Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number | Candidate Number

Thursday 22 October 2020

Afternoon (Time: 1 hour 30 minutes) | Paper Reference **8ST0/02**

Statistics

Advanced Subsidiary
Paper 2

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 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 ▶

P62298A

1/1/1/

P 6 2 2 9 8 A 0 1 2 4

Pearson

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

1 A class of students is studying for AS-level Statistics.

The students sit two mock examinations, Paper 1 and Paper 2. Each paper is scored out of 60 marks.

The students' results are shown in **Figure 1**.

Student	Paper 1	Paper 2
A	13	23
B	25	27
C	27	26
D	31	36
E	32	28
F	24	31
G	42	50
H	28	25
I	29	30
J	46	48

Figure 1

Making any necessary assumptions, use a Wilcoxon signed-rank test to investigate whether there is a difference in the average level of difficulty between Paper 1 and Paper 2.

(8)

Question 1 continued

(Total for Question 1 is 8 marks)

2 During January 2019, a Prom Party committee of Year 11 students was formed at a large school.

In March 2019, the committee published its proposed plans for the 2019 Prom Party.

An alternative group of Year 11 students disagreed with the proposed plans and the way decisions had been made by the Prom Party committee. These students formed a new committee and called this the Star Prom committee. They decided to develop alternative plans for the 2019 Prom Party.

The Prom Party committee then launched a survey, titled '2019 Prom Party Plans Feedback Survey', intended for Year 11 students to complete, and placed copies of the survey in the school entrance.

Two questions from this survey are given below.

1. How would you rate the wonderful plans that the 2019 Prom Party committee has published?

   ☐ Fantastic

   ☐ Great

   ☐ Good

7. Do you believe that the Star Prom committee should support the wonderful Prom Party plans instead of obstructing them?

   ☐ Yes

   ☐ No

   ☐ No opinion

Discuss whether the results of this survey are likely to give a representative viewpoint of students in Year 11 about plans for the 2019 Prom Party.

(4)

Question 2 continued

(Total for Question 2 is 4 marks)

3 The number of pairs of women's shoes sold in the UK in 2018 was approximately 127.5 million. The number of pairs of women's shoes sold **in each shoe size** in the UK in 2018 is shown in the diagram in **Figure 2**.

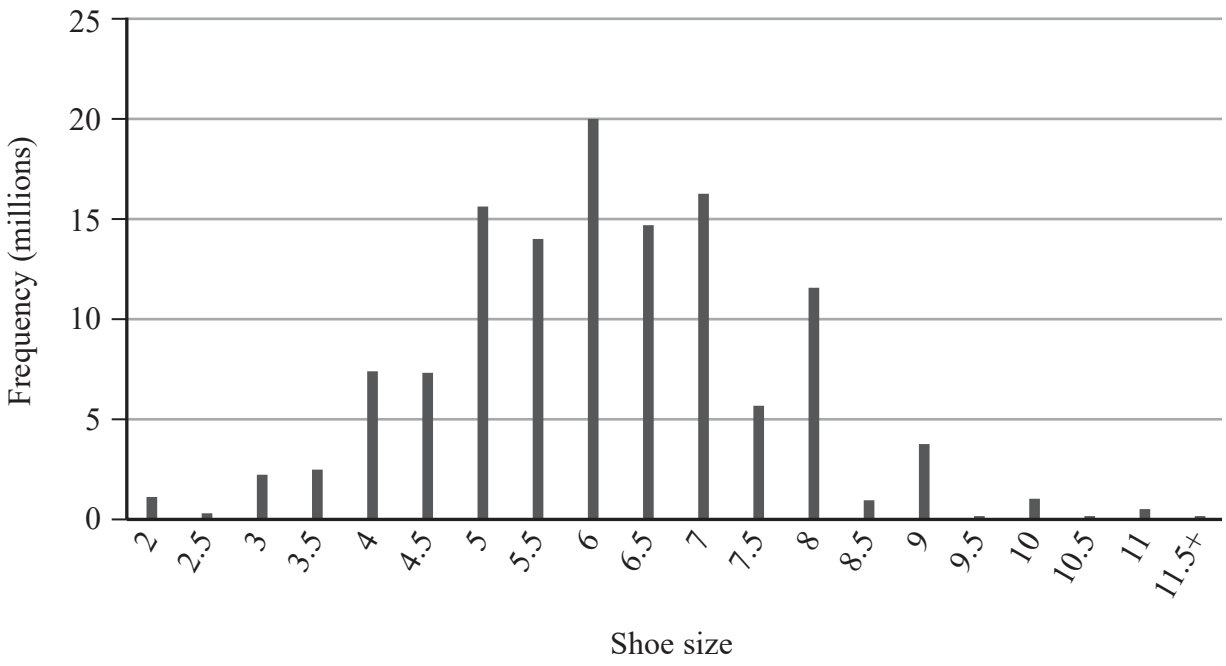

Figure 2: The number of pairs of women's shoes sold in the UK in 2018, by size

It was observed that there are two distinct curve shapes in the data in **Figure 2**.

