



Oxford Cambridge and RSA

Wednesday 14 October 2020 – 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



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** Graph 1.1 is based on historic data on the number of arctic lynx caught annually. The data allow the population of these animals to be estimated. The graph shows the estimated population, in suitable units, every 2 years from 1857 to 1905 inclusive.

© Encyclopædia Britannica, 2012. Item removed due to third party copyright restrictions.

Graph 1.1

- (i) State two features of the population. [2]
- (ii) Estimate the median value of the lynx population from 1857 to 1905, showing your reasoning. [2]
- (iii) Estimate the percentage increase in the lynx population from 1871 to 1887. [3]
- (iv) In 1985, the badger population of England was estimated to be about 100 000. Another study in 1995, using a different methodology, suggested the much larger figure of about 230 000. At the time a newspaper claimed that this showed that badgers were out of control.

Give two reasons why the newspaper's claim may have been false. [2]

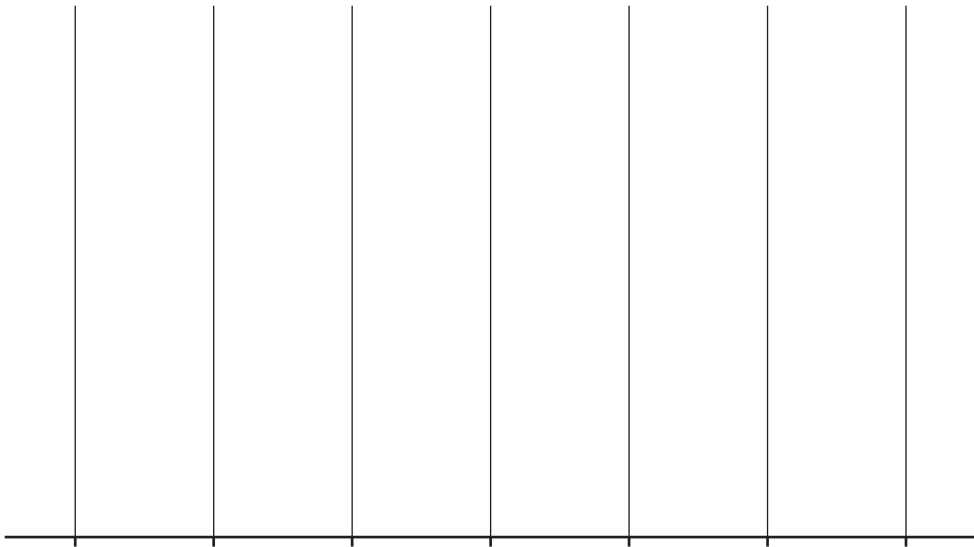
1 (i)	Feature 1
	Feature 2

1 (ii)	
1 (iii)	
1 (iv)	Reason 1
	Reason 2

2 A Normal distribution has mean 110 minutes and standard deviation 10 minutes.

(i) Sketch this distribution on the grid provided, placing suitable values along the horizontal axis. [3]

(ii) What percentage of observations from this distribution would you expect to be less than 100 minutes? [2]

2 (i)										
2 (ii)	<table border="1" data-bbox="245 1077 1465 1559"><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr></table>									

A particular weekday train journey is scheduled to take 100 minutes. Table 2.1 gives a summary of the times that it actually took on the 120 journeys during a period of 24 weeks.

Time, t mins	$98 \leq t < 100$	$100 \leq t < 110$	$110 \leq t < 120$	$120 \leq t < 130$	$130 \leq t < 160$	$t = 188$
Frequency	2	72	41	2	2	1

Table 2.1

These figures are used to give estimates of 110 minutes for the mean time and 10 minutes for the standard deviation.

- (iii) Show how the estimate of 110 minutes for the mean time can be obtained. [2]
- (iv) What percentage of these observations are less than 100? [1]
- (v) Explain briefly why you should not be surprised by the difference between your answers to parts (ii) and (iv). [1]

2 (iii)	
2 (iv)	
2 (v)	

- 3 Sally is a university psychology student. She is conducting an investigation into public attitudes towards crime. She carries out the same experiment with groups of students at 4 universities (A, B, C and D). She describes a recent court case in which someone was found guilty and given a prison sentence of 2 years. She asks them whether they think the sentence was too light, about right or too harsh. Their responses are given in Table 3.1.

Observed frequency, f_o	Too light	About right	Too harsh	Total
A	30	5	5	40
B	8	10	16	34
C	30	5	5	40
D	12	10	24	46
Total	80	30	50	160

Table 3.1 Observed frequencies

- (i) State which of the following best describes Sally's sample.

Simple random, Opportunity, Stratified, Self-selected, Cluster, Quota.

[1]

- (ii) Sally says "I have proved that exactly half of the country want the courts to give more severe sentences."

Give 2 different criticisms of Sally's statement.

[2]

3 (i)	
3 (ii)	Criticism 1
	Criticism 2

Sally's tutor suggests that she uses her data to carry out a χ^2 goodness of fit test.

- (iii)** Which one of the following is a correct statement of the null hypothesis for this test?
- The numbers of students in the 3 groups are the same in each university.
 - The proportions of students in the 3 groups are different in the 4 universities.
 - Students are typical of people in the country as a whole.
 - People are equally divided between the 3 groups.
 - The proportions of students in the 3 groups are the same in the 4 universities. [1]
- (iv)** Complete Table 3.2 and show that the test statistic $X^2 = 42.1$ to 3 significant figures. [4]
- (v)** Complete the χ^2 test using a 5% significance level, stating your conclusion clearly. [4]

3 (iii)																																		
3 (iv)	<table border="1" style="margin: auto; border-collapse: collapse;"> <thead> <tr> <th style="width: 15%;">Expected frequency, f_e</th> <th style="width: 20%;">Too light</th> <th style="width: 20%;">About right</th> <th style="width: 20%;">Too harsh</th> <th style="width: 25%;">Total</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;">A</td> <td style="text-align: center;">$\frac{80 \times 40}{160} = 20$</td> <td></td> <td></td> <td style="text-align: center;">40</td> </tr> <tr> <td style="text-align: center;">B</td> <td style="text-align: center;">17</td> <td style="text-align: center;">6.375</td> <td style="text-align: center;">10.625</td> <td style="text-align: center;">34</td> </tr> <tr> <td style="text-align: center;">C</td> <td></td> <td></td> <td></td> <td style="text-align: center;">40</td> </tr> <tr> <td style="text-align: center;">D</td> <td style="text-align: center;">23</td> <td style="text-align: center;">8.625</td> <td style="text-align: center;">14.375</td> <td style="text-align: center;">46</td> </tr> <tr> <td style="text-align: center;">Total</td> <td style="text-align: center;">80</td> <td style="text-align: center;">30</td> <td style="text-align: center;">50</td> <td style="text-align: center;">160</td> </tr> </tbody> </table>				Expected frequency, f_e	Too light	About right	Too harsh	Total	A	$\frac{80 \times 40}{160} = 20$			40	B	17	6.375	10.625	34	C				40	D	23	8.625	14.375	46	Total	80	30	50	160
Expected frequency, f_e	Too light	About right	Too harsh	Total																														
A	$\frac{80 \times 40}{160} = 20$			40																														
B	17	6.375	10.625	34																														
C				40																														
D	23	8.625	14.375	46																														
Total	80	30	50	160																														
	Table 3.2 Expected frequencies																																	
	$X^2 = 5.0000$	+	+																															
	+ 4.7647	+ 2.0613	+ 2.7191																															
	+	+	+																															
	+ 5.2609	+ 0.2192	+ 6.4446																															
	=	= 42.1 (3 sf)																																

3 cont

3 (v)	

BLANK PAGE

PLEASE DO NOT WRITE ON THIS PAGE

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 (i) Find the population density of Egypt in people per square kilometre. [2]

(ii) In the pre-release data, the land area for Holy See (Vatican City) is recorded as 0.

