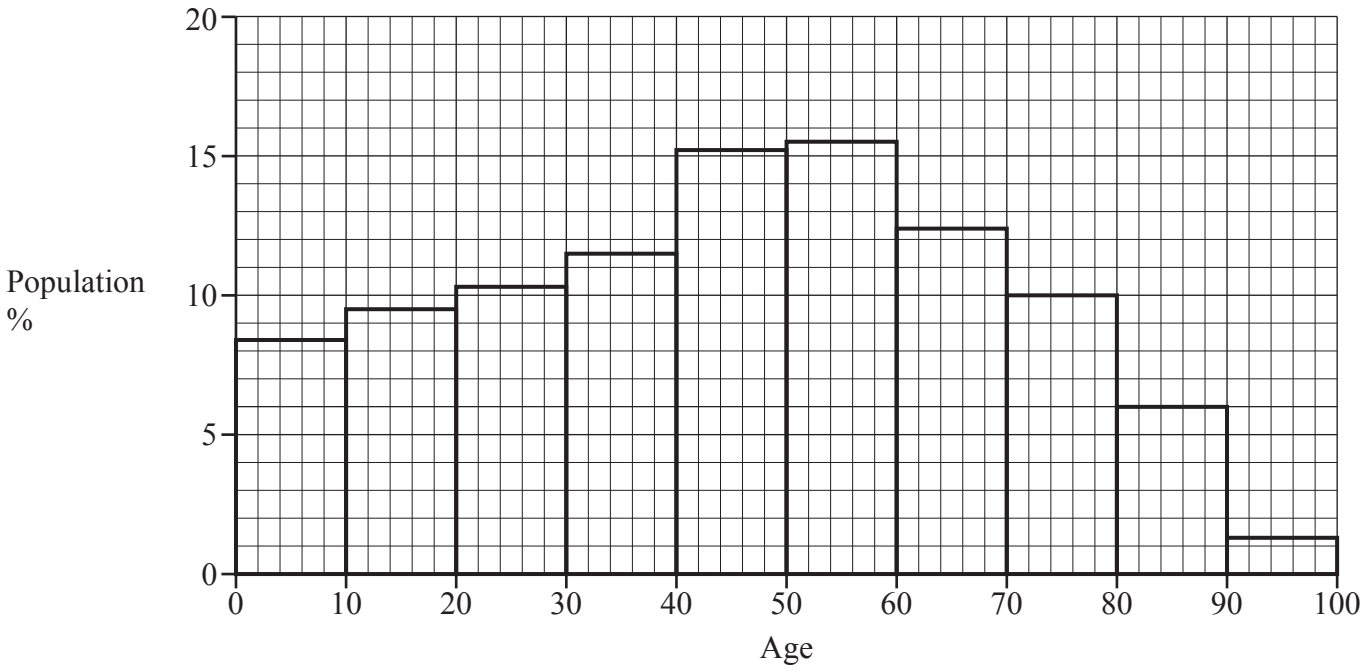
(c) Show that the position of the median line is consistent with the values on the chart.

Use calculations in your answer. **[2]**

4(a) **Italy** **Median age**

Annual population growth rate

4(b) **Fig. 4.1**



4(c)	

(d) State **one** feature of **Fig. 4.1** that is consistent with Italy's annual population growth rate. **[1]**

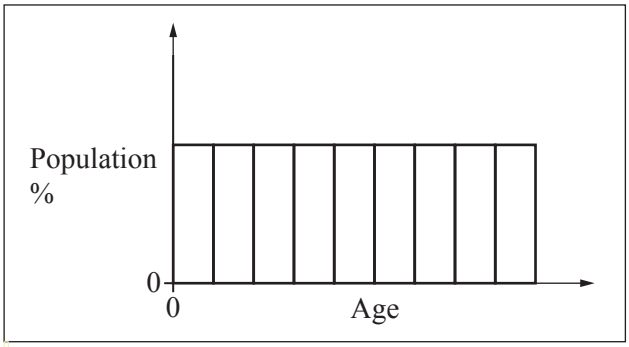
4(d)	

(e) For Pakistan, the median age is 22 and the annual population growth rate is 1.91%.

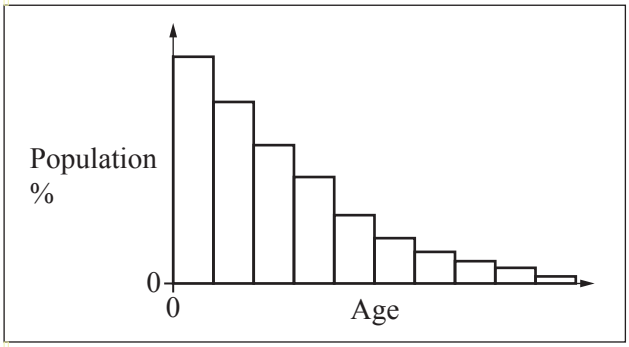
Which of these charts could show the complete population profile of Pakistan?

Tick (✓) **one** box. **[1]**

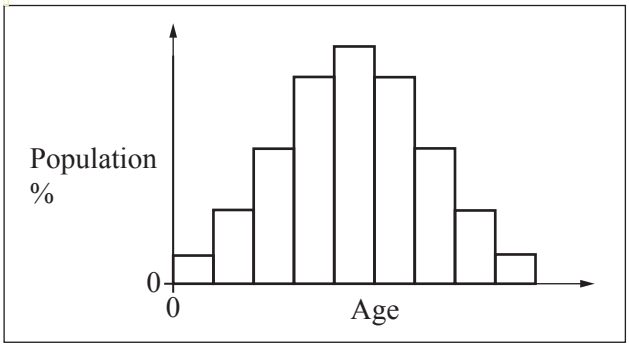
4(e)



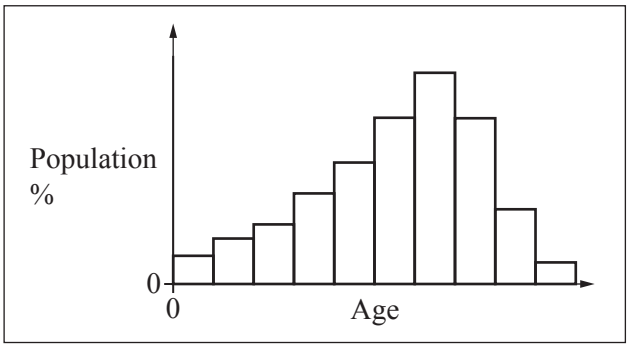
☐



☐



☐



☐

BLANK PAGE

DO NOT WRITE ON THIS PAGE

Turn over for the next question

5 **(a)** Libya is in North Africa.

(i) Explain how you know that the forest percentage in Libya is one of the lowest in Africa.

Suggest why this is the case. **[2]**

(ii) State the proportion of Libya's population that live in towns and cities. **[1]**

5(a)(i)	
5(a)(ii)	

Jane makes this statement:

'Countries with a low forest percentage are likely to have a high urban population percentage'.

She selects 10 countries for a pilot investigation of her statement.

Fig. 5.1 shows the countries she selects.

(b) Jane carries out a hypothesis test using Spearman's rank correlation coefficient.

The null and alternative hypotheses are:

H_0: There is no association between forest percentage and urban population percentage.

H_1: There is a negative association between forest percentage and urban population percentage.

(i) Is the test 1-tail or 2-tail? **[1]**

(ii) Fill in the empty cells in **Fig. 5.1**. **[2]**

(iii) Calculate the value of r_s and show that the test is significant at the 5% level. **[4]**

5(b)(i)	Tick (✓) **one** box. The test is 1-tail. ☐ The test is 2-tail. ☐

5(b)(ii) **Fig. 5.1**

Country	Urban population %	Urban rank, x	Forest %	Forest rank, y	$d = x - y$	d^2
Tunisia						
Botswana	72.9	6	19.8	8	–2	
Belize	46.6	9	60.6	1	8	
Greenland	87.9	3	0.0	10	–7	
Chile	88.0	2	21.9	7	–5	
Indonesia	58.6	8	51.7	3	5	
Russia	75.3	5	49.4	4	1	
Cuba	77.5	4	27.3	5	–1	
Belgium	98.2	1	22.4	6	–5	
Samoa	17.5	10	60.4	2	8	
				Total, ∑	**0**	**262**

5(b)(iii)

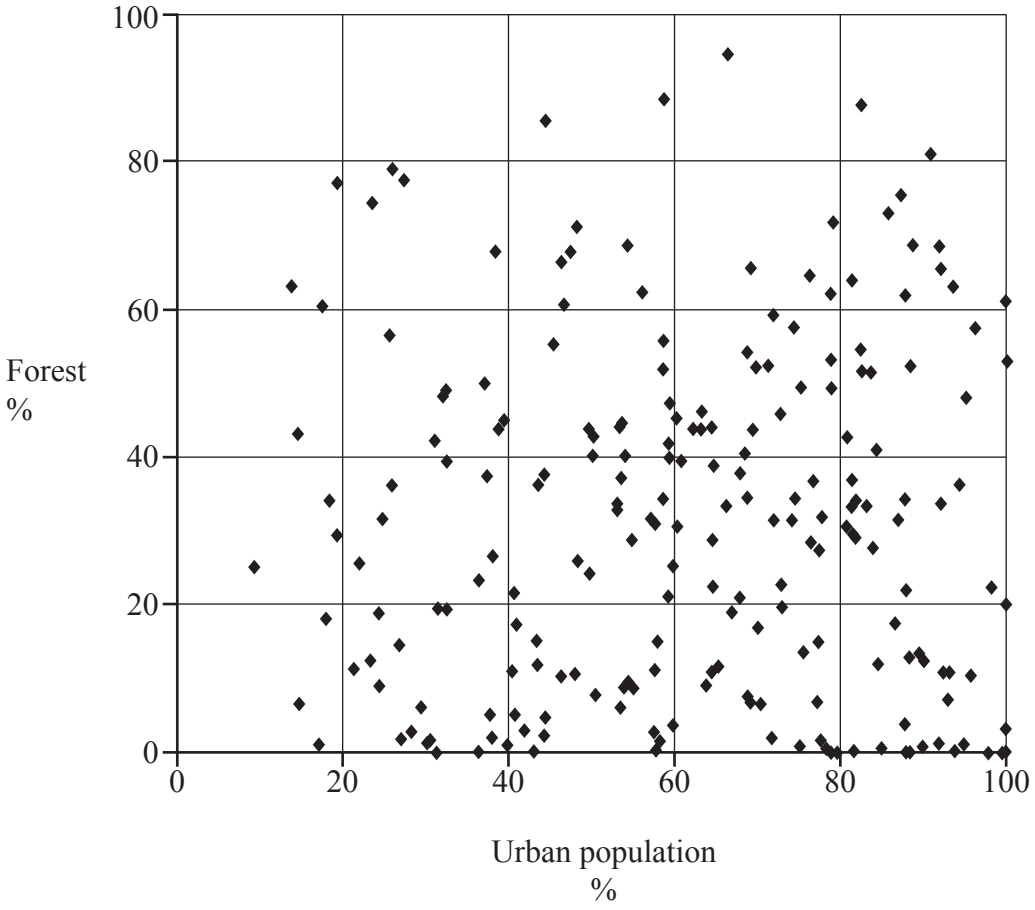
(c) Suggest how Jane selected these countries. **[1]**

5(c)

(d) Jane cleans the data from the whole pre-release spreadsheet appropriately.

She uses it to draw the scatter diagram of forest % against urban population % in **Fig. 5.2**.

Fig. 5.2



(i) Decide whether these statements about the data in **Fig. 5.2** are **true** or **false**.

Tick (✓) **one** box in each row. **[1]**

5(d)(i)

	True	False
The higher the urban population percentage, the lower the forest percentage.		
There are some countries where almost everyone lives in towns or cities.		

(ii) Jane then uses the cleaned pre-release data to work out the Pearson's product moment correlation coefficient for the data in columns D and F.

The answer is –0.0087 ... .

Compare the Pearson's product moment correlation coefficient to the Spearman's rank correlation coefficient that you found in part **(b)(iii)**.

You should state which you consider to be the more reliable answer and suggest a reason for the difference between the two. **[2]**

5(d)(ii)	

6 (a) (i) Write down the **population** and **CO_2 emissions** of India. **[1]**

(ii) Calculate the average CO_2 emissions per person in India. **[1]**

(b) Show that the average CO_2 emissions per person is nearly 10 times higher in the United States than it is in India. **[2]**

6(a)(i)	**India**	**Population**
		CO_2 emissions
6(a)(ii)		
6(b)		

Sasha uses the spreadsheet of pre-release data to investigate the impact that different countries are having on the environment.

He starts new columns P, Q and R and then:

- copies column C, Population, into column P
- copies column H, CO_2 emissions, into column Q
- enters " = Q2/P2" into cell R2
- copies cell R2 down to R236
- formats column R to give the numbers in the cells to **2** decimal places.

(c) (i) Show that the number in cell R2 is 3.39 and state what it represents. **[2]**

(ii) The number in cell R209 is 5.84.

What does this tell you? **[1]**

6(c)(i)	
6(c)(ii)	

Before he can continue, Sasha needs to clean the data.

In some rows, column Q contains "#N/A" or the number zero. In those rows he deletes the content of cells Q **and** P so that they are both blank.

(d) State **one** country for which the cells in columns P and Q are now both blank. **[1]**

After Sasha's data cleaning, the total of the entries in column Q is 35 532 499 700.

The total for column P is 7 978 557 003.

(e) Determine whether people in the United Kingdom have above or below average CO_2 emissions when compared with the rest of the world.

Show your reasoning. **[2]**

6(d)	
6(e)	

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.