(a) Describe this unusual pattern in the distribution.

Give a reason, in context, why this pattern may exist.

(2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 3 continued

(b) If you were to model this data, which distributions would you recommend as a good fit?

Give reasons for your choices.

(3)

(c) If a women's shoe sold in the UK in 2018 is picked at random, use **Figure 2** to estimate the probability that it is a size 3 shoe.

(2)

(d) Explain why the probability found in part (c) is **not necessarily** the same as the probability that a randomly chosen woman in the UK wears a size 3 shoe.

(1)

For the remainder of the question, you may assume that the two probabilities described in parts (c) and (d) are approximately equal.

(e) Using your answer to (c), estimate the expected number of women with a size 3 shoe in a UK town with a population of 2 500 **people**.

(3)

Question 3 continued

In the story of Cinderella, Prince Charming's method for finding Cinderella is to check every woman in a town to see whose foot fits into a small size shoe, a glass slipper.

Assume the town is the same as that in part (e).

(f) Using your answer to part (e), evaluate the effectiveness of Prince Charming's method.

(2)

(Total for Question 3 is 13 marks)

BLANK PAGE

4 Sharon is analysing some data from the 2011 UK census. She is looking for patterns between levels of qualifications that people hold in the UK.

The dataset contains data from all 393 local authority areas (census-merged) in the UK.

(Data source: 2011 census, dataset DC5104EWla)

Sharon uses a spreadsheet program to compare the following two variables:

- Proportion of the population of the local authority area that is **male**, with an **apprenticeship** qualification, working **full-time** with **two or more** dependent children in the family,
- Proportion of the population of the local authority area with a **university** qualification, working **part-time** with only **non-dependent** children in the family.

She calculates Pearson's product-moment correlation coefficient between the two variables to be $r = -0.10385$

She then uses the spreadsheet program to find the 5% critical values of a two-tailed test of this correlation coefficient for n = 393

These are ± 0.09893

She concludes with the following statement:

There is an association between the two populations:

- UK males with an apprenticeship qualification, working full-time with two or more dependent children in the family,
- UK residents with university qualifications, working part-time with only non-dependent children in the family.

Make **four** criticisms of Sharon's analysis and/or conclusion.

(4)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 4 continued

(Total for Question 4 is 4 marks)

5 In the game of cricket, a 'one-day international' (ODI) is a match between two national teams that lasts a single day. A match can be won, lost or result in a tie.

Most matches are played in the home country of one of the two nations taking part. That nation's team is referred to as the 'home team'.

At the start of the match a fair coin is tossed, and the outcome of the toss is guessed by one of the captains. The winner of the toss chooses whether his or her team will be first or second to bat.

Between January 2006 and July 2017, there were 1000 ODIs between male cricket teams where:

- one of the teams was a 'home team', and
- one of the teams won the match (no ties).

Some details of these 1000 matches are given in the table in **Figure 3**.

Winner of the toss wins match	480
Second team to bat wins match	522
'Home team' wins match	576

(Data source: cricsheet.org)

Figure 3

(a) State **two** assumptions necessary for modelling the number of matches won by the 'home team' with a binomial distribution.

(2)

(b) Making any necessary assumptions:

(i) give a reason why it is **not** necessary to carry out a hypothesis test to investigate whether winning the toss gives an advantage in terms of winning a match,

(1)

Question 5 continued

(ii) investigate whether being second to bat gives an advantage in terms of winning a match,

(4)

(iii) investigate whether playing at home gives an advantage in terms of winning a match.

(2)

Question 5 continued

(c) Comment on the validity of your two assumptions given in (a).

(3)

(Total for Question 5 is 12 marks)

BLANK PAGE

6 An on-demand television service sells advertising to companies at a cost that relates to the probability that their advertisement is shown.

Each advertisement break within programmes shows 3 randomly-sampled advertisements.

(a) From the perspective of the **television viewers**, is it more desirable to use sampling **without** replacement or **with** replacement when selecting the three advertisements for an advertisement break?

Explain your answer.

(2)

Company A pays the service a monthly fee, as do a number of other companies.

In a given month, the probability that Company A's single advertisement is shown **first** in an advertisement break is equal to Company A's fee for that month as a proportion of the total fees paid that month by all companies.

In February, Company A pays the service a fee of £25 000 for a single advertisement.

The service received a total of £1 000 000 in fees that month.

(b) Explain why the probability that Company A's advertisement is shown first in an advertisement break in February equals $\frac{1}{40}$

(1)

Question 6 continued

A simple model is constructed which assumes that all companies who advertise pay the same fee (£25 000) in February for a single advertisement.

(c) Write down how many companies are paying for a single advertisement in February.

(1)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 6 continued

A tree diagram for the probabilities that Company A's advertisement is shown during a randomly chosen advertisement break using this simple model is given in **Figure 4**.

S represents the event that Company A's advertisement is selected.

(d) Explain how **Figure 4** illustrates that sampling without replacement has been used.

(1)

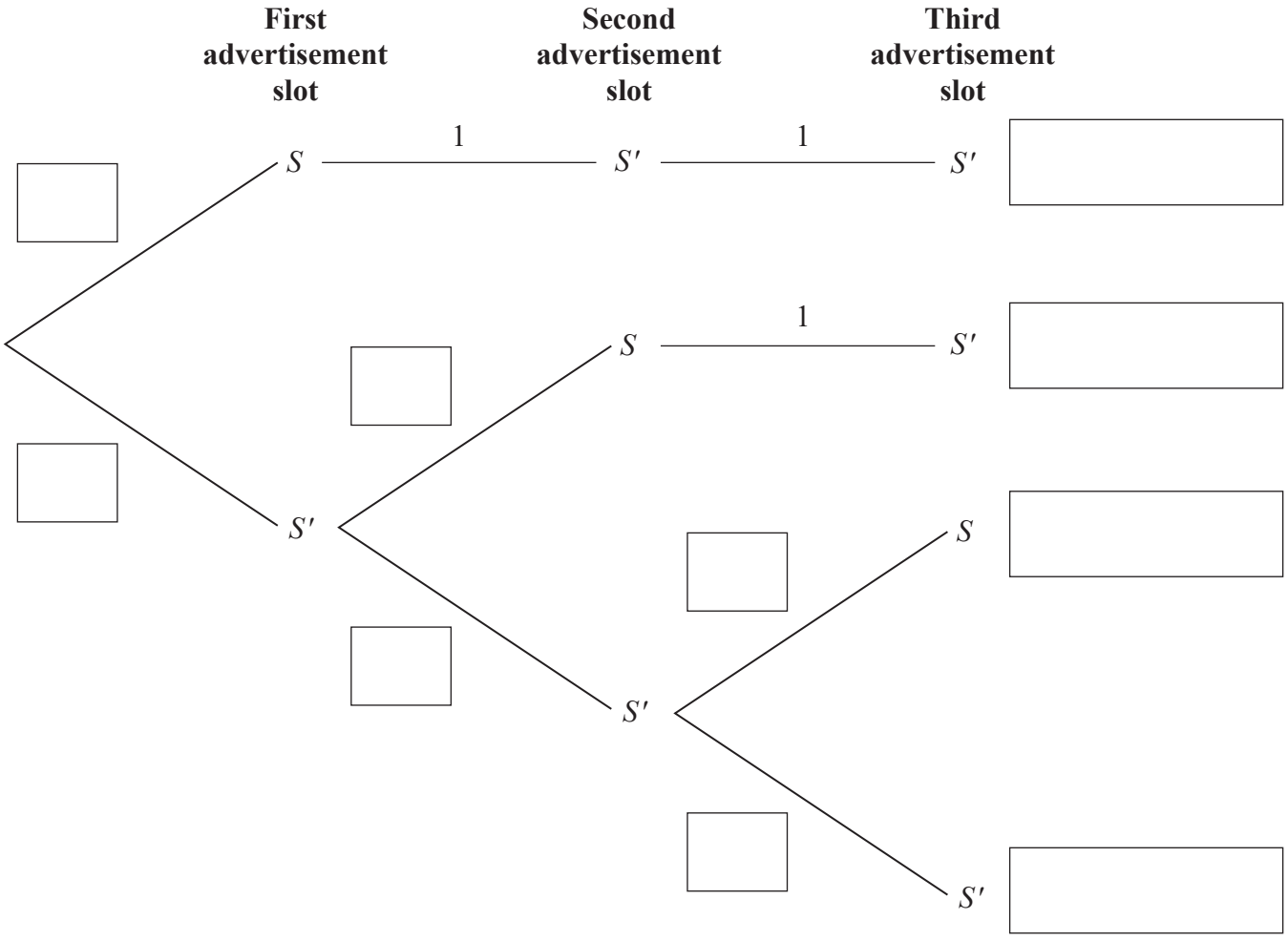

Figure 4

Question 6 continued

The probabilities of any one of the remaining advertisements being selected for the second slot are equal. This also applies for those remaining for the third slot.

(e) Fill in the missing probabilities in the boxes given in **Figure 4**.

(5)

Space for working

(f) Using this simple model, find the probability that Company A's advertisement is shown in a randomly selected advertisement break.

(2)

Question 6 continued

(g) From the perspective of a **company** advertising with the service, using this simple model, is it more desirable to use sampling **without** replacement or **with** replacement when selecting the three advertisements for an advertisement break?

Explain your answer, using further calculations where appropriate.

(4)

Question 6 continued

Using a different model, the probability that Company B's advertisement is shown in a randomly selected advertisement **break** is 0.06

The service offers a film that plays with 7 advertisement breaks.

(h) Calculate the probability that Company B's advertisement will play **at least** twice during the film.

Showings of Company B's advertisements are independent between breaks.

(3)

(Total for Question 6 is 19 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

BLANK PAGE

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE