OCR
Oxford Cambridge and RSA

Wednesday 8 June 2022 – 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.

BLANK PAGE

PLEASE DO NOT WRITE ON THIS PAGE

Answer **all** the questions.

Section A (30 marks)

1 Jane is hoping to open a charity shop to support a local animal shelter. She will need more people to help with the work but before trying to recruit anyone she feels she needs more information.

She asks the 12 other people in her yoga group to fill in the questionnaire in **Fig. 1.1** anonymously.

Helping animals

For each question please circle one response

A Are you in favour of work to support animals?

1 Not at all   2 A little   3 On the whole   4 Strongly   5 Very strongly

B Would you welcome a new charity shop on the high street raising money for animal welfare?

1 Not at all   2 Prefer not   3 On the whole   4 Overall, yes   5 A great idea

C Would you be prepared to donate your unwanted possessions to an animal charity shop?

1 Never   2 Occasionally   3 Sometimes   4 Often   5 Always

D How many hours a week would you be prepared to spend helping in an animal charity shop?

1 None   2 One to Two   3 Three to Five   4 Six to Ten   5 Over Ten

Fig. 1.1

(a) Choose the term in the answer space which describes Jane's sample best. **[1]**

1(a)	**Circle one of the following terms**
	Cluster Opportunity Quota Self-selected Simple random Stratified

1 continued

Jane records the scores of 1, 2, 3, 4 or 5 in the table in **Fig. 1.2**. She treats them as numerical scores.

Question \ Score	1	2	3	4	5
A	5	0	1	2	4
B	6	0	0	3	3
C	4	1	0	4	3
D	6	0	1	2	2

Fig. 1.2

(b) Complete the table in **Fig. 1.3** in the answer space. **[1]**

1(b)

Score	**1**	**2**	**3**	**4**	**5**
Frequency			2		

Fig. 1.3

(c) Show that one person did not answer all the questions. **[1]**

1(c)

1 **(d)** Choose the term in the answer space which describes the distribution of the scores best. **[1]**

1(d)	**Circle one of the following terms**
	Positively skewed Negatively skewed Uniform Normal Bimodal

(e) **(i)** Find the mean score per person per question answered, giving your answer to **2** decimal places. **[3]**

(ii) Give one reason why this figure is not helpful to Jane. **[1]**

1(e)(i)	
1(e)(ii)	**Reason**

2 A large number of seeds of a forest tree are planted. Many years later, the heights of a sample of the now mature trees are reported to have mean 40 feet and standard deviation 8 feet.

(a) **(i)** Convert the mean and standard deviation of these heights into metres.
[1 foot = 30.5 cm to 3 s.f.] **[2]**

(ii) It is suggested that the mean and standard deviation reported are only approximate.

Give a reason to support this view. **[1]**

2(a)(i)	**Mean**
	Standard deviation
2(a)(ii)	**Reason**

In a larger study it is found that the heights of the mature trees grown under usual conditions are Normally distributed with mean 12 m and standard deviation 2 m. This distribution is shown in **Fig. 2.1** in the answer space for part **(b)**.

(b) Number the scale on the horizontal axis in **Fig. 2.1**. **[2]**

(c) Find the percentage of mature trees that are over 10 m tall. **[3]**

2(b)

