Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number	Candidate Number

Monday 22 June 2020

Afternoon (Time: 2 hours) | Paper Reference **9ST0/03**

Statistics

Advanced

Paper 3: Statistics in Practice

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).
- **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 6 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 ▶

P62707A

©2020 Pearson Education Ltd.

1/1/1/1/1/

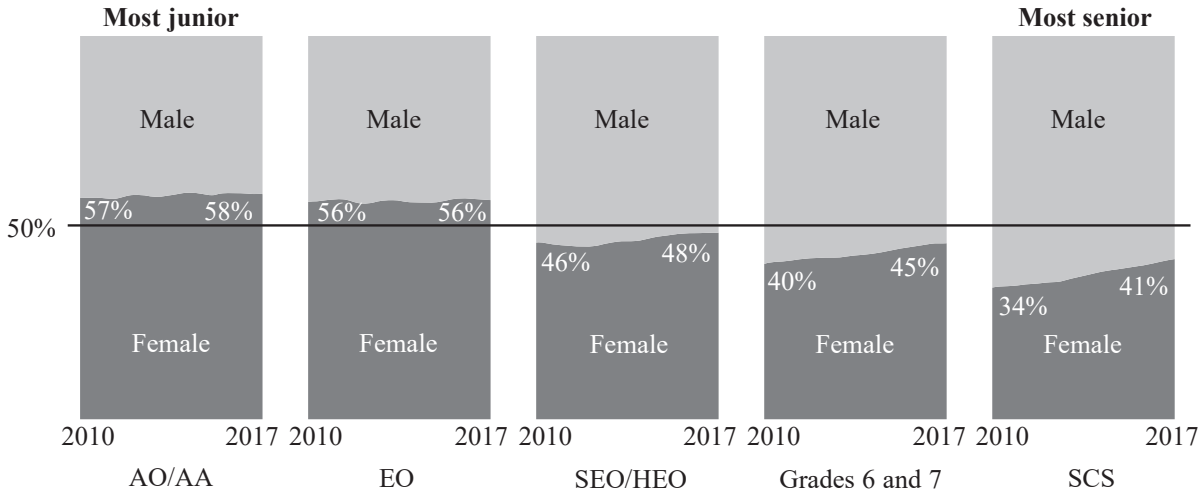
P 6 2 7 0 7 A 0 1 2 8

Pearson

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

1 Teru is looking at some graphs from an official UK government publication that show patterns of employment for civil servants (central government employees) over time.

Figure 1 shows the gender balance of each civil service grade between 2010–2017.



(Source: https://www.instituteforgovernment.org.uk/publications/whitehall-monitor-2018)

Figure 1

Teru says that **Figure 1** shows that the total number of female civil servants has definitely increased between 2010 and 2017.

(a) Explain whether Teru is correct or not.

(2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 1 continued

(b) Describe **two** important trends shown in **Figure 1**.

(2)

DO NOT WRITE IN THIS AREA

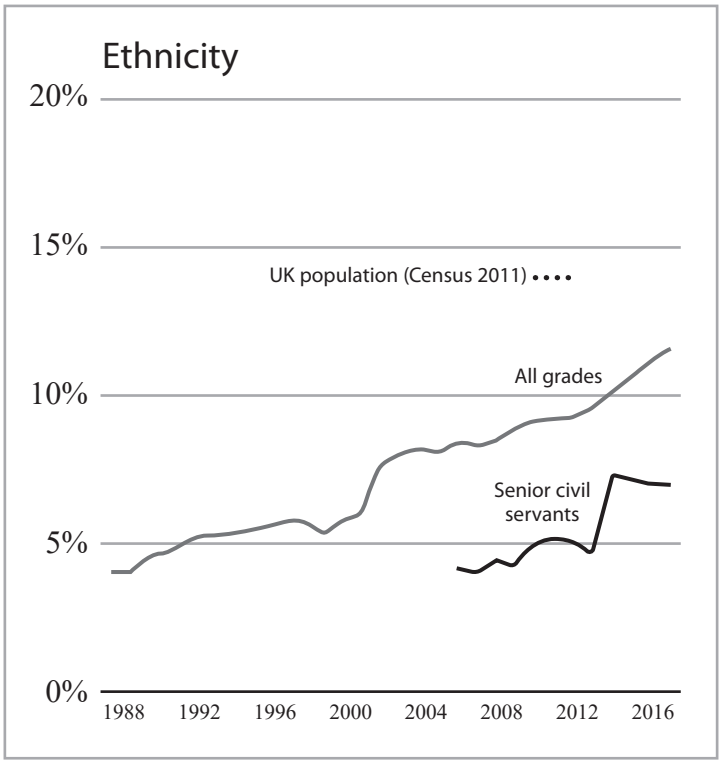
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 1 continued

Teru has another graph, **Figure 2**, which is taken from the same official UK government publication. It shows the **percentage** of civil servants from an **ethnic minority** for 1988 to 2017.

Figure 2 also shows the percentage of **senior** civil servants (SCS) from an ethnic minority between 2006 and 2017. These were the only years for which records were available for SCS.



(Source: https://www.instituteforgovernment.org.uk/publications/whitehall-monitor-2018)

Figure 2

One additional fact, **A**, is also shown on **Figure 2** by a short dotted line.
This is:

A The percentage of the UK population from an ethnic minority in 2011 taken from the UK Census.

(c) Give **one** reason why the additional fact **A**, included on **Figure 2**, provides useful information.

(1)

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

Question 1 continued

(d) Write a short report for a news website on the pattern of employment of civil servants from an ethnic minority, based on the information in **Figure 2**.

Make **three** distinct comments in an appropriate style.

(4)

(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

2 Snowy tree crickets are insects that live in various states of the USA. Male snowy tree crickets make a noise called a chirp by rubbing their wings together.

Thomas made a study of snowy tree crickets living in the wild in California. When he encountered a snowy tree cricket, he recorded the rate of the cricket's chirp (X chirps per minute) and the air temperature (Y °C). His results are shown in **Figure 3**.

Snowy tree cricket	Rate of chirp x	Air temperature y
A	50	9.2
B	50	9.4
C	110	15.6
D	127	17.8
E	129	18.3
F	133	18.6
G	154	19.7
H	159	19.7

(Data source: Annals of the Entomological Society of America, 55(3), 303–322.)

Figure 3

(a) Use the data in **Figure 3** to calculate Pearson's product moment correlation coefficient for rate of a snowy tree cricket's chirp and air temperature.

(1)

(b) Conduct an appropriate hypothesis test using the value you calculated in part (a).

You should make any necessary assumptions.

(4)

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

Question 2 continued

It has been suggested that the rate of a snowy tree cricket's chirp can be used as a reliable means of estimating air temperature.

(c) Comment on this suggestion with reference to your answers in parts (a) and (b).

(2)

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

Question 2 continued

Juliana entered the data in **Figure 3** in a spreadsheet and calculated the following least squares regression line for air temperature (Y °C) on rate of snowy tree cricket's chirp (X chirps per minute) in the form $y = a + bx$.

$$y = 4.44 + 0.102x$$

(d) Interpret, in context, the values of a and b in Juliana's least squares regression line.

(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

A graph of the data in **Figure 3** with Juliana's least squares regression line is shown in **Figure 4.**

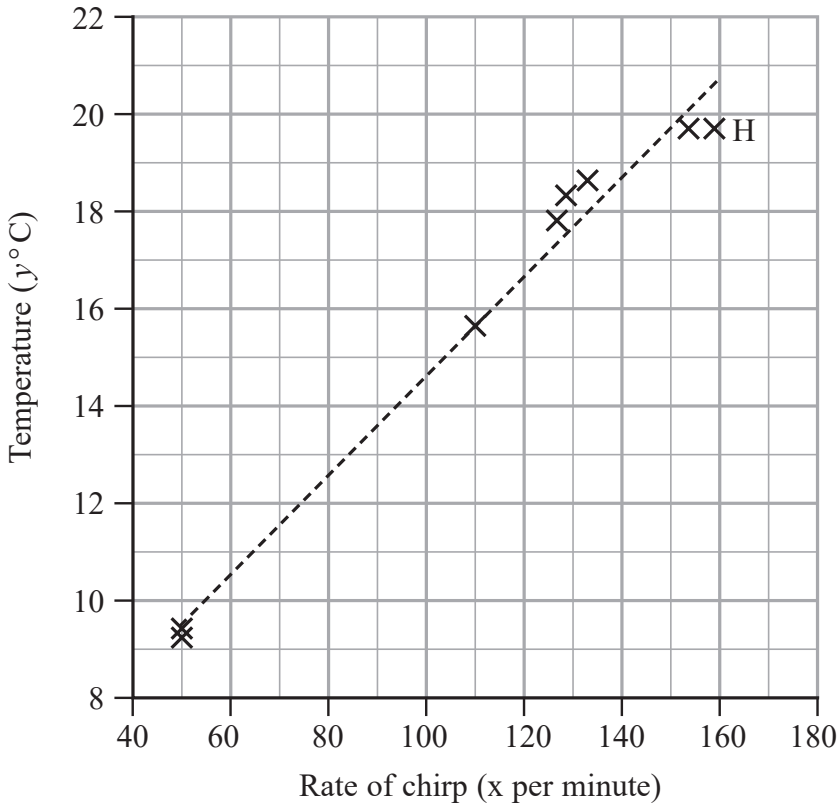


Figure 4

(e) **Calculate** the residual for cricket H.

Give your answer to **2** significant figures.

(2)

Question 2 continued

(f) Make a comment about the validity of using Juliana's regression line to estimate air temperature from the rate of a snowy tree cricket's chirp:

(i) with regard to the residuals of the points plotted in **Figure 4**,

(1)

(ii) with regard to any other piece of evidence.

(1)

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

Question 2 continued

(g) Make **one** suggestion for the next stage of the research.

Explain how your suggestion would improve the validity of the results.

(2)

(Total for Question 2 is 16 marks)

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

3 People buy and sell shares in companies to try to make money. Share prices can change daily and the daily percentage change of a share price can be calculated.

Chidi has bought some shares in the company BASF.

He has a sample of the daily percentage change of the BASF share price, summarised below:

Sample size	Mean	Standard deviation
251	–0.163%	1.37%

Using his sample and the null hypothesis H_0: $\mu = 0$, he performs a **parametric** hypothesis test to find out whether there is evidence that the BASF share price has a **negative** daily percentage change, on average.

(a) Complete Chidi's hypothesis test. You **do not** need to conclude in context.

(5)

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

Question 3 continued

Chidi's friend, Jason, discovers that Chidi's sample is, in fact, the daily percentage change of the BASF share price for **all** of the 251 days in 2018 on which shares were bought and sold.

Jason makes two criticisms of Chidi's decision to perform his hypothesis test on this data:

A Chidi did **not** need to do a hypothesis test to find out whether the average daily percentage change of the BASF share price was negative in 2018.

B Chidi cannot use the outcome of his hypothesis test to reliably make a judgement about the average daily percentage change of the BASF share price **in the future.**

(b) Explain why **each** of Jason's criticisms is correct.

(2)

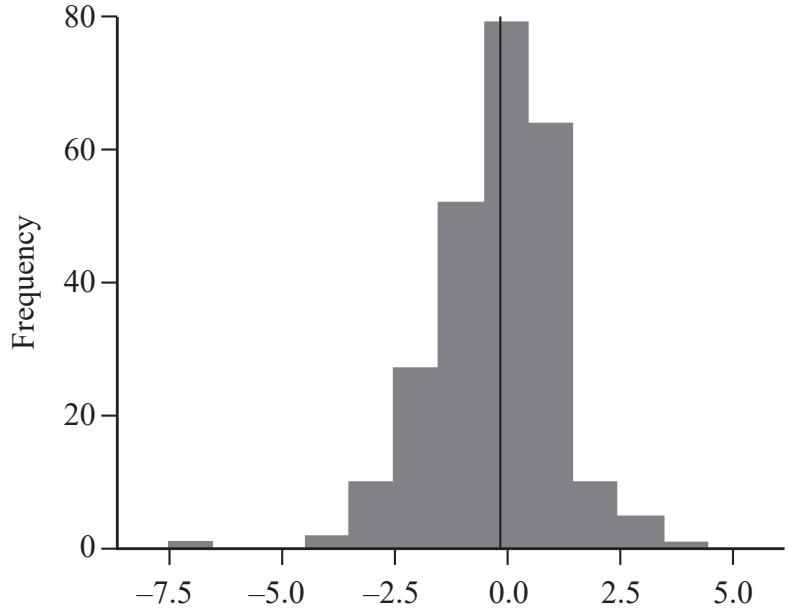
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 3 continued

Chidi looks at the daily price of BASF shares over several years in the past and decides that he will keep his shares.

He wants to apply a statistical model to his complete 2018 dataset of the daily percentage change of the BASF share price so that he can make comparisons between 2018 and other years.

He uses statistical software to plot a histogram of the daily percentage change of the BASF share price for 2018 and marks the mean value with a vertical line, **Figure 5**.



(Data source: https://www.basf.com)

Figure 5

As well as calculating the mean of –0.163% and the standard deviation of 1.37%, Chidi finds that:

183 of the 251 values in his dataset are within 1 standard deviation of the mean

241 of the 251 values in his dataset are within 2 standard deviations of the mean

(c) Comment on the validity of using the model N(–0.163, 1.37^2) for the daily percentage change of the BASF share price in 2018.

Use Chidi's calculations above and **Figure 5** to support your comments.

(4)

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

Question 3 continued

(Total for Question 3 is 11 marks)

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

4 Donal works in the head office of a large law firm. The law firm employs both qualified solicitors and 'agents' to resolve legal cases.

Each legal case is assigned to an agent for one month to find out if it can be resolved easily. If not, the legal case has to be assigned to a solicitor.

Donal is writing a report on the use of work targets for agents at the law firm.

To collect data from his colleagues, he used the following sampling method.

- Randomly select 3 of the law firm's 10 offices.
- Take a random sample of 20% of the agents from each of the 3 offices selected.

Donal visited the 3 offices selected to collect detailed information from the agents in his sample.

(a) Write down the name of Donal's sampling method.

(1)

(b) For Donal's sampling method write down

(i) **two** advantages,

(2)

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

Question 4 continued

(ii) **one** disadvantage.

(1)

At the start of the month each agent is assigned 40 legal cases.

From past experience, Maxine, a manager, claims that each agent resolves 40% of their legal cases by the end of the month.

(c) If Maxine's claim is correct, explain why a binomial model might be appropriate for the number of legal cases resolved by an agent by the end of the month.

(3)

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

Question 4 continued

Agents have the target of resolving at least 15 of their 40 cases by the end of the month.

Assume that the model B(40, 0.4) is appropriate for the number of cases resolved by an agent by the end of the month.

(d) Calculate the probability that an agent selected at random **does not** meet their target by the end of any given month.

(1)

Agents are assessed quarterly (every three months). Those agents that do not meet the target for **all three months** in the quarter are given a formal warning.

(e) Calculate the probability that an agent selected at random will be given a formal warning at the end of any given quarter.

(2)

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

Question 4 continued

Donal obtained records for the number of legal cases resolved by the end of the month for each of the 90 agents in his sample.

These are shown in **Figure 6.**

Number of resolved cases	Frequency	Probability		
$x \leqslant 11$	9	0.071		
$12 \leqslant x \leqslant 14$	26	0.246		
$15 \leqslant x \leqslant 17$	36			
$18 \leqslant x \leqslant 20$	17			
$x \geqslant 21$	2			
Total	90			

Figure 6

(f) Use a χ^2 goodness-of-fit test to investigate whether Donal's records imply that the model B(40, 0.4) is indeed appropriate for the number of legal cases resolved by an agent by the end of the month.

You may use the empty columns in **Figure 6** for your working, if you require them.

(10)

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

Question 4 continued

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

Question 4 continued

(g) Explain the meaning of a Type II error in the context of this question.

(2)

(Total for Question 4 is 22 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

5 A palaeontologist studies fossils. A "dig" is a site where palaeontologists try to find fossils.

Sharon is a palaeontologist interested in studying an extinct rhinoceros called a **teleoceras** that lived in North America in the Miocene era (about 23 million to 5 million years ago).

She has been offered two places on digs, one in California and the other in Montana. Both are new digs on sites dating from the Miocene era.

Sharon has obtained some information about previous Miocene era fossil digs in California and Montana. This is shown in **Figure 7**.

		Number of digs	Number of digs where **teleoceras** fossils were found
Location	California	947	12
	Montana	75	4

(Data source: https://paleobiodb.org/)

Figure 7

Using the information in **Figure 7**, carry out a hypothesis test of proportion.

Use the results of your hypothesis test to advise Sharon whether she should accept the place on the dig in California or the place on the dig in Montana.

(10)

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

Question 5 continued

(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

6 Betsan surveyed a sample of 22 knitters. 12 of the knitters said they regularly attended knitting groups where they shared their hobby with others. 10 said they did not.

The knitters were asked to rate **how calming** they found knitting on a scale from **0 'not at all'** to **100 'extremely'**.

Betsan decides to perform a Wilcoxon rank-sum test to find out whether knitters who attend knitting groups find knitting **more** calming, on average, than knitters who do not attend knitting groups.

Betsan's data are shown in **Figure 8** along with a partial ranking of the ratings.

Attends knitting groups		**Does not attend knitting groups**	
Calmness rating	Rank	Calmness rating	Rank
35	3	25	1
50	6	30	2
58	8	37	4
62	10	40	5
69	12	55	7
75	14	60	9
79	15	65	11
85	17	70	13
90		80	16
95		93	
98			
100			

Figure 8

(Data source: adapted from DOI:10.4276/030802213X13603244419077)

(a) Complete the ranking in **Figure 8**.

(1)

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

Question 6 continued

(b) Complete Betsan's Wilcoxon rank-sum test.

(7)

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

Question 6 continued

(c) State **two** reasons why Betsan's choice to use a Wilcoxon rank-sum test rather than the equivalent parametric hypothesis test might have been appropriate.

(2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 6 continued

Betsan also asked the knitters how happy knitting makes them feel on a rating of **0 'not at all'** to **100 'extremely'**.

She performed another Wilcoxon rank-sum test on the results to find out whether knitters who attend knitting groups find knitting makes them happier, on average, than knitters who do not attend knitting groups.

The result of her test was to reject H_0 at the 5% significance level.

Betsan claims that her test shows that knitting in a group **makes** people happier than knitting on your own.

(d) Suggest **two** other possible explanations for Betsan's results.

Assume that none of the knitters lied to Betsan and that she did not make any mistakes in her calculations.

(2)

(Total for Question 6 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

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.