

Please check the examination details below before entering your candidate information

Candidate surname

Other names

Centre Number

Candidate Number

Pearson Edexcel Level 3 GCE

Wednesday 7th June 2023

Morning (Time: 2 hours)

Paper
reference

9ST0/01

Statistics

Advanced

PAPER 1: Data and Probability

You must have:

Statistical formulae and tables booklet, Calculator

Total Marks

**Candidates may use any calculator allowed by Pearson regulations.
Calculators must not have retrievable mathematical formulae stored in them.**

Instructions

- Use **black** ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B).
- Use **black** ink or ball-point pen.
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions and ensure that your answers to parts of questions are clearly labelled.
- Answer the questions in the spaces provided
– *there may be more space than you need.*
- You should show sufficient working to make your methods clear.
Answers without working may not gain full credit.
- Unless otherwise stated, inexact answers should be given to three significant figures.
- Unless otherwise stated, statistical tests should be carried out at the 5% significance level.

Information

- A booklet 'Statistical formulae and tables' is provided.
- There are 7 questions in this question paper. The total mark for this paper is 80.
- The marks for **each** question are shown in brackets
– *use this as a guide as to how much time to spend on each question.*

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.
- If you change your mind about an answer, cross it out and put your new answer and any working underneath.

Turn over ►

P72909A

©2023 Pearson Education Ltd.
N:1/1/1/1/1/




Pearson

Answer ALL questions. Write your answers in the spaces provided.

- 1 Yin is the manager of a busy fast-food restaurant. He wants to get customer feedback on the quality of the service at the restaurant.

Three different sampling methods are considered by Yin.

Method 1

A staff member stands at the main exit to survey customers as they leave the restaurant.

- (a) For **Method 1**, give

(i) **one** advantage,

(1)

(ii) **one** disadvantage.

(1)

Method 2

An internet link to the survey is printed on the back of customer receipts.

- (b) For **Method 2**, give

(i) **one** advantage,

(1)

(ii) **one** disadvantage.

(1)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 1 continued

Method 3

Customers are asked for their email address when they pay their bill, and a survey is emailed to them afterwards.

(c) For **Method 3**, give

(i) **one** advantage,

(1)

(ii) **one** disadvantage.

(1)

Yin would like the survey to be ongoing. He wants to minimise expense to himself and inconvenience to customers.

(d) Select which one of the three methods, 1, 2 or 3, you would recommend for Yin to use.

Briefly explain your selection.

(1)

Yin would like the response rate to this survey to be as high as possible.

(e) Make **two** suggestions for Yin to improve the response rate of his survey.

(2)

(Total for Question 1 is 9 marks)



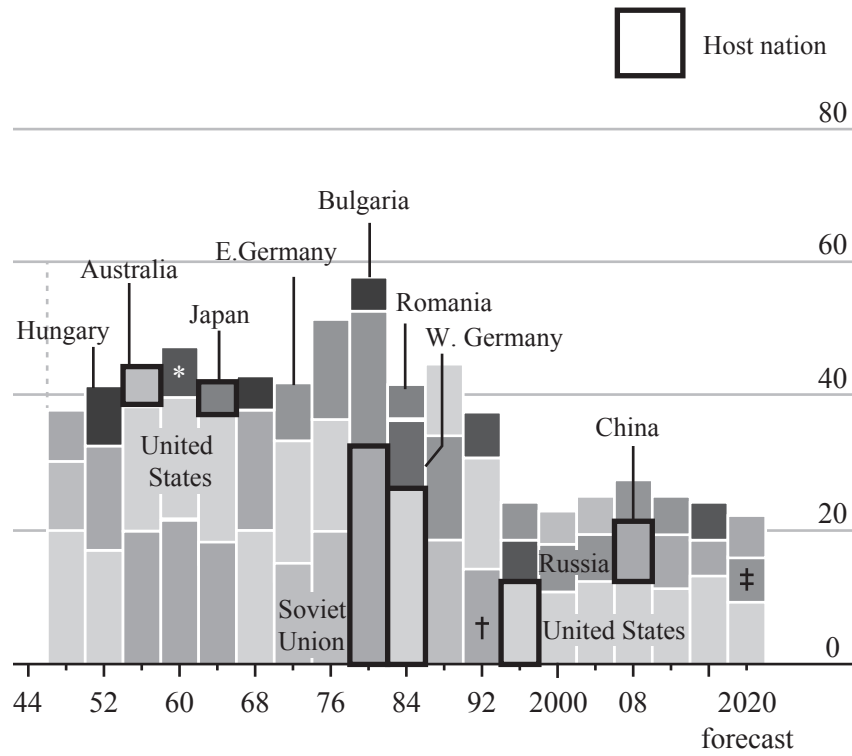
P 7 2 9 0 9 A 0 3 2 4

2 A graph was shown in an article about Summer Olympic teams in *The Economist* in July 2021

A section of the graph is shown in **Figure 1**.

Home advantage

Share of Olympic medals won, top three teams at each games, %



*United Team of Germany † Unified Team (Soviet Union)
‡ Russian Olympic Committee

Figure 1

(a) Estimate the highest proportion of medals won by a single nation in one Summer Olympic Games.

(1)

(b) State how many Summer Olympic Games saw the host nation score **more** medals than any other team.

(1)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 2 continued

In the original article, the graph in **Figure 1** was printed in colour.

- (c) Explain how colour could be used to make the graph clearer and easier to understand.

(1)

- (d) State **four** criticisms, other than the lack of colour, of the graph.

(4)

(Total for Question 2 is 7 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



3 Many websites allow various filters, such as product features, to be applied to searches by online shoppers.

In November 2021, the Currys website had 400 models of television for sale.

Three different filters for televisions were considered. They are described below.

8K resolution (R) A very high-resolution screen with many pixels.
This technology is fairly new and relatively expensive.

Screen size $\geq 75''$ (S) The screen size is at least 75 inches.
Generally a larger screen is more expensive.

Voice control (V) The television can be controlled using voice commands.
This technology is fairly new.

The three filters were applied singly and in combination with other filters. The number of models of television found on the website after these filters were applied can be seen in **Figure 2**.

Filter applied to			Models of television
8K resolution (R)	Screen size $\geq 75''$ (S)	Voice control (V)	
✓			16
	✓		74
		✓	304
✓	✓		9
✓		✓	16
	✓	✓	69
✓	✓	✓	9

(Date source: www.currys.co.uk, accessed 25/11/2021)

Figure 2

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

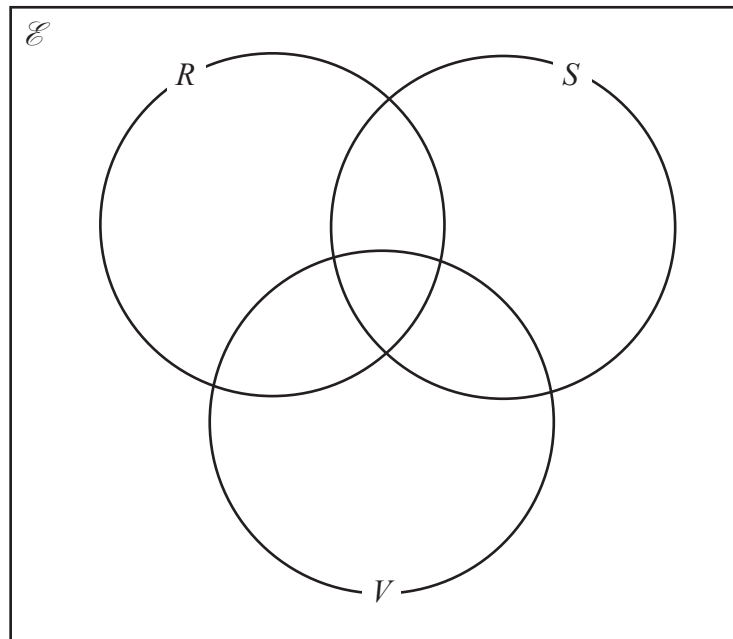


Question 3 continued

(a) Use the data in **Figure 2** to complete the Venn diagram frequencies below.

You are reminded that there are 400 models in total.

(6)



One model of television is selected at random from those listed on the website.

(b) Find the probability that the television

(i) has 8K resolution

(1)

(ii) has a screen size **smaller** than 75 inches and voice control

(2)

(iii) will not be displayed if any of the filters are applied

(1)



- 4 A video game is being created by a game developer. In the game, zombies are spawned into the game at random (a zombie will appear in the game at random).

Zombies are spawned one at a time.

In Version 1 of the game, the time between spawns (in seconds) follows the probability distribution in **Figure 3**.