Given that this figure is not an error, state what it tells you about the country's land area. [1]

(iii) The population density of Monaco was recently estimated to be 19 250 people per square kilometre. Show that this is consistent with the information in the pre-release data. [3]

4 (i)	
4 (ii)	

4 (iii)	

5 Jo thinks it would be good for the environment if people used less electricity. He carries out an investigation to find out the countries where people use the most electricity.

On the spreadsheet containing the pre-release data, he sets up a new column, Q, for average electricity consumption per person.

The command for cell Q2 is

$= E2 * 1000000 / C2$

(i) State the number that appears in cell Q2.

State precisely what it represents, including its units. **[2]**

Jo plans to copy cell Q2 down column Q so that he can carry out investigations relating to this quantity in different countries. Before doing this, he cleans the data in columns C and E to ensure that the cells in column Q contain only numbers.

(ii) State one thing that Jo does when cleaning the data in columns C and E. **[1]**

The country with the highest value cell in column Q is Iceland; that with one of the lowest values is Liberia.

(iii) Calculate the values of the cells in column Q corresponding to Iceland and Liberia.

Suggest an explanation of why Liberia has a low value. **[3]**

Jo finds the correlation coefficients between the entries in column Q and those in each of columns D (total area), J (birth rate) and L (GDP per capita).

Jo writes down the answers but forgets which pairs of variables give rise to them. In numerical order they are -0.429, 0.0835, 0.638.

(iv) State, with a reason, which correlation coefficient you think matches which pair of variables. In answering this you may need to apply your general knowledge. **[3]**

5 (i)	Number in cell Q2
	What it represents
5 (ii)	

5 (iii)	Iceland
	Liberia
	Reason for Liberia's low value
5 (iv)	Q and D (Total area)
	Q and J (Birth rate)
	Q and L (GDP per capita)

6 Sunni is carrying out an investigation into countries' birth rates and their rates of population growth.

As a pilot project he calculates Spearman's rank correlation on the sample of countries given in Table 6.1, below.

- (i) Look at the countries in Table 6.1 and describe how Sunni chose the sample. [1]
- (ii) Complete Table 6.1. [3]
- (iii) Calculate the value of r_s . [2]
- (iv) Sunni uses the value of r_s as the test statistic in a 2-tail hypothesis test at the 10% significance level. State the null and alternative hypotheses and the outcome of the test. [4]

6(i)							
6(ii)	Country	Birth rate	Birth rate rank, x_i	Population growth rate	Growth rate rank, y_i	$d_i = x_i - y_i$	d_i^2
	Algeria						
	Angola	44.2		3.52			
	Belize	24		1.8			
	Bermuda	11.3	9	0.45	7	2	4
	Argentina	16.7	6	0.91	6	0	0
	Afghanistan	37.9		2.36			
	Russia	11	10	-0.08	9	1	1
	Anguilla	12.5	8	1.97		5	25
	Albania	13.2	7	0.31	8	-1	1
	American Samoa	19.6	5	-1.3	10	-5	25
						Total	0

Table 6.1

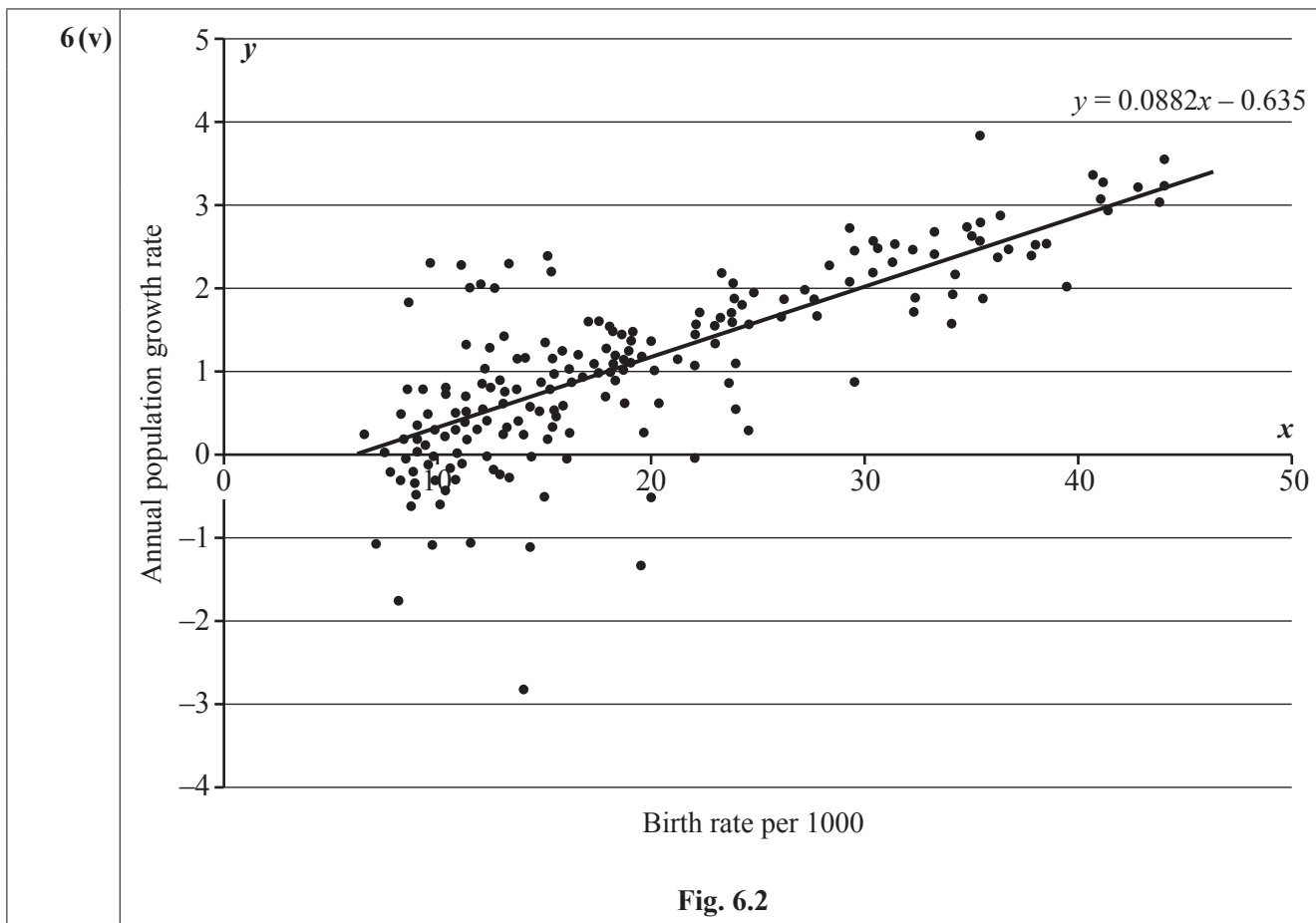
6 (iii)	
6 (iv)	

6 cont

Sunni continues his investigation on a spreadsheet. Fig. 6.2 shows the scatter diagram and line of best fit for all the countries for which the data are available. The variables are Birth rate per 1000, x , and Annual population growth rate, $y\%$.

(v) Ring the point on Fig. 6.2 corresponding to the Cook Islands. [1]

(vi) State one factor, other than birth rate and death rate, that influences a country's rate of population increase. [1]



6(vi)	

The equation of the line of best fit is $y = 0.0882x - 0.635$.

This line can be used as a model of the relationship between the variables.

(vii) Use the equation of the line to calculate the coordinates of the point where it crosses the x -axis.

State what information this gives you, according to the model. **[2]**

(viii) Suggest a reason why the line has not been drawn far enough to the left to cross the y -axis. **[1]**

6 (vii)	Intersection with the x-axis
6 (viii)	Information

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).

This block contains a large area of lined paper for writing answers. It features a vertical margin line on the left side and horizontal dotted lines for writing. The lines are evenly spaced and extend across the width of the page.

Ruled area for writing, consisting of a solid vertical line on the left and horizontal dotted lines extending across the page.



Copyright Information

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download from our public website (www.ocr.org.uk) after the live examination series. If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact The OCR Copyright Team, The Triangle Building, Shaftesbury Road, Cambridge CB2 8EA.

OCR is part of the Cambridge Assessment Group; Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.