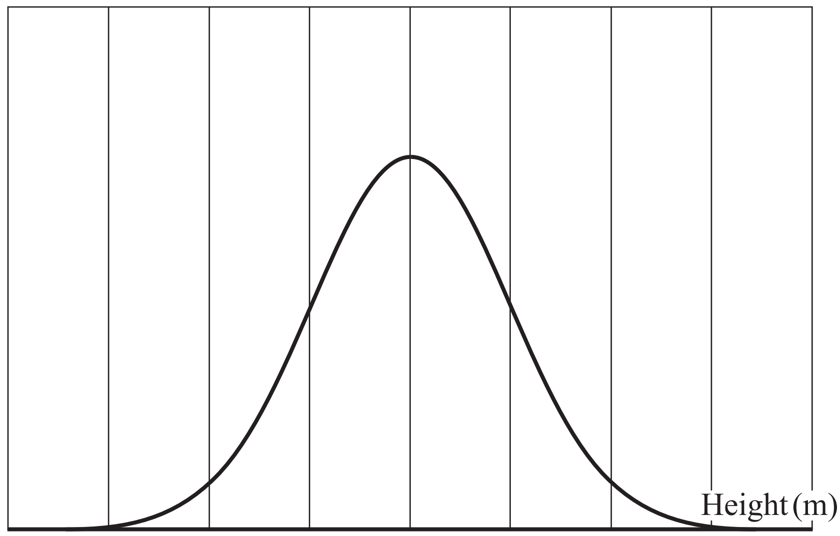

Fig. 2.1

2(c)

2 continued

In one particular plantation 254 seeds grew into young trees. They were then thinned out. The smallest 50% of the young trees were cut down.

Many years later all the surviving trees were measured. The distribution of their mature heights is shown in the frequency chart in **Fig. 2.2**.

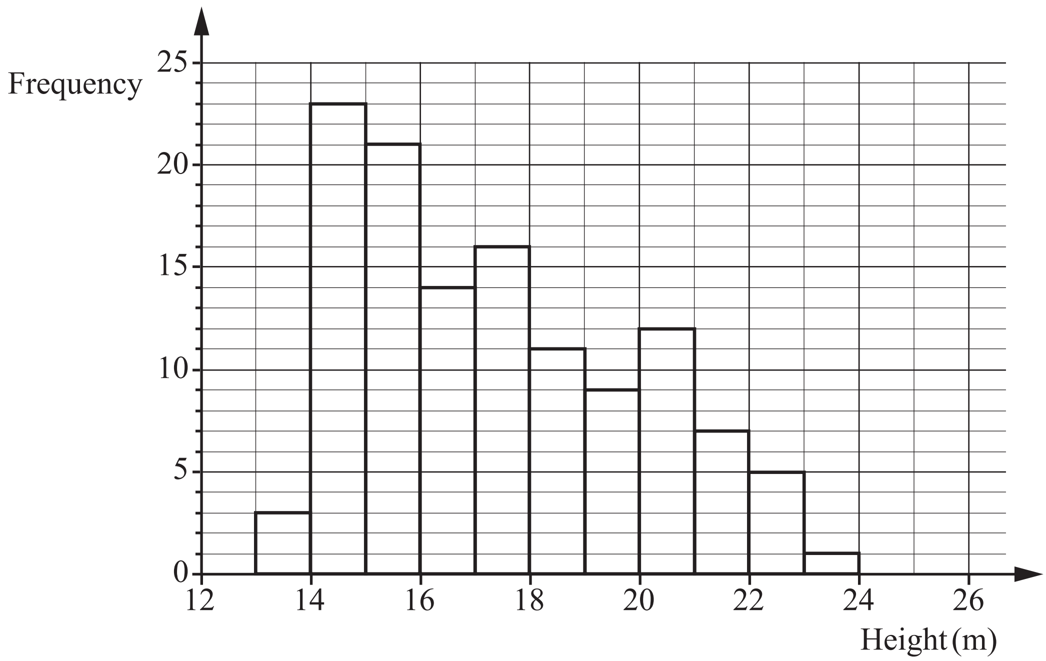

Fig. 2.2

(d) How many of the trees that were not cut down when the thinning took place had subsequently died? **[2]**

(e) Comment on the distribution of the heights of the mature trees. **[1]**

2(d)	

2(e)	

3 A shop sells devices for counting the number of steps you take each day. **Fig. 3.1** shows their own special notice for advertising them.

Statistically Proved

Taking more steps in the day makes you sleep longer at night

Fig. 3.1

A customer challenges this and asks the shop for details of the proof.
Fig. 3.2 shows the shop manager's written reply.

For a week I recorded the number of steps I took each day and how long I slept for that night. That gave me my theory. Look at my figures.

Day	*Mon*	*Tue*	*Wed*	*Thu*	*Fri*	*Sat*	*Sun*
Steps	6459	6871	7762	6967	12678	12511	12413
Sleep (hours)	6.4	6.8	7.4	5.9	7.7	8.1	7.3

So I used my data to do a proper hypothesis test using the Spearman coefficient. It showed there is a significant correlation. You can see I have proved statistically that taking more steps makes you sleep longer at night.