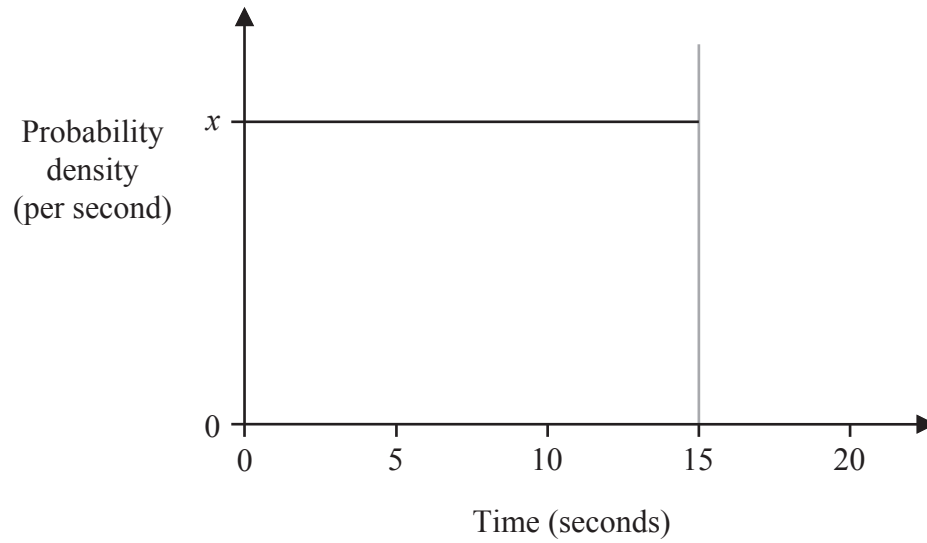


Figure 3

- (a) Write down the value of x

(1)

Ahmed is playing Version 1 of the game, and a zombie has just spawned into the game.

- (b) Calculate

- (i) the probability that no zombies will spawn in the next 10 seconds,

(1)

- (ii) the probability that **each** of the next three zombies will spawn more than 10 seconds after the last.

(2)



Question 4 continued

In Version 2 of the game, the time between spawns (in seconds) follows the exponential distribution with **mean** 5 seconds.

Zara is playing Version 2 of the game, and a zombie has just spawned into the game.

(c) Calculate the probability that

(i) at least one zombie will spawn in the next 5 seconds,

(3)

(ii) no zombies will spawn in the next 10 seconds,

(2)

(iii) no more than 10 zombies will spawn in the next minute.

(4)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 4 continued

The game developer decides that the probability found in (c)(iii) is too low.

She decides the probability that no more than 10 zombies will spawn within a minute of another zombie should be as close to 0.95 as possible, but no less than 0.95

- (d) Using this information, find, to the nearest integer, the average number of zombies that will spawn in the next minute.

You may use trial and improvement.

(1)

- (e) Using your answer to (d) state the new mean value for the exponential distribution in (c)

(1)

(Total for Question 4 is 15 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



P 7 2 9 0 9 A 0 1 1 2 4

5 In 1950, W.S. Robinson conducted a statistical analysis of some data from the 1930 census in the USA. Part of his analysis investigated correlation between rates of literacy and place of birth.

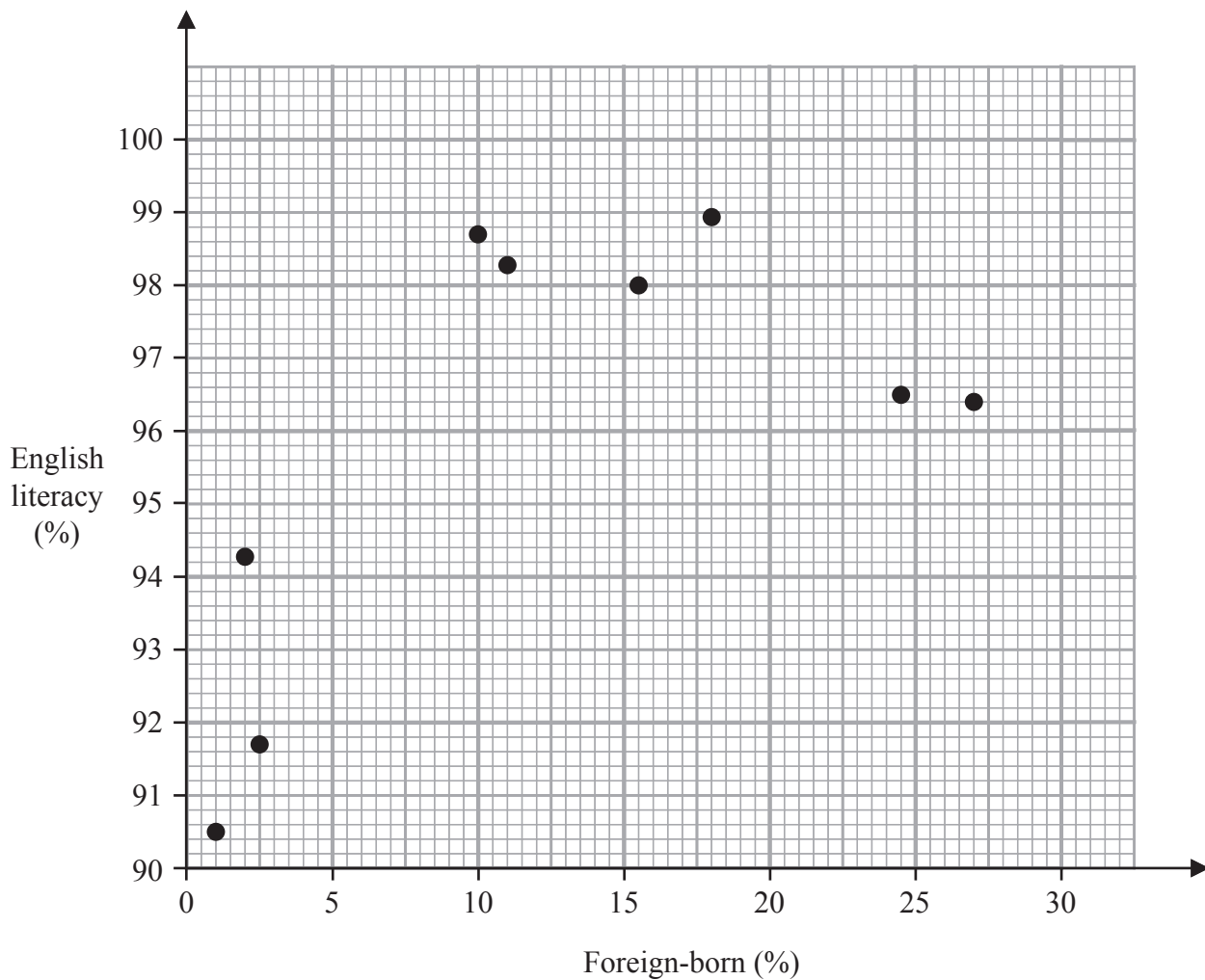
In 1930, USA census data was split into 9 geographical regions.

For each region, the following two percentages were calculated.

Foreign-born Percentage of the adult population born outside the USA.

English literacy Percentage of the adult population able to read English to a standard sufficient to function in society.

In Robinson's paper, the full data was not provided, though a scatter diagram was given similar in form to the one in **Figure 4**.



[source: <https://doi.org/10.2307/Data/2087176>]

Figure 4

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 5 continued

(a) Complete the table below using **Figure 4**.

(2)

Foreign-born (%)	1	2	2.5	10	11	15.5	18	24.5	27
English literacy (%)									

Robinson calculated a correlation coefficient for this data.

(b) Calculate the value for the Pearson's product-moment correlation coefficient between Foreign-born (%) and English literacy (%) for adults in the 9 regions.

(1)

(c) State **one** reason why it may be inappropriate to calculate the Pearson's product-moment correlation coefficient between Foreign-born (%) and English literacy (%) for adults in the 9 regions.

(1)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



P 7 2 9 0 9 A 0 1 3 2 4

Question 5 continued

In the same paper, Robinson considers correlation between Foreign-born (%) and English literacy (%) for **all of the adults** in the census, considered **at individual level** (so each data point represents one person, not one region).

He produces the following statistic.

$$r = -0.118$$

- (d) Explain why this statistic may be unexpected when compared with the result obtained in (b) (2)

- (e) Using your understanding of the context, give a short explanation to explain how the unexpected situation regarding the results in (b) and (d) may have occurred. (4)

(Total for Question 5 is 10 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE



P 7 2 9 0 9 A 0 1 5 2 4

6 Asbestos is a type of mineral used in construction that has been linked to health risks.

Riddhi works for an asbestos removal company. She is researching the removal of two types of asbestos,

- White asbestos (or chrysotile)
- Blue asbestos (or crocidolite)

When there is an asbestos find during an inspection, its location and type are recorded.

An asbestos find is called an **instance** of asbestos.

Figure 5 shows instances of the types of asbestos as a proportion of **total** instances of asbestos recorded by the company in 2022.

Asbestos type	Proportion of total instances
White	0.972
Blue	0.018
Other	0.010

Figure 5

Not all instances of asbestos are removed, as sometimes it is safer to cover the asbestos up and leave it in place.

Figure 6 shows the proportion of instances in 2022 that were removed, by type of asbestos.

Asbestos type	Proportion of instances that are removed
White	0.015
Blue	0.971
Other	0.153

Figure 6

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

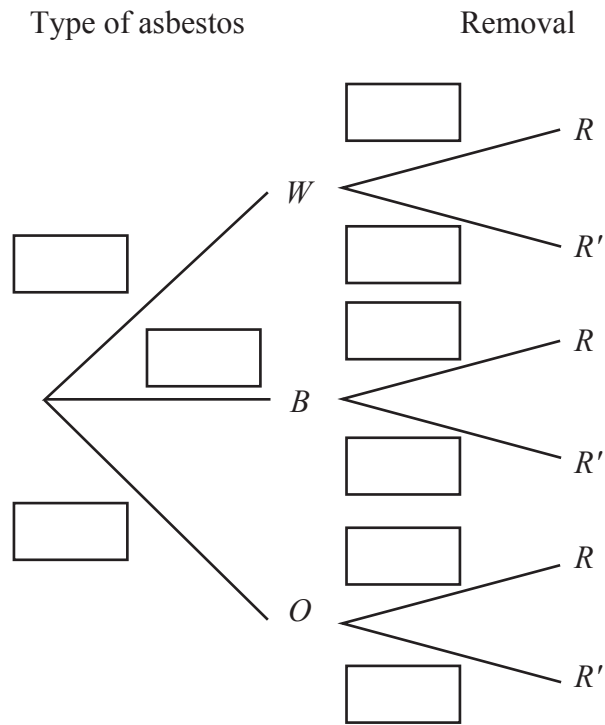
DO NOT WRITE IN THIS AREA



Question 6 continued

(a) Enter the correct probability values into each box in the tree diagram below.

(2)



(b) Calculate the probability that a randomly selected instance was removed.

(2)



Question 6 continued

(c) Find the probability that a randomly selected instance is blue asbestos, given that it was removed by the company.

(4)

(d) Suggest a possible reason, in context, why the probability found in (c) is greater than the probability that a randomly selected instance is blue asbestos.

(1)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 6 continued

The removal of different types of asbestos requires different types of safety equipment.

Riddhi states that the company should invest mostly in safety equipment for removing white asbestos, as more than 97% of asbestos instances are of this type.

- (e) Explain whether you agree with Riddhi’s statement.

You should include **numerical justification** from (c)

(3)

Handwriting practice lines for answer (e)

(Total for Question 6 is 12 marks)



- 7 Divya carries out a large-sample study to investigate the 2D:4D digit ratio for adult females

The 2D:4D digit ratio is calculated by dividing the length of the index finger (second digit) by the length of the ring finger (fourth digit) on their dominant hand.

Previous studies suggest that the 2D:4D digit ratio for adult females is normally distributed.

Figure 7 shows the values of the **sextiles** in Divya's large-sample study.

[A **sextile** is similar to a quartile, but the data has been split into six equal proportions instead of 4.]

Sextile	Approximate percentile	Digit ratio
S_1	17th	0.942
S_2	33rd	0.958
S_3	50th	0.971
S_4	67th	0.984
S_5	83rd	1.000

Figure 7

- (a) State the approximate value of the mean of the digit ratio in Divya's study.

Explain your answer.

(2)

- (b) Explain why the approximate value of the standard deviation of the digit ratio in Divya's study is 0.029

(3)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 7 continued

Divya also carries out a second large-sample study to investigate the 2D:4D digit ratio for adult males.

Previous studies suggest that the 2D:4D digit ratio for adult males is also normally distributed.

Her second large-sample study yields the following statistics:

$$\begin{aligned}\bar{x} &= 0.952 \\ s &= 0.034\end{aligned}$$

(c) Find the probability that a randomly chosen male adult has a 2D:4D digit ratio

(i) greater than 1 **(1)**

(ii) between 0.9 and 1 **(1)**

(d) Find the probability that a randomly chosen adult male has a 2D:4D digit ratio lower than that of a randomly chosen adult female.

You should use your answers to (a) and (b) **(5)**

(Total for Question 7 is 12 marks)

TOTAL FOR PAPER IS 80 MARKS





DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE



P 7 2 9 0 9 A 0 2 3 2 4

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

