Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number	Candidate Number

Pearson Edexcel Level 3 GCE

Time 2 hours

Paper reference **9ST0/01**

Statistics

Advanced
PAPER 1: Data and Probability

You must have:	Total Marks
Statistical formulae and tables booklet Calculator	

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).
- **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 ▶

©2022 Pearson Education Ltd.

Pearson

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

1 The journey times of commuters for a selection of those travelling into two London train stations are given in **Figure 1** and **Figure 2**.

Journey times for London King's Cross

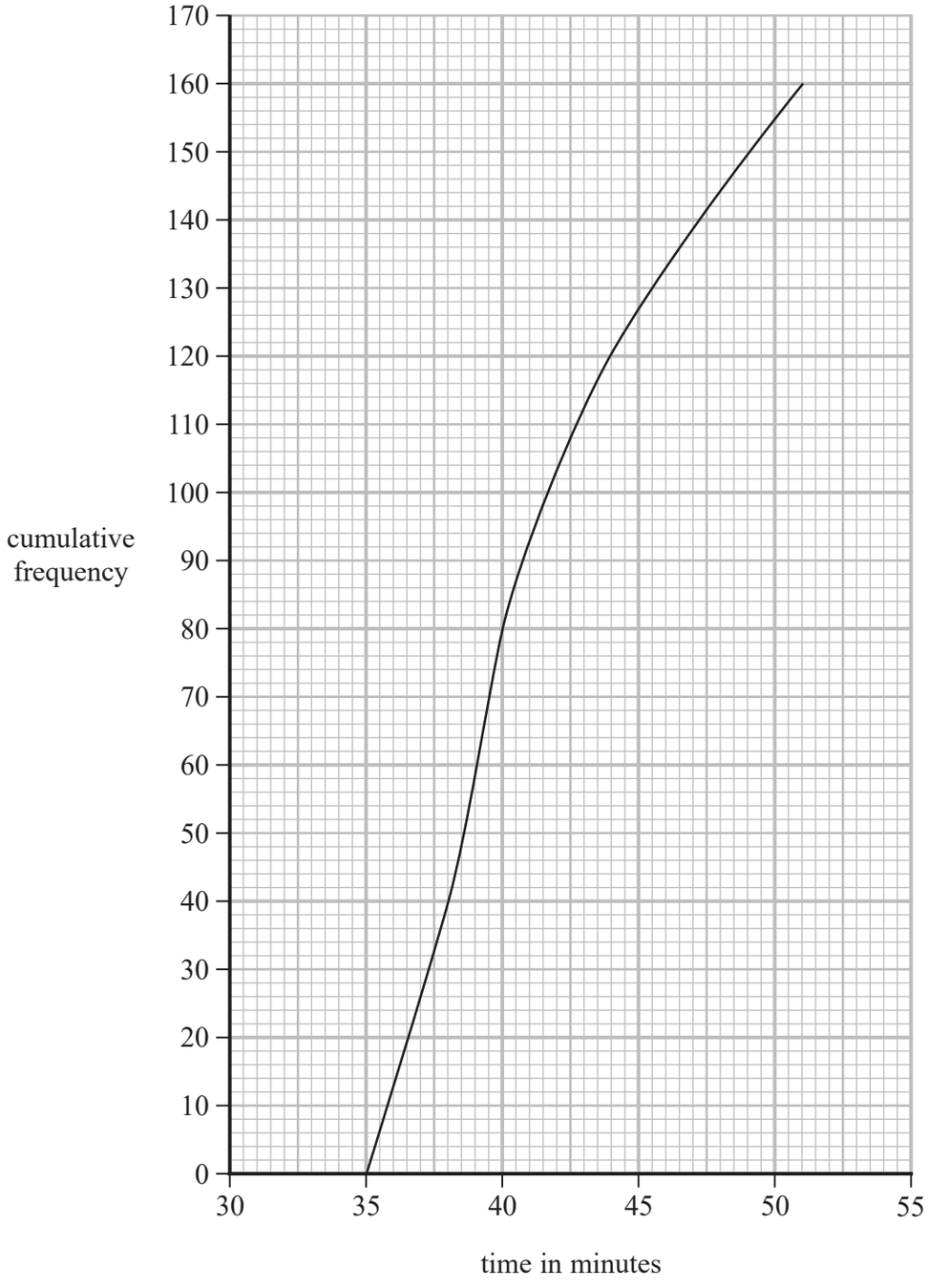

Figure 1

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

Journey time (min) for London Waterloo

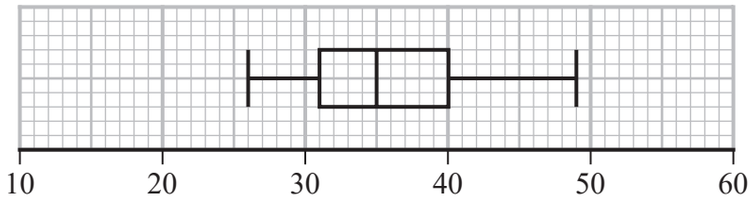

Figure 2

These times are the time taken from **leaving their house** to arriving at their destination station.

(a) Compare the average and spread of journey times for commuters travelling to these two stations.

(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 1 continued

These journeys are made up of two parts.

The first part of the total journey is the journey travelling from the commuter's home to their local station, where they take the train to London King's Cross or London Waterloo. This part of the journey may include waiting for this train to arrive.

The second part of the total journey is the journey by train to London King's Cross or London Waterloo.

All passengers in **Figure 1** took the 7:15 train from Stevenage to London King's Cross.

All passengers in **Figure 2** took the 7:45 train from Wimbledon to London Waterloo.

All the data is from the same day.

Lyra looks at the information in **Figure 1** and **Figure 2** and concludes that the train journey time from Wimbledon to London Waterloo is shorter than that from Stevenage to London King's Cross.

(b) State an assumption that Lyra has made in order for this conclusion to be valid.

(1)

(c) Identify **two** sources of variability for these commuters from Wimbledon to London Waterloo?

(2)

(Total for Question 1 is 9 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

2 Alana is researching show dogs.

Show dogs have a desired range of heights.

The maximum and minimum desired male and female heights, in inches, for a random sample of breeds of show dog are shown in **Figure 3**.

Alana plots the minimum desired male heights and minimum desired female heights in **Figure 4**.

		Male		Female	
		Minimum desired height	Maximum desired height	Minimum desired height	Maximum desired height
Breed	Affenpinscher	9	11.5	9	11.5
	Afghan Hound	26	28	24	26
	Akita	26	28	24	26
	American Eskimo Dog	9	19	9	19
	Australian Shepherd	20	23	18	21
	Bearded Collie	21	22	20	21
	Beauceron	25.5	27.5	24	26.5
	Belgian Malinois	24	26	22	24
	Bernese Mountain Dog	25	27.5	23	26
	Black Russian Terrier	27	30	24	26.5
	Cattle Dog	18	20	17	19
	Chesapeake Bay Retriever	21	24	23	26
	Chinese Shar-Pei	18	20	18	20
	Doberman Pinscher	26	28	24	26
	Flat-Coated Retriever	23	24.5	22	23.5
	Foxhound	22	25	21	24
	Greater Swiss Mountain Dog	25.5	28.5	23.5	27
	Irish Red and White Setter	24.5	26	22.5	24
	Irish Wolfhound	32	No max	30	No max
	Italian Greyhound	13	15	13	15
	Keeshound	17	19	16	18
	Neapolitan Mastiff	26	31	24	29
	Pointer	25	28	23	26
	Shetland Sheepdog	13	16	13	16
	Staffordshire Terrier	18	19	17	18
	Water Spaniel	15	18	15	18
	Welsh Springer Spaniel	18	19	17	18
	Wirehaired Pointing Griffon	22	24	20	22

[Source: American Kennel Club Breed Statistics]

Figure 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 2 continued

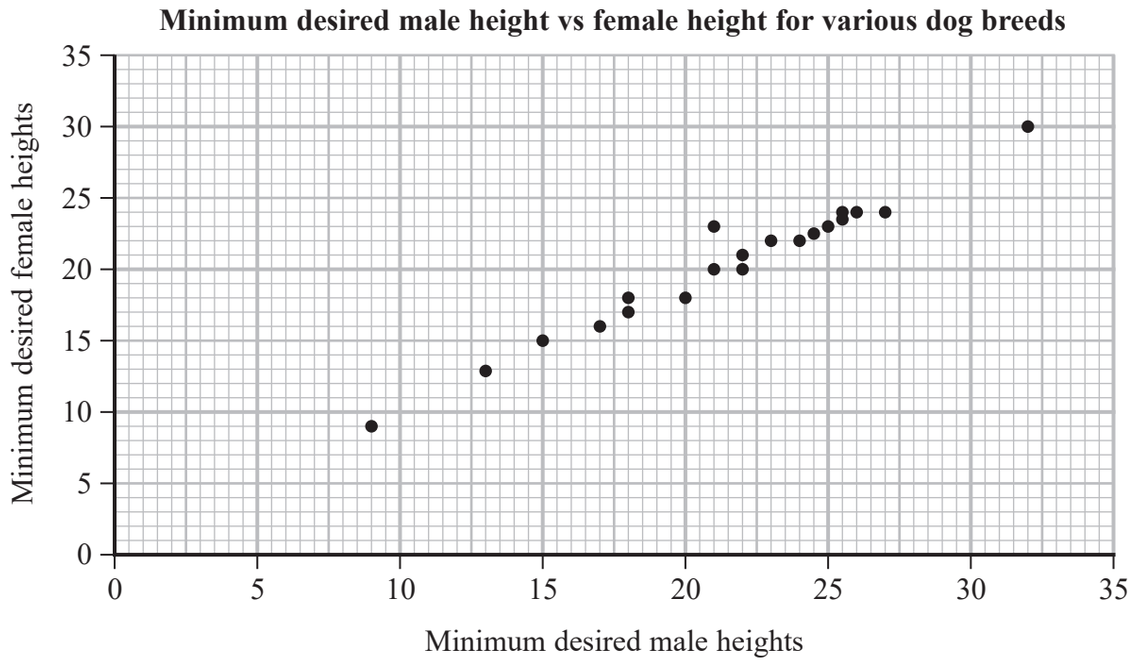

Figure 4

(a) Explain why Alana might have chosen to plot her data before calculating a correlation coefficient.

(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 2 continued

After examining **Figure 4** Alana realises that one of the dog breeds has had the male and female columns recorded the wrong way round.

(b) State, with a reason, which breed you believe has been recorded incorrectly.

(2)

The PMCC between the minimum desired height for males and females was calculated as r = 0.988

Alana says that the PMCC will be higher if the heights were measured in centimetres.

(c) Explain why Alana is incorrect.

(1)

Alana edits the data, swapping the values she realised had been recorded incorrectly.

(d) How would you expect the PMCC of her corrected data to compare to the PMCC of the original data, r = 0.988?

(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 2 continued

Alana states that it would be impossible for there to be a negative correlation between the **minimum desired female height** and **maximum desired female height** for all breeds of show dog.

(e) Discuss Alana's statement.

You should include reasoned comments on whether or not you believe she is correct.

(4)

(Total for Question 2 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

3 In 2011 Boston College psychologists, Angelina Hawley-Dolan and Ellen Winner, investigated whether paintings of professional abstract artists could be differentiated from those of a five-year-old child.

Leo, a psychology student, decides to do his own version of this experiment.

Suggest how Leo might set up his experiment.

You should include comments about

- how he will source the data
- the data he should collect
- how he could minimise bias.

(7)

(Total for Question 3 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
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

4 Shai has created a video game. She makes her game free to play, with players able to buy gems to spend in the game for bonuses.

She creates a model to predict the income that she can expect from her game.

Shai's data shows that the majority of players do not spend any money on gems.

A small number of players, known as 'whales', make up the majority of the revenue for games by buying the gems.

[Source: https://deltadna.com/blog/how-whales-spend/]

(a) Give a reason why the normal distribution is unlikely to be a suitable model for the amount a randomly selected player spends in a given month.

(1)

Shai instead chooses to model the spend, in pounds, of a randomly selected whale in a given month. She uses the distribution $W \sim N(50, 100)$

She assumes that the amount a whale spends each month is independent from their spend in previous months.

(b) Find $P(W < 40)$

(1)

(c) Using Shai's model, calculate the probability that a randomly selected whale has a mean spend of less than £40 a month over a year.

(2)

(d) Explain why the probability calculated in part (c) **must** be smaller than the probability calculated in part (b).

(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 4 continued

(e) Explain why $P(W \leqslant 40)$ and $P(W < 40)$ are equal.

(1)

(f) Calculate the probability that a randomly selected whale spends **more** than £40 in each of four consecutive months.

(2)

(g) Calculate the probability that a randomly selected whale spends more than £40 in at least 8 of the months in a one-year period.

(3)

(h) Calculate the probability that a randomly selected whale spends less than £40 in a month that they spent more than £30

(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 4 continued

In Shai's game, a bundle of gems costs £5. The game also offers unlimited gems for any player spending £60 a month.

(i) Discuss the appropriateness or otherwise of the normal distribution to model the spend of a whale in light of this information.

(3)

(Total for Question 4 is 17 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

5 Tom owns a busy cinema with two screens. Screen 1 has 120 seats, and Screen 2 has 300 seats. He shows a different film on each screen. Children and pensioners are eligible for a concession ticket that is cheaper than a standard ticket.

From past experience Tom estimates that the proportion of concession tickets sold for the most popular 7 pm showing for the film on Screen 1 is 35% and the proportion for the film on Screen 2 is 15%.

Both screens show films at 7 pm one night.

The random variable X denotes the number of concession tickets purchased for the 7 pm showing of the film on **Screen 1**.

The random variable Y denotes the number of concession tickets purchased for the 7 pm showing of the film on **Screen 2**.

The random variable T denotes the **total** number of concessions tickets purchased for **both** showings of films at 7 pm that night.

Tom assumes that all the tickets will be purchased, and that the probability of a concession ticket being purchased is independent from one customer to the next.

(a) By modelling each of X and Y as an independent binomial distribution,

(i) show that the distribution of T can be approximated as N(87, 65.55)

(4)

(ii) estimate the probability that at least 100 concessions tickets are purchased in total.

(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 5 continued

(b) Assuming that it is valid to model X and Y as binomial distributions, give **two** reasons to justify the use of the **approximation** in part (a).

(2)

(c) Give **two** reasons in context why the assumptions that Tom made may **not** be justified.

(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

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

Tom charges £4 for a concession ticket, and £6 for a regular ticket. It costs him £200, in total, to run the cinema for the 7pm showing.

(d) Assuming that all tickets are sold for both screens,

(i) write an expression for the number of regular tickets sold in terms of T

(1)

(ii) find the probability Tom earns at least £2 125 of profit.

Note that profit is total income from sales minus costs.

(5)

(Total for Question 5 is 16 marks)

6 Caroline runs a hairdressing salon in London. She reads the following article.

How much hair loss is normal in a day

According to the American Academy of Dermatologists, an adult will normally lose anywhere from 50 to 100 strands of hair per day. For people with longer hair strands, losing them may be more noticeable. Since there are 100,000 hair follicles – or more – on each person's scalp, the loss of 100 or so hair strands a day doesn't make a big difference in appearance.

Average hair loss in men and women

Women tend to lose more hair strands per day than men. There's no way to measure the difference objectively, because daily heat styling and frequent hair colouring plays a big part in how much of your hair sheds. Approximately 40 percent of women lose extra hair every day because of the way they style it.

What causes hair to fall out

Some daily hair loss is normal. Increased hair loss can be a result of stress or a health condition.

[Source: https://www.healthline.com/health/how-much-hair-loss-is-normal]

Caroline states that the random variable $X \sim B(100\,000, 0.001)$ could be used to model the number of hairs lost by a randomly selected customer during one day.

(a) Explain how Caroline chose her parameters for the distribution of X

(2)

(b) Comment on the assumptions underlying Caroline's model indicating whether or not these are reasonable.

Your answer should include **three** different statements.

(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 6 continued

(Total for Question 6 is 5 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

7 Animal Crossing is a video game in which players visit desert islands. A single animal lives on each desert island. A player visiting an island will always meet the animal that lives there. Each animal is either male or female.

Jonny plays a simplified version of Animal Crossing in which there are only four animal species. Data on the number of animals of each species is shown in Figure 5.

		Species				
		Alligator	**Anteater**	**Bear**	**Bird**	**Total**
Gender	**Male**	6	3	10	15	34
	Female	3	7	6	6	22
	Total	9	10	16	21	56

[Source: https://animalcrossing.fandom.com/wiki/Animal_Crossing_Wiki]

Figure 5

Jonny visits a desert island at random.

(a) Write down the probability that the animal he meets is a female anteater.

(1)

(b) Write down the probability that the animal he meets is a male, given that it is an alligator.

(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 7 continued

Jonny assumes that once he has visited a particular desert island, he will never visit that island again.

(c) Under Jonny's assumption, find the probability that when Jonny visits four islands, he meets two birds and two bears, given that all the animals he meets are either birds or bears.

(5)

In the full version of Animal Crossing there are **more** than four species of animal.

Jonny's assumption in part (c) is untrue, and it is possible to meet the same animal **multiple times**.

The data for alligators, anteaters, bears and birds in **Figure 5** are correct for both versions of the game.

(d) Explain **why**, in the full version of the game,

(i) the probability in part (a) would be smaller than the calculated value

(1)

(ii) the probability in part (b) would be unchanged

(1)

(iii) the probability in part (c) would be smaller than the calculated value.

(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 7 continued

In Animal Crossing players are able to buy and sell turnips each week.

The prices at which players can sell turnips during a week follow one of three patterns:

Spike Fluctuating Decreasing

The probability of the prices following a certain pattern one week depends only on the previous week's pattern, as shown in **Figure 6**.

		This week's pattern		
		Spike	**Fluctuating**	**Decreasing**
Previous week's pattern	**Spike**	0.3	0.5	0.2
	Fluctuating	0.65	0.2	0.15
	Decreasing	0.7	0.25	0.05

[Source: https://animalcrossing.fandom.com/wiki/Animal_Crossing_Wiki]

Figure 6

During the first week of May, the prices of Jonny's turnips are following the fluctuating pattern.

(e) Find the probability that, during the **third** week of May, the prices of Jonny's turnips are following the fluctuating pattern again.

(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 7 continued

During the first week of May, the prices of Jonny's turnips were following the fluctuating pattern.
During the third week of May, the prices of Jonny's turnips were following the fluctuating pattern again.

(f) Find the probability that during the **second** week of May, the prices of Jonny's turnips were following the spike pattern.

(4)

(Total for Question 7 is 16 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

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

BLANK PAGE

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA