

Write your name here

Surname

Other names

Pearson Edexcel
Level 3 GCE

Centre Number

--	--	--	--	--

Candidate Number

--	--	--	--	--

Statistics

Advanced Subsidiary
Paper 2

Friday 22 June 2018 – Morning
Time: 1 hour 30 minutes

Paper Reference

8ST0/02

You must have:

Statistical formulae and tables booklet
Calculator

Total Marks

--

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. 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). Coloured pencils and highlighter pens must not be used.
- **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, statistical tests should be carried out at the 5% significance level.
- When a calculator is used, the answer should be given to three significant figures unless otherwise stated.

Information

- A booklet 'Statistical formulae and tables' is provided.
- There are 7 questions in this question paper. The total mark for this paper is 60.
- 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 ►

P58390A

©2018 Pearson Education Ltd.

1/1/1/



Pearson

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

1

Random	Systematic	Experimental	Unbiased
Snowball	Simple	Judgmental	Disproportional
Unrestricted	Census	Sampling	Non-random
Biased	Cluster	Proportional	Stratified

Figure 1: Words used to describe sampling techniques

Using any of the words in **Figure 1**, describe the following sampling techniques.

- (a) Debbie wants to send a questionnaire to a sample of people at her golf club. The club has 461 members, 129 of whom are female.

She sends questionnaires to 25 randomly selected male members and 25 randomly selected female members.

(2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

2



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(b) Graham has a list of 965 email addresses, which he has numbered 1 to 965.

He uses a spreadsheet program to generate 50 random numbers between 1 and 965, then he looks up the email addresses corresponding to these numbers. These email addresses form his sample.

(2)

(Total for Question 1 is 4 marks)



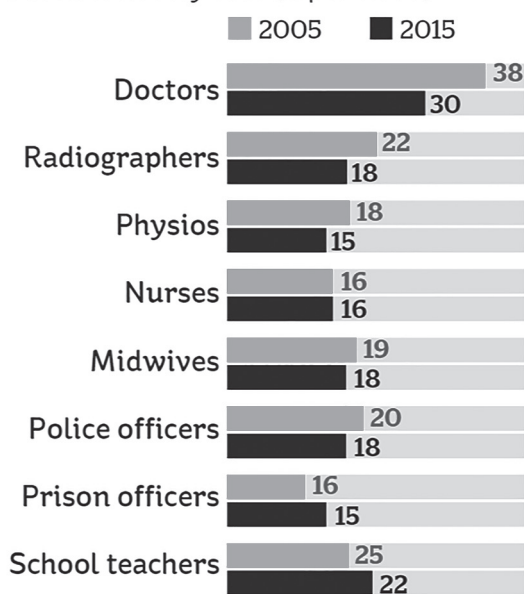
- 2 The diagrams in **Figure 2** are taken from an article in the *i newspaper* titled ‘May set to end seven-year cap on pay rises for public servants’.

The article describes how Theresa May, the Prime Minister of the UK (in September 2017), may be about to remove the seven-year cap on pay rises for public servants.

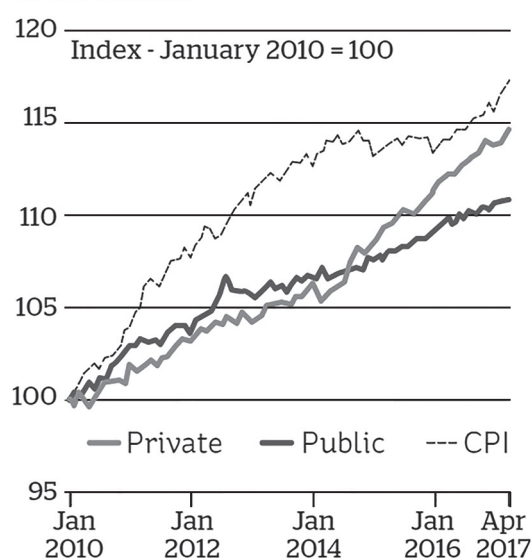
No further explanation is given for the diagrams.

How wages have changed

Median hourly rate (£ per hour)



Comparison with private sector pay and inflation



SOURCE: ONS

Source: The *i newspaper*, 5th September 2017

Figure 2

Explain, in the context of the question, how **Figure 2** may be misleading to a reader.

(4)



- 3 Bhupathy's purple frog is a newly discovered species of frog located in the Western Ghats mountain range in India.

Ramesh is studying the size of these frogs, and has recorded the body length of the 15 specimens he found. This data is presented in **Figure 3**.

Specimen	Sex	Body length (mm)
A	Male	60
B	Male	58
C	Female	57
D	Male	55
E	Male	62
F	Female	72
G	Male	67
H	Male	53
I	Female	77
J	Male	55
K	Male	67
L	Female	86
M	Male	54
N	Male	55
O	Male	56

Adapted from research data

Figure 3: Measurements of Ramesh's sample of Bhupathy's purple frogs

Ramesh believes that the female Bhupathy's purple frog has a longer body, on average, than the male.

- (a) Use the Wilcoxon rank-sum test to investigate Ramesh's belief.

(9)



Question 3 continued

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

One of Ramesh’s colleagues, Gill, points out that the sample is very imbalanced between the sexes. She believes that there may be a biological reason for this.

- (b) Find the probability that a random sample of 15 Bhupathy’s purple frogs would have this level of sex imbalance **or worse**.

You may assume that the population of Bhupathy’s purple frogs is half male and half female.

(3)

- (c) Do you agree with Gill’s belief? Explain your answer.

(1)

(Total for Question 3 is 13 marks)



- 4 A Hertzsprung-Russell diagram is a scatter diagram showing the relationship between a star's colour and its absolute magnitude (a measure of the brightness of the star).

The colour is measured on a scale called the 'Blue minus visible colour index' (B-V index).

Mina is researching star colours, and has found a Hertzsprung-Russell diagram in a textbook. It includes all stars less than 3 parsecs (an astronomical unit of distance) from the Sun. This diagram is presented in **Figure 4**.

The 'absolute magnitude' axis is presented with numbers **in reverse order**.

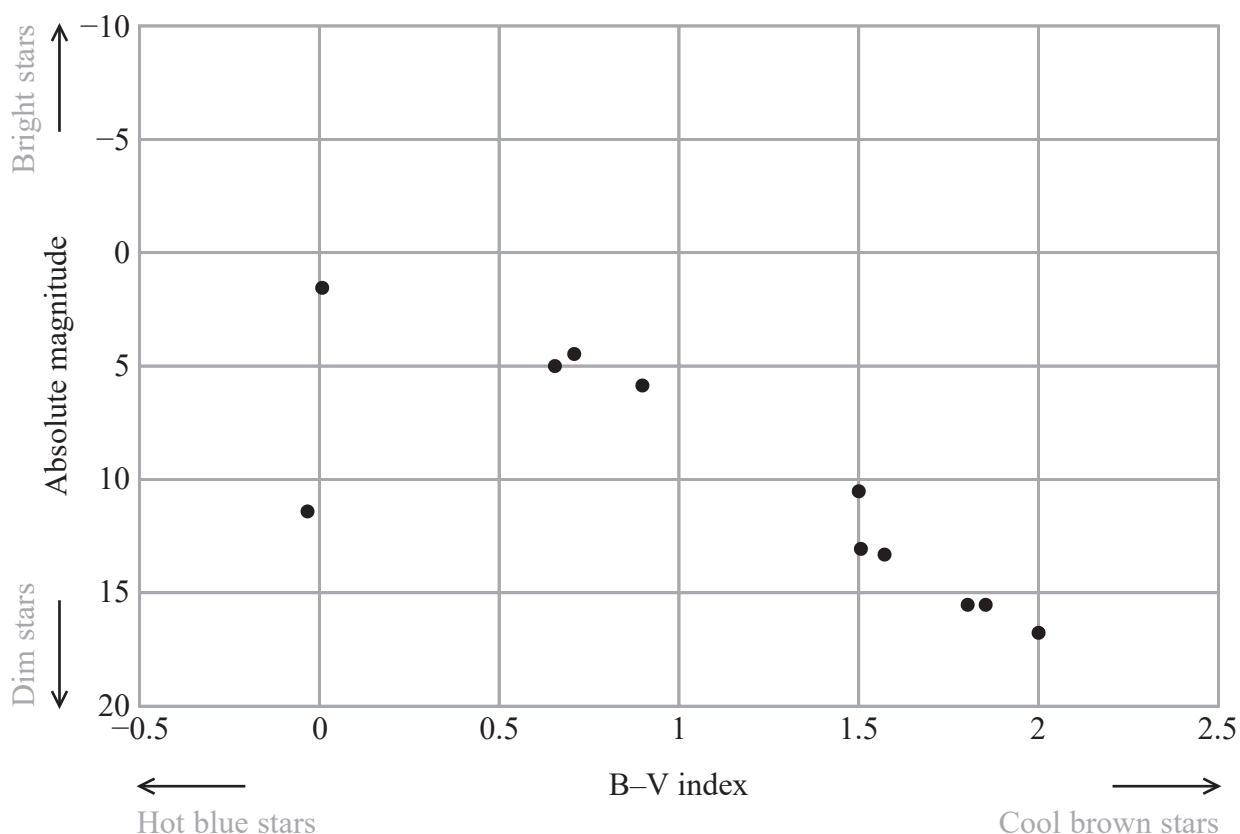


Figure 4: Hertzsprung-Russell diagram of stars less than 3 parsecs from the Sun



Mina believes that the relationship between B–V index and absolute magnitude may be linear.

Using the data given in **Figure 5**, she decides to calculate a least-squares regression line to predict absolute magnitude from B–V index.

Star name	Distance from the Sun (parsecs)	B–V index	Absolute magnitude
Sol	0	0.656	4.9
Proxima Centauri	1.2959	1.907	15.4
Alpha Centauri A	1.3248	0.71	4.4
Alpha Centauri B	1.3248	0.9	5.7
Barnard’s Star	1.8238	1.57	13.2
Wolf 359	2.3906	2	16.6
Lalande 21185	2.5469	1.502	10.5
BL Ceti	2.6267	1.85	15.5
UV Ceti	2.6267		15.6
Sirius A	2.6371	0.009	1.5
Sirius B	2.6371	-0.03	11.3
Ross 154	2.9698	1.51	13.0

Data source: The HYG Database

Figure 5: Data for stars less than 3 parsecs from the Sun

She decides to remove the data for the star with no recorded B–V index. She also decides to remove the data for the obvious outlier, assuming it to be an error in the data collection.

- (a) Calculate the equation of the least-squares regression line of absolute magnitude on B–V index, using Mina’s remaining data.

Give your equation with coefficients correct to 3 decimal places.

(3)



Question 4 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

- (b) Assuming that this sample is representative of all stars in the galaxy, use this equation to estimate the mean absolute magnitude of stars in the galaxy with a B–V index of 1.2 (2)

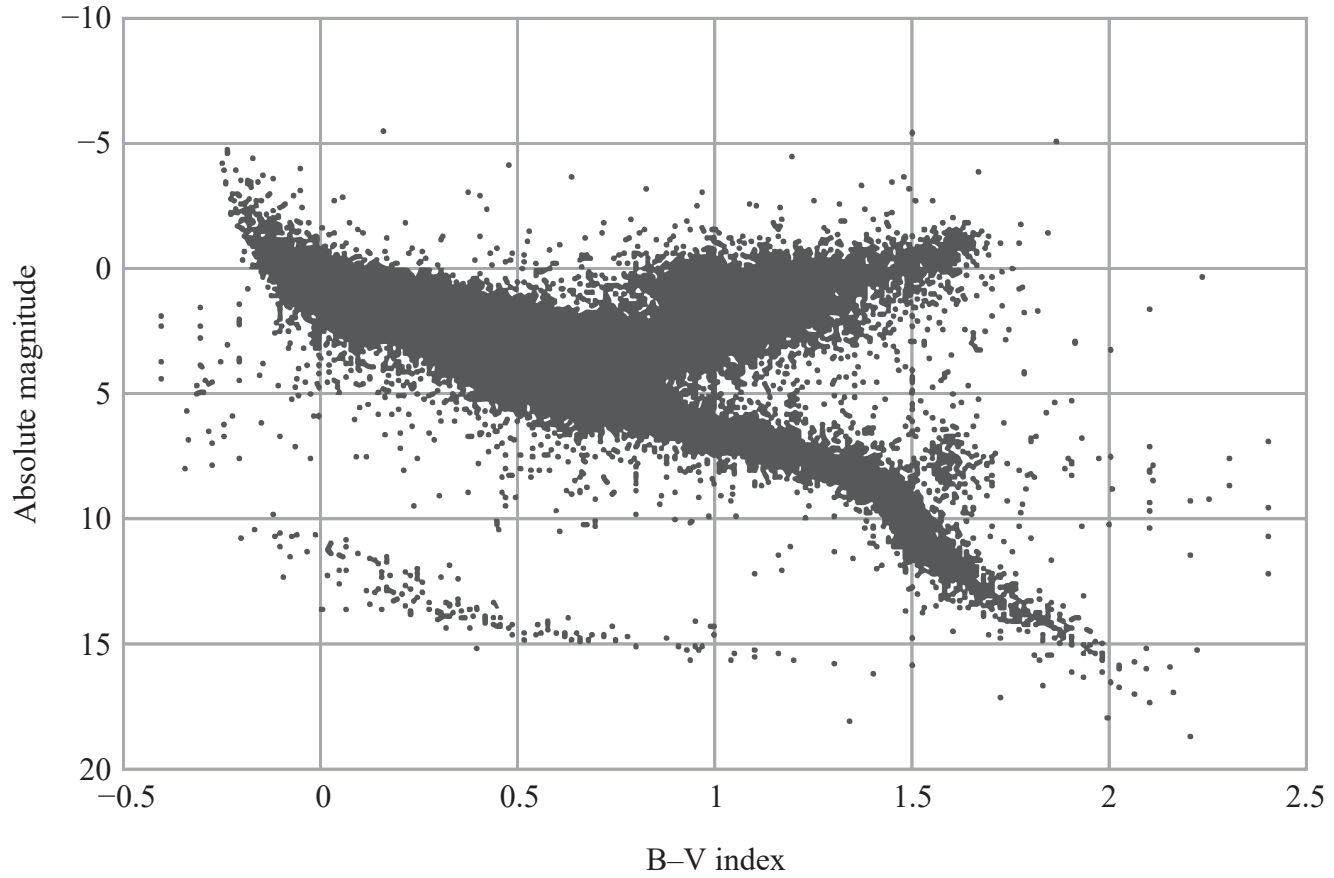


DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

At a later stage, Mina does some research online and finds a database with a much larger sample of stars. She then uses graph-plotting software to produce the Hertzsprung-Russell diagram in **Figure 6**.



Data source: The HYG Database

Figure 6: Hertzsprung-Russell diagram of stars less than 200 parsecs from the Sun

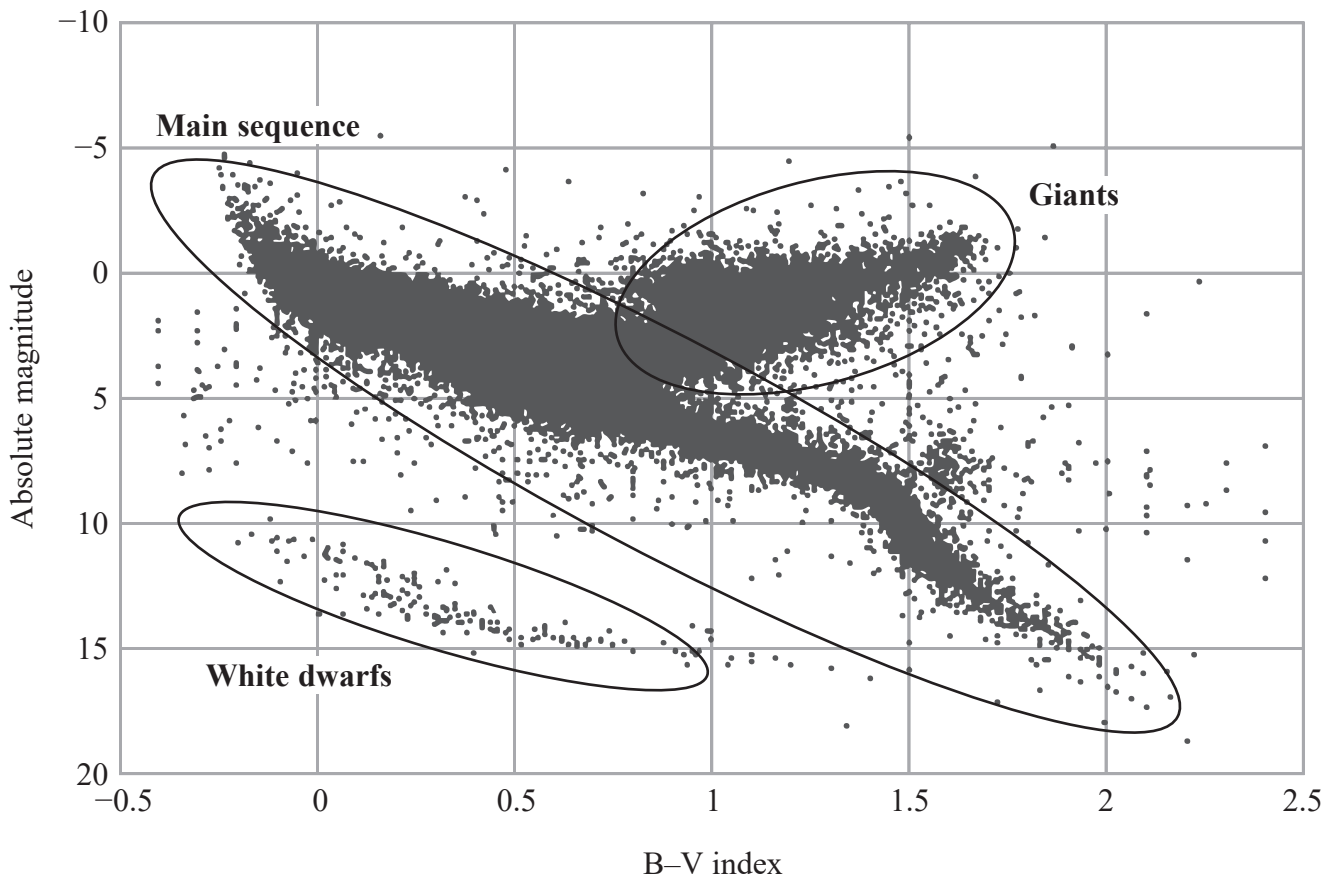
(c) Explain why your answer to part (b) may not be reliable.

(1)



Mina's research leads her to find that the strange shape in the distribution is caused by the existence of several different types of star.

These are labelled in **Figure 7**.



Data source: The HYG Database

Figure 7: Hertzsprung-Russell diagram with star types labelled

(d) Make suggestions to Mina to improve her model for predicting the mean absolute magnitude from B-V index.

(2)



Question 4 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 4 is 8 marks)



- 5 The **Freedom of Information Act (2000)** is an Act of the UK Parliament that ensures members of the public have the right to access information held by public authorities.

The government receives many thousands of these freedom-of-information (FOI) requests every year. Some of these are routine requests that are handled quickly. Others are non-routine, and can take extra time to process.

The government aims to respond to any FOI request within 20 days.

Figure 8 shows the timeliness of responses to non-routine FOI requests from 1 Jan–31 March 2017.

These requests are further categorised into requests to a ‘Department of State’, and requests to ‘Other monitored bodies’.

	Total requests received (excluding on-hold and lapsed)	Timeliness of response		
		20-day deadline met	Permitted extension to deadline	Late response (i.e. deadline missed)
TOTAL	12 308	10 749	626	933
Departments of State	8 387	7 195	430	762
Other monitored bodies	3 921	3 554	196	171

Source: Gov.uk

Figure 8: Timeliness of response to non-routine information requests received by monitored bodies from 1 Jan–31 March 2017



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

For a non-routine request randomly selected from 1 Jan – 31 March 2017, the events S and L are defined as follows:

S : Request was directed to a ‘Department of State’

L : Request had a late response

(a) Find

(i) $P(S)$ (1)

(ii) $P(L)$ (2)

(b) Construct a Venn diagram to show the **frequency** data for events S and L using the data in **Figure 8**. (3)



(c) Find $P(S|L')$

(2)

(d) Are the events S and L' statistically independent? Explain your answer.

(2)

(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

BLANK PAGE



P 5 8 3 9 0 A 0 1 9 2 8

- 6 Imhotep is investigating the number of insurance claims made by car drivers and motorcycle riders in a given year.

He contacts a motorcycle insurance company that sends him the summary table presented in **Figure 9**.

Number of claims	Proportion of customers (correct to 4 decimal places)
0	0.8127
1	0.1573
2	0.0248
3	0.0051
4	0.0001
5+	0.0000

Figure 9: Number of claims made by motorcycle insurance customers in 2016

(a) Calculate

- (i) the expected number of claims per motorcycle insurance customer in 2016, (1)

- (ii) the variance of the number of claims per motorcycle insurance customer in 2016. (2)



The data from the motoring website that Imhotep used was presented in the following form on the website.

Feedback from 2016 questionnaire

Based on sample of 139 responses

⋮

Question 18

How many car insurance claims have you made in the last 12 months?

Average of responses = 0.146

Standard deviation of responses = 0.486

(c) Give three reasons why your conclusions drawn in (b) might not be reliable.

(3)

(Total for Question 6 is 9 marks)



- 7 Stock markets are often affected by political events. On 23/06/2016, the UK voted to decide whether to leave the European Union, with the result declared in the early hours of 24/06/2016.

A random sample of 10 companies from the London Stock Exchange (LSE) is presented in **Figure 10**, with the share prices of each company for these dates.

The distribution of company share prices is known to be asymmetric (not symmetrical).

Company name	Symbol	23/06/2016		24/06/2016	
		Open	Close	Open	Close
Vodafone Group Plc	VOD.L	218.2	217.9	203.25	219.3
Worldpay Group plc	WPG.L	291.8	294.7	280	280.3
Sky plc	SKY.L	888.5	893.5	770	834.5
TUI AG	TUI.L	1039	1042	847.5	954
St. James's Place plc	STJ.L	919.5	924	680	774.5
Direct Line Insurance Group PLC	DLG.L	374.5	374.1	350	351.2
SSE plc	SSE.L	1550	1550	1375	1420
Schroders plc	SDR.L	2660	2711	2173	2382
Smiths Group plc	SMIN.L	1120	1120	1028	1086
Royal Dutch Shell plc	RDSA.L	1823	1852	1752.5	1872

Data source: Yahoo Finance

Figure 10: Share prices for 10 LSE companies on 23-24 June 2016

You are in charge of a small team of journalists investigating this data. One of your team, Nadia, tests the data and finds that the drop in the average share price between 'Close' on 23/06/2016 and 'Open' on 24/06/2016 is statistically significant.

Nadia suspects that the average share price was still lower at 'Close' on 24/06/2016 than at 'Close' on 23/06/2016.

- (a) Use the sign test to investigate Nadia's suspicion.

Present your conclusion, together with Nadia's findings, in a short report for publication in a broadsheet newspaper.

(10)



(b) You get permission to spend more time analysing the data. Suggest two ways you could extend your analysis in order to further investigate stock prices after the vote. (2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 7 is 12 marks)

TOTAL FOR PAPER IS 60 MARKS



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

BLANK PAGE



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

Every effort has been made to contact copyright holders to obtain their permission for the use of copyright material. Pearson Education Ltd. will, if notified, be happy to rectify any errors or omissions and include any such rectifications in future editions.

