Please check the examination detail	ls below before ent	ering your candidate information
Candidate surname		Other names
Pearson Edexcel Level 3 GCE	Centre Number	Candidate Number
Thursday 20 J	une 20	019
Morning (Time: 1 hour 30 minutes	s) Paper R	deference 8ST0/02
Statistics Advanced Subsidiary Paper 2	/	
You must have: Statistical Formulae and Tables bo Calculator	ooklet	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).
- **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 8 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





	Answer ALL questions. Write your answers in the spaces provided.		
1	You have been asked by your local district council to conduct a survey of local businesses.		
	You need to generate a stratified sample of businesses that is representative of all businesses located in the district.		
	Suggest two possible stratifying factors for your sample.		
	For each factor, suggest a set of suitable strata.		
		(3)	
_			

Question 1 continued	
	(Total for Question 1 is 3 marks)



- 2 For each of the following random variables, decide which of the following apply:
  - The binomial distribution is a **suitable** model. In this case, you should write down the value of *n* and *p* (if known).
  - The suitability of the binomial distribution as a model is **dependent** on some additional factor(s).

    In this case, you should state what factor(s) the suitability depends on.
  - The binomial distribution is **unsuitable** as a model.

In this case, you should explain why it is unsuitable.

(a)  $X = \text{total number of questions answered correctly, when the answer to each of 10 multiple-choice questions is randomly guessed.$ 

10 multiple-choice questions is randomly guessed.	(2)
(b) A raffle sells 300 tickets and has 20 prizes.	
Prize tickets are drawn one by one randomly from a large bowl and then discarded.	
Y = number of prizes won by a random sample of 12 tickets.	(8)
	(2)

(Total for Question 2 is 4 marks)





(1)

3 Benford's law is a statistical observation about the **first digits** of numbers in many real-life datasets.

One such dataset that obeys Benford's law is the set of Surface Areas of British Lakes (in hectares).

If X represents the **first digit** of the surface area of a randomly selected British lake, then Benford's law states that X will have the probability distribution given in **Figure 1**.

So, for example, the probability that a randomly chosen British lake has a surface area with first digit 8 (such as 81 hectares or 824.5 hectares) is approximately 0.05

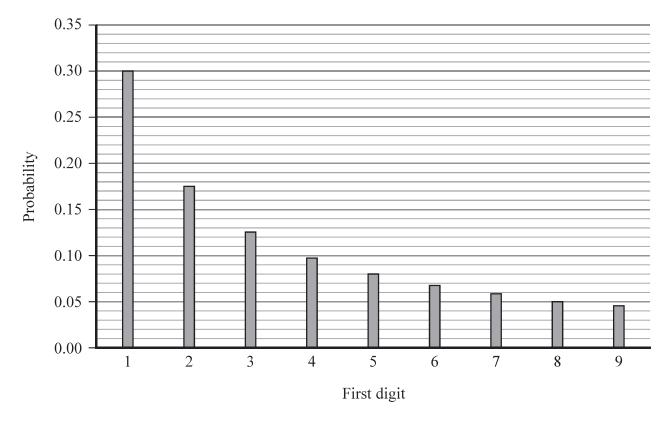


Figure 1: Probability mass function for Benford's law

(a) If a British lake is chosen at random, **estimate** the probability that the first digit of its surface area is greater than 7.

Question 3 continued	
(b) If a random sample of 5 British lakes is chosen, <b>estimate</b> the probability that at least one of the first digits of their surface areas is greater than 7.	(2)
	(2)
(c) Using <b>estimated</b> probabilities from <b>Figure 1</b> , estimate the value of $E(X)$ .	(3)



Question 3 continued			
(d) Explain what $E(X)$ represents in the context of the question.	(2)		
(Total for Question 3 is 8 mark	xs)		



4 A sample taken from a normally-distributed data population produces the following summary statistics on a calculator.

_	
×	=8.0848
$\Sigma \times$	=404.24
$\Sigma \times^2$	=3420.593
S²X	=3.110064
SX	=1.763537
n	=50
min(x)	=4.083804
Q1	=6.885888
Med	=8.093566
03	=9.126168
max(x)	=12.08374

Stating your hypotheses, test whether the population mean equals 8.5

(	6	)	

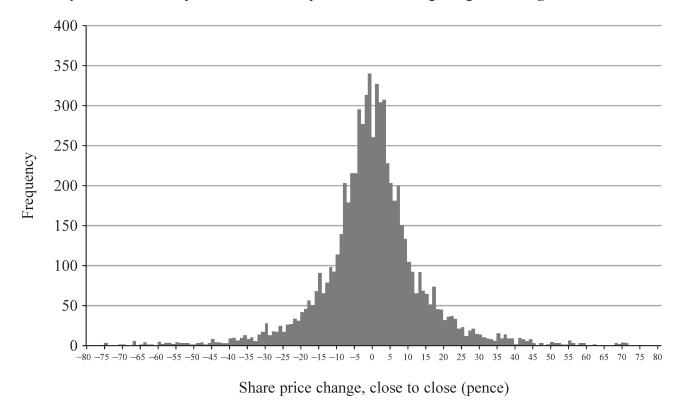
·	
[ ·	

Question 4 continued	
(To	tal for Question 4 is 6 marks)



5 Borghild is studying the changes in Pearson PLC's share price on the London Stock Exchange.

She has calculated the daily **change** in its share price for each weekday from 1st July 1988 to 30th September 2017, and produced the histogram given in **Figure 2**.



Source from: https://uk.finance.yahoo.com/

Figure 2: Histogram of daily changes to Pearson PLC's share price

Borghild models this data as a normal distribution with the following parameters:

$$\mu = 0.03$$
  $\sigma = 8.73$ 

Question 5 continued			
(a) Using Borghild's model, estimate the probability that on a randomly chosen day in the date range,			
(i) Pearson PLC's share price did not change by more than 5p from the day before,	(1)		
(ii) Pearson PLC's share price increased in value by more than 35p from the day before	ore. (1)		
	(1)		



Question 5 continued	
(b) Using your answers to part (a) and <b>Figure 2</b> , sta Borghild's model.	ate one strength and one weakness of
_ 0.8 0 0	(2)
	(Total for Question 5 is 4 marks)



**6** Rebecca works in the Maths and Statistics Faculty of a prestigious university. She is in charge of managing the application process for Maths and Statistics degree courses.

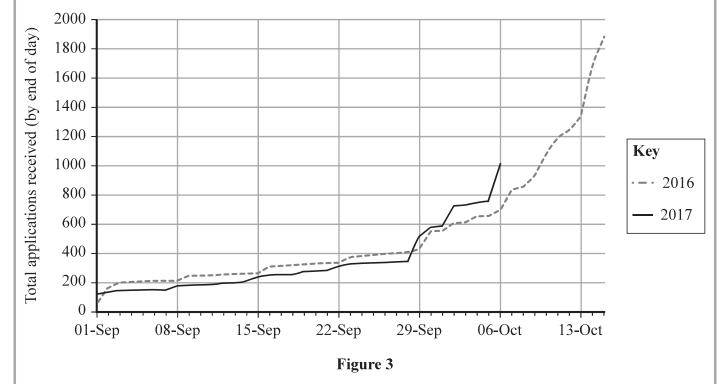
Applications open each year on the 1st of September and close on the 15th of October.

Rebecca keeps a daily record of the number of applications received.

On the 7th of October 2017, she used spreadsheet software to produce the graph shown in **Figure 3**.

**Figure 3** illustrates the **total** number of applications received since the start of the application period, up to the end of each day in the application period.

**Figure 3** shows data for the whole 2016 application period, and for the 2017 period up to the 6th of October 2017.



(a) Approximately how many applications did the university receive **on** the 29th September 2017?

(2)

(b) How many more days did it take in 2016 for the total applications to reach 1000 than it did in 2017?

(1)



#### **Question 6 continued**

The head of faculty asks Rebecca to estimate the total number of applications the faculty will receive by the end of the application period.

Rebecca completes the following calculations:

By end of 6th October,

2016 applications = 700

2017 applications = 1025

$$\frac{1025}{700}$$
 = 1.4643 (4 d.p.)

≈ 46 % increase

Total applications in 2016 = 1885

Estimate of total applications in 2017

= 1885 × 1.4643 = 2760 (0 d.p.)

(c) Explain, with reference to Figure 3, why Rebecca's estimation may not be reliable.

Suggest how the actual 'total applications' in 2017 may differ from her estimate.

(3)

(Total for Question 6 is 6 marks)

7 A journalist presents a story to the editor stating that:

'District councils in England are having to cut costs to ever greater levels. From the tax year 2015-16 to the tax year 2016-17, the expenditure on employee salaries, benefits and expenses has reduced by 6%, on average.'

The journalist is unable to quote a reliable source, and the editor suspects that the 'reduced' figure of 6% is too high. The editor has very limited statistical knowledge, so puts Diane, who has studied some statistics, in charge of investigating these figures.

Diane is unable to find complete data for the whole of England on district council expenditure. She uses data from a sample of district councils whose accounts she has found online.

Diane's collected data is presented in the table in **Figure 4**.

District Council	15-16 employee expenditure (£000)	16-17 employee expenditure (£000)	Percentage change (%)
A	25 059	24328	-2.92
В	20 967	18263	-12.90
С	195 953	204 564	4.39
D	11 367	11 359	-0.07
Е	32 706	33 361	2.00
F	19825	18852	-4.91
G	10825	10925	0.92
Н	9 2 0 2	8300	-9.80
I	28 633	27961	-2.35

(Data source: Data from nine district council accounts)

Figure 4: Employee expenditure for Diane's sample of district councils

(a) Write down the employee expenditure of District Council E in 2016-17.

**(1)** 

Question 7 continued			
(b) Use a Wilcoxon signed-rank test to investigate the editor's suspicion.			
You should state one assumption that you need to make about the population in order for the test to be valid.			
Present your findings in a form suitable for the editor to understand.  (9)			



Question 7 continued				

uestion 7 co	ntinued	
	ater finds out that not all district councils store their accounts online, ble to the public.	
Explair	n why this might affect the validity of the test performed in (b).	(1)
	the month of May each year, all district councils are legally required to have ecounts available online, accessible to the public.	
	n how a spreadsheet program could be used during May to improve the ng method.	
•		(3)
	(Total for Question 7 is 14 ma	rks)



**8** Andrew is a sports medicine researcher who is investigating correlation between different types of physical aptitude.

He takes a random sample of female gym members and assesses their physical aptitudes using the following activities.

#### Beep test

This records the number of successful runs between two points 20 metres apart. The next point must be reached before the beep for the run to be considered successful. Each successful run is called a lap.

Level 1 is at the lowest beep speed. Each level has 10 laps at the same speed.

The speed levels get gradually faster for each level up to Level 10.

Andrew records, for each gym member, the **highest level reached**, along with the number of **completed laps at that level**.

### Maximum leg press

Andrew records, for each gym member, the maximum mass (in kg) that their legs can push.

Andrew's recorded results are presented in **Figure 5**.

	Beep test		Marinanna	
Member	Highest level reached	Completed laps at highest level	Maximum leg press (kg)	
A	7	0	145	
В	4	6	160	
С	9	5	222	
D	1	8	114	
Е	3	6	291	
F	6	1	143	
G	9	1	81	
Н	2	6	92	
I	3	0	61	
J	2	9	122	
K	8	9	233	

(Data source: Adapted from research data)

Figure 5



a) For Andrew's data in the form presented in <b>Figure 5</b> ,		
(i) explain why this data cannot be used to investigate the correlation between the results of the beep test and the maximum leg press using Pearson's product-moment correlation coefficient,		
	(2)	
(ii) suggest two possible solutions to this problem.		
(ii) suggest two possible solutions to this problem.	(2)	



# **Question 8 continued**

two tests for Andrew's sample.

	Beep test		Marinanna lag	
Member	Highest level reached	Completed laps at highest level	Maximum leg press (kg)	
A	7	0	145	
В	4	6	160	
С	9	5	222	
D	1	8	114	
Е	3	6	291	
F	6	1	143	
G	9	1	81	
Н	2	6	92	
I	3	0	61	
J	2	9	122	
K	8	9	233	

(Data source: Adapted from research data)

Figure 5

(b) Making any necessary assumptions, use one of your suggestions in (a)(ii) that **does not** involve discarding data to investigate for an association between the results of the

(7)

Question 8 continued				



Quest	ion 8 continued				
(c)	Andrew identifies two potential outliers in his dataset.				
	He interviews the corresponding gym members at athletic interest.	He interviews the corresponding gym members and asks each to state their main athletic interest.			
	Explain, with reasons in context, which member you believe described herself as:				
	(i) an indoor sprint cyclist,				
			(2)		
	(ii) a marathon runner.		(2)		
			(2)		
	(Total for Question 8 is 15 marks		arks)		
_	7	TOTAL FOR PAPER IS 60 MA	RKS		
	•	COLUMN TO THE EAST OF THE STATE			





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.