Fig. 3.2

The customer decides to check the shop manager's test.

(a) Write down the null hypothesis and the alternative hypothesis for a 1-tail test using Spearman's rank correlation coefficient. **[2]**

(b) Complete the table in the answer space and calculate the value of r_s. **[4]**

(c) Carry out the test, using a 5% significance level, and state the conclusion. **[3]**

(d) Make **two** criticisms of the manager's reply to the customer in **Fig. 3.2**. **[2]**

3(a)	**Null hypothesis**
	Alternative hypothesis

3(b)

Day	Steps	Steps rank, x	Sleep	Sleep rank, y	$d = x - y$	d^2
Mon	6459		6.4			
Tue	6871	6	6.8	5	1	1
Wed	7762	4	7.4	3	1	1
Thu	6967		5.9			
Fri	12678	1	7.7	2	–1	1
Sat	12511	2	8.1	1	1	1
Sun	12413	3	7.3	4	–1	1
				$\sum$		

Spearman's coefficient, r_s =

3(c)

3(d) **Criticism 1**

Criticism 2

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 the Seychelles, its Birth rate per thousand population and its Annual population growth (%). **[1]**

(b) **(i)** Find how many babies are born in a year in the Seychelles. **[1]**

(ii) Find by how many people the population of the Seychelles increases over a year. **[2]**

(c) The death rate for the Seychelles is 7.1 per thousand population.

Show that there is a net immigration of about 100 people per year into the Seychelles. **[3]**

4(a)	**Population** .... **Birth rate** .... per 1000
	Growth rate .... %
4(b)(i)	**Babies**
4(b)(ii)	**Increase in population**
4(c)	

5 Some countries are described as "young", meaning that a high proportion of the people who live there are young. Similarly countries where a high proportion of the population are not young can be described as "old".

(a) (i) Write down the median ages of France, Japan, the United States and Zambia. **[1]**

(ii) Place the four countries in order of median age, from youngest to oldest. **[1]**

(b) Explain briefly why the median age is a better measure of how young or old a country is than the life expectancy. **[1]**

5(a)(i)	**France**	**Japan**
	United States	**Zambia**
5(a)(ii)	**Order (youngest first)**	
5(b)		

Rishi is interested in the differences between typical median ages in different continents or regions.

He uses the command =AVERAGE(... : ...) to find the mean values of the median ages.

(c) Fill in the blank spaces in the command needed to find the mean of the median ages for the countries in Africa. **[1]**

5(c)	=AVERAGE(:)

5 continued

Rishi uses the spreadsheet to draw the bar chart in **Fig. 5.1** below showing the mean values of the median ages for each continent or region.

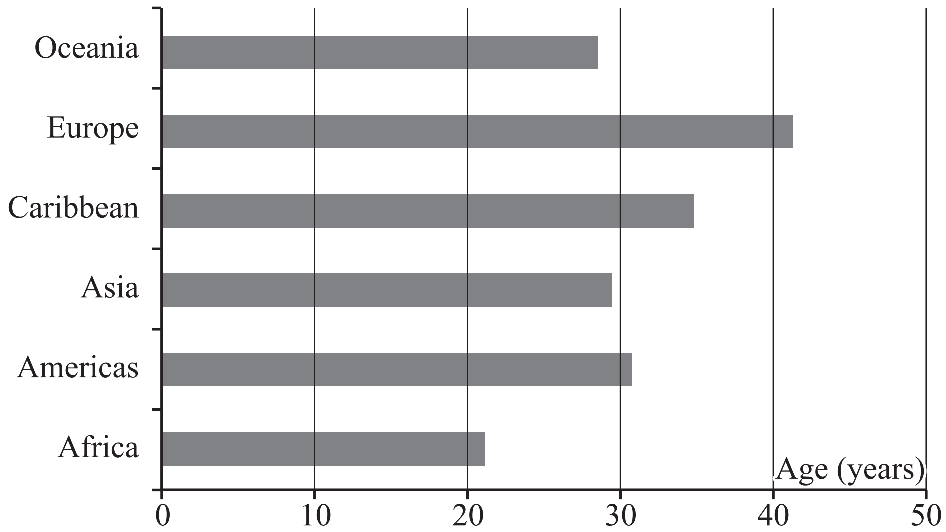

Fig. 5.1

He also uses the spreadsheet to calculate the standard deviations of the median ages for the different continents. The continent with the lowest standard deviation is Europe with 3.8 years and the continent with the highest is Asia with 6.6 years.

(d) For Africa the mean is 21.2 years and the standard deviation is 5.3 years.

Use these figures to identify the median ages of countries that can be considered outliers for Africa.

Identify **3** countries that are outliers for Africa, justifying your answer. **[2]**

5(d)	

Rishi looks more carefully at the median ages for Africa. He wishes to compare them with those from the other continents.

He notices that the figures for Northern Africa and the island countries are quite different from those for the rest of Africa. He decides to treat all those countries as outliers and to remove them from his investigation.

This leaves only countries in mainland Southern Africa. For these 42 countries the mean is 19.2 and the standard deviation is 2.5.

(e) Give one reason in favour of removing all those countries and one reason against doing so. **[2]**

(f) There are several issues that the government of a young country may regard as problems. Use your general knowledge to state one of them briefly. **[1]**

5(e)	**Reason in favour of removing the countries**
	Reason against removing the countries
5(f)	

6 Casey is investigating the rate of population growth around the world.

She starts by cleaning the data, excluding countries where the growth rate figure looks unreliable. Then she places countries for each region or continent into three groups according to their annual growth rate: Negative (less than 0%), Low (between 0% and 1.5%) and Fast (over 1.5%).

Her results are shown in **Fig. 6.1**.

	Observed frequency, f_o			
	Negative	**Low**	**Fast**	**Total**
Africa	0	10	46	**56**
Americas	2	20	4	**26**
Asia	5	27	18	**50**
Caribbean	5	14	4	**23**
Europe	18	29	1	**48**
Oceania	6	14	4	**24**
Total	**36**	**114**	**77**	**227**

Fig. 6.1

Casey notices the differences between the regions and wonders if they are statistically significant. She decides to carry out a χ^2 test on the data.

(a) Select which one of the following statements is the null hypothesis. **[1]**

A The regions are all different.

B Most regions have a Low growth rate.

C The proportions of Negative, Low and Fast growth rates are independent of the region.

D There is a correlation between regions and annual growth rate.

E The proportions of Negative, Low and Fast growth rates depend on the region.

6(a)	**Circle one of the statements for the null hypothesis**				
	A	B	C	D	E

6 **(b)** Show that the Expected frequency of Negative growth rate for the Americas is less than 5. **[3]**

6(b)

Because the Expected frequency for this cell is so low, Casey decides to combine the figures for Negative and Low under a new heading of "Slow".

(c) Complete **Fig. 6.2** in the answer space. **[2]**

6(c)

	Observed frequency, f_o				Expected frequency, f_e		
	Slow	**Fast**	**Total**		**Slow**	**Fast**	**Total**
Africa	10	46	**56**		37.004 ...	18.995 ...	**56**
Americas							
Asia	32	18	**50**		33.039 ...	16.960 ...	**50**
Caribbean	19	4	**23**		15.198 ...	7.801 ...	**23**
Europe	47	1	**48**		31.718 ...	16.281 ...	**48**
Oceania	20	4	**24**		15.859 ...	8.140 ...	**24**
Total	**150**	**77**	**227**				

Fig. 6.2

6 **(d)** **(i)** Complete the calculation of X^2 in the answer space. Give your working correct to **2** decimal places and your final answer to **1** decimal place. **[2]**

(ii) State which element makes the smallest contribution to the value of X^2. **[1]**

(e) Carry out the test at the 1% significance level and state the conclusion. **[5]**

6(d)(i)	$X^2 = 19.71 + 38.39$ $+ 1.35 + 2.63$ $+ 0.03 + 0.06$ $+ 0.95 + 1.85$ $+ 7.36 + 14.34$ $+ \ldots + \ldots$ $= \ldots$ (1 dp)
6(d)(ii)	**Element with smallest contribution**

6(e)	

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).