Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

Tuesday 21 May 2024

Morning (Time: 1 hour 30 minutes)

Paper reference **WST01/01**

Mathematics

International Advanced Subsidiary/Advanced Level
Statistics S1

You must have:
Mathematical Formulae and Statistical Tables (Yellow), calculator

Total Marks

Candidates may use any calculator permitted by Pearson regulations. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or 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.
- Values from the statistical tables should be quoted in full. If a calculator is used instead of the tables, the value should be given to an equivalent degree of accuracy.
- Inexact answers should be given to three significant figures unless otherwise stated.

Information

- A booklet 'Mathematical Formulae and Statistical Tables' is provided.
- There are 6 questions in this question paper. The total mark for this paper is 75.
- 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 ▶

P75720RA

F:1/1/1/1/

P 7 5 7 2 0 R A 0 1 2 4

Pearson

1. A researcher is investigating the growth of two types of tree, Birch and Maple. The height, to the nearest cm, a seedling grows in one year is recorded for 35 Birch trees and 32 Maple trees. The results are summarised in the back-to-back stem and leaf diagram below.

Totals	Birch		Maple	Totals
(2)	9 8	2	5 7 7 8 9	(5)
(8)	9 9 9 6 5 3 1 1	3	0 2 6 6 8 9 9	(7)
(9)	9 8 8 7 6 3 1 1 1	4	1 1 1 ***k*** 7 8	(6)
(9)	7 7 7 5 4 3 2 1 0	5	0 1 2 3 4 4 4	(7)
(3)	7 6 5	6	3 4 6	(3)
(3)	6 5 4	7	0 7	(2)
(1)	5	8	0 0	(2)

Key: 5 | 6 | 3 means 65 cm for a Birch tree and 63 cm for a Maple tree

The median height that these **Maple** trees grow in one year is 45 cm.

(a) Find the value of ***k***, used in the stem and leaf diagram.

(1)

(b) Find the lower quartile and the upper quartile of the height grown in one year for these **Birch** trees.

(2)

The researcher defines an outlier as an observation that is

$$\text{greater than } Q_3 + 1.5 \times (Q_3 - Q_1) \text{ or less than } Q_1 - 1.5 \times (Q_3 - Q_1)$$

(c) Show that there is only one outlier amongst the Birch trees.

(2)

The grid on page 3 shows a box plot for the heights that the Maple trees grow in one year.

(d) On the same grid draw a box plot for the heights that the Birch trees grow in one year.

(4)

(e) Comment on any difference in the distributions of the growth of these Birch trees and the growth of these Maple trees.
State the values of any statistics you have used to support your comment.

(1)

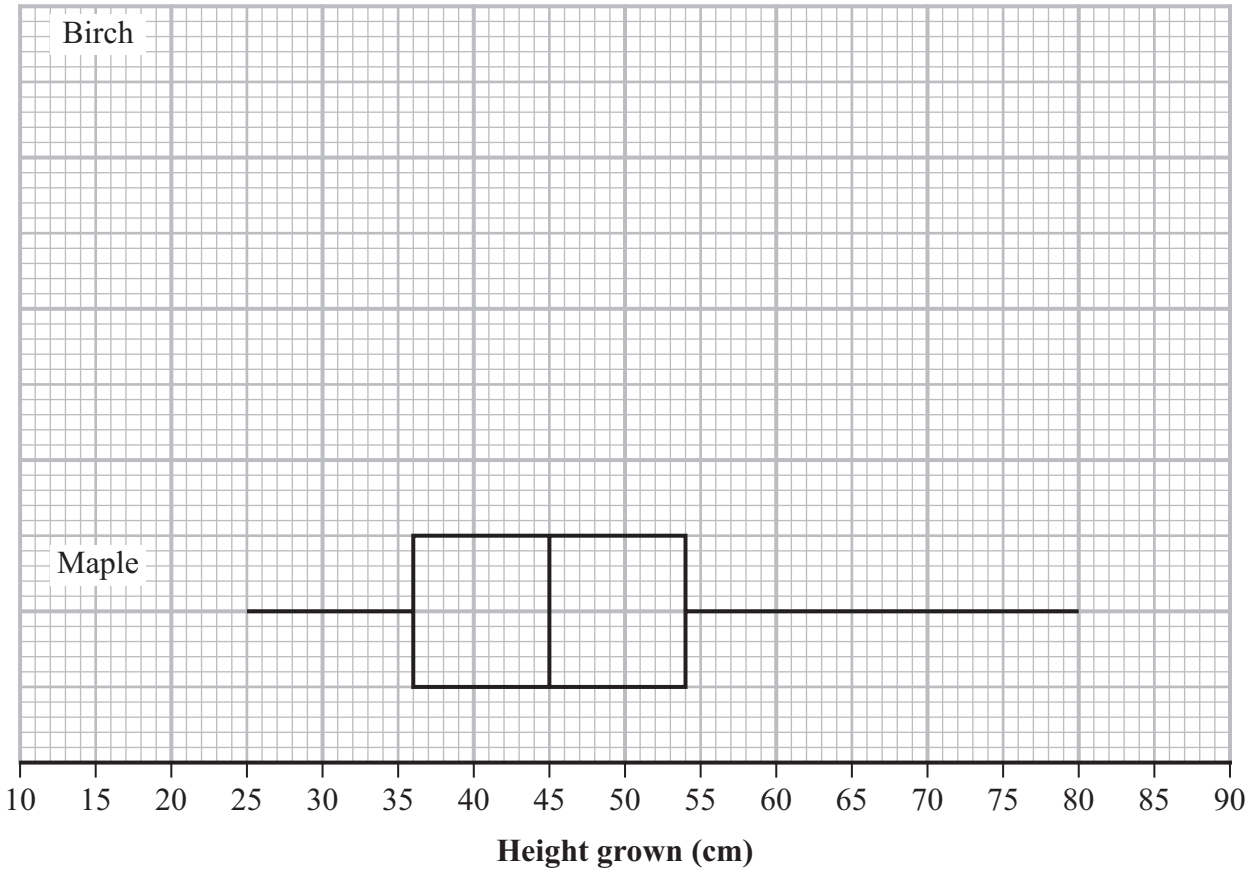
The researcher realises he has missed out 4 pieces of data for the **Maple** trees. The heights each seedling grows in one year, to the nearest cm, in ascending order, for these 4 Maple trees are 27 cm, a cm, 48 cm, $2a$ cm.

Given that there is no change to the box plot for the **Maple** trees given on page 3

(f) find the range of possible values for a
Show your working clearly.

(3)

Question 1 continued



Turn over for spare grid if you need to redraw your answer for part (d)

Question 1 continued

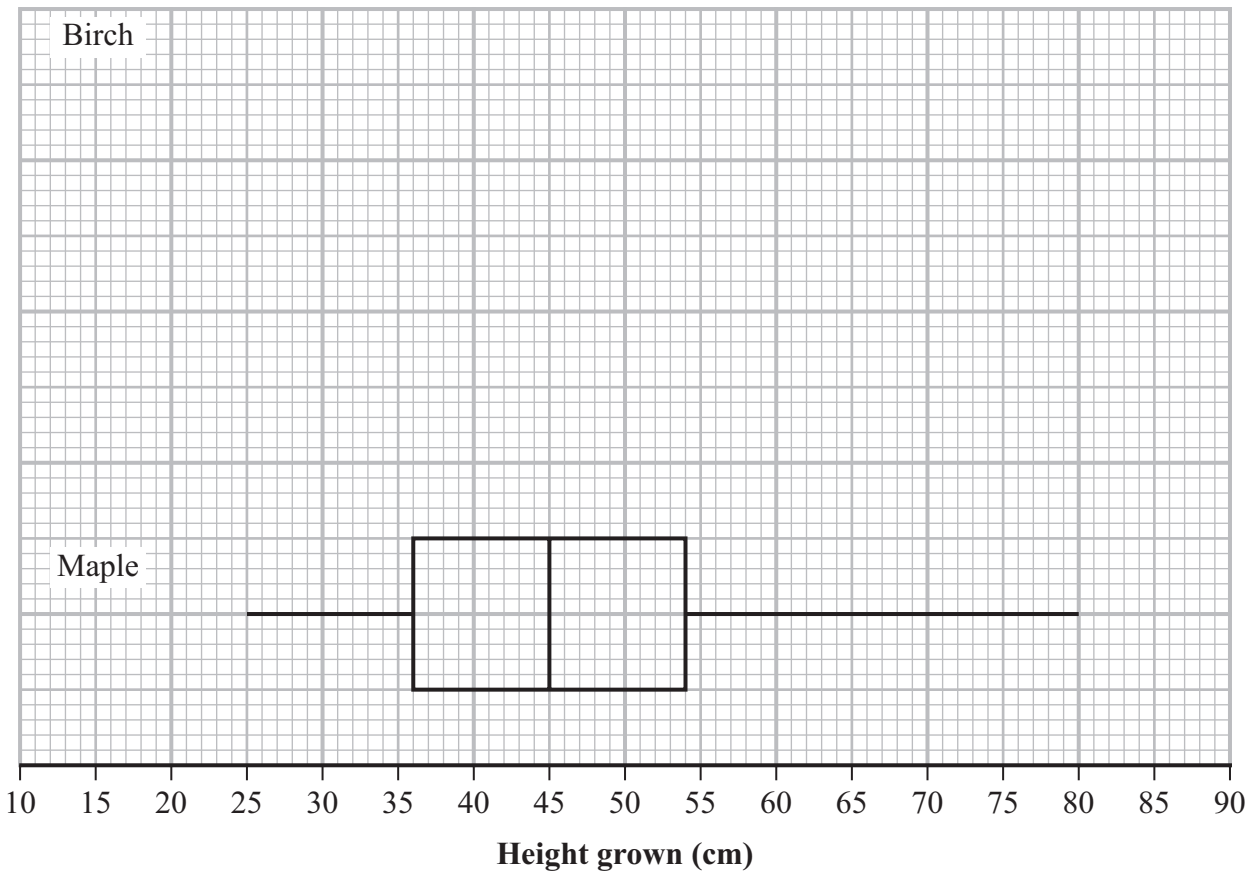
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 1 continued

Only use this grid if you need to redraw your answer for part (d)



(Total for Question 1 is 13 marks)

2. A spinner can land on the numbers 2, 4, 5, 7 or 8 only.
The random variable X represents the number that this spinner lands on when it is spun once. The probability distribution of X is given in the table below.

x	2	4	5	7	8
$\mathbf{P}(X = x)$	0.25	0.3	0.2	0.1	0.15

(a) Find $P(2X - 3 > 5)$

(1)

Given that $E(X) = 4.6$

(b) show that $Var(X) = 4.14$

(3)

The random variable $Y = aX - b$ where a and b are positive constants.

Given that

$$E(Y) = 13.4 \quad \text{and} \quad Var(Y) = 66.24$$

(c) find the value of a and the value of b

(4)

In a game Sam and Alex each spin the spinner once, landing on X_1 and X_2 respectively.

Sam's score is given by the random variable $S = X_1$
Alex's score is given by the random variable $R = 2X_2 - 3$

The person with the higher score wins the game. If the scores are the same it is a draw.

(d) Find the probability that Sam wins the game.

(4)

Question 2 continued

Question 2 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 2 continued

(Total for Question 2 is 12 marks)

3. The lengths, x mm, of 50 pebbles are summarised in the table below.

Length	Frequency
$20 \leqslant x < 30$	2
$30 \leqslant x < 32$	16
$32 \leqslant x < 36$	20
$36 \leqslant x < 40$	8
$40 \leqslant x < 45$	3
$45 \leqslant x < 50$	1

A histogram is drawn to represent these data.

The bar representing the class $32 \leqslant x < 36$ is 2.5 cm wide and 7.5 cm tall.

(a) Calculate the width and the height of the bar representing the class $30 \leqslant x < 32$

(3)

(b) Using linear interpolation, estimate the median of x

(2)

The weight, w grams, of each of the 50 pebbles is coded using $10y = w - 20$
These coded data are summarised by

$$\sum y = 104 \qquad \sum y^2 = 233.54$$

(c) Show that the mean of w is 40.8

(2)

(d) Calculate the standard deviation of w

(4)

The weight of a pebble recorded as 40.8 grams is added to the sample.

(e) Without carrying out any further calculations, state, giving a reason, what effect this would have on the value of

(i) the mean of w

(ii) the standard deviation of w

(3)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 3 continued

Question 3 continued

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 14 marks)

4. A biologist is studying bears. The biologist records the length, d cm, and the girth, g cm, of 8 bears. The biologist summarises the data as follows

$$\sum d = 1456.8 \qquad \sum g = 713.2 \qquad \sum dg = 141\,978.84 \qquad \sum g^2 = 72\,675.98$$

$$S_{dd} = 16\,769.78$$

(a) Calculate the exact value of S_{dg} and the exact value of S_{gg}

(3)

(b) Calculate the value of the product moment correlation coefficient between d and g

(2)

(c) Show that the equation of the regression line of g on d can be written as

$$g = -42.3 + 0.722d$$

where the values of the intercept and gradient are given to 3 significant figures.

(3)

(d) Give an interpretation, in context, of the gradient of the regression line.

(1)

Using the equation of the regression line given in part (c)

(e) (i) estimate the girth of a bear with a length of 2.5 metres,

(ii) explain why an estimate for the girth of a bear with a length of 0.5 metres is not reliable.

(2)

Using the regression line from part (c), the biologist estimates that for each x cm increase in the length of a bear there will be a 17.3 cm increase in the girth.

(f) Find the value of x

(2)

Question 4 continued

Question 4 continued

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 13 marks)

5. A competition consists of two rounds.

The time, in minutes, taken by adults to complete round one is modelled by a normal distribution with mean 15 minutes and standard deviation 2 minutes.

(a) Use standardisation to find the proportion of adults that take less than 18 minutes to complete round one.

(2)

Only the fastest 60% of adults from round one take part in round two.

(b) Use standardisation to find the longest time that an adult can take to complete round one if they are to take part in round two.

(3)

The time, T minutes, taken by adults to complete round two is modelled by a normal distribution with mean μ

Given that $P(\mu - 10 < T < \mu + 10) = 0.95$

(c) find $P(T > \mu - 5 \mid T > \mu - 10)$

(5)

Question 5 continued

Question 5 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

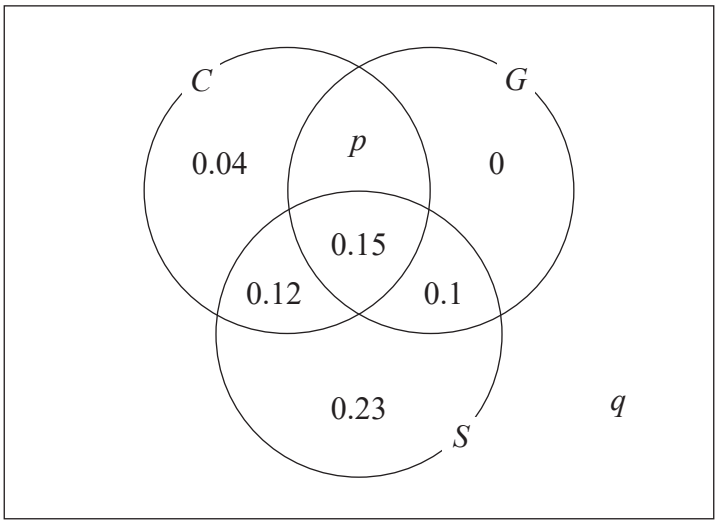
Question 5 continued

(Total for Question 5 is 10 marks)

6. The Venn diagram shows the probabilities related to teenagers playing 3 particular board games.

C is the event that a teenager plays Chess
S is the event that a teenager plays Scrabble
G is the event that a teenager plays Go

where p and q are probabilities.



(a) Find the probability that a randomly selected teenager plays Chess but does not play Go.

(1)

Given that the events C and S are independent,

(b) find the value of p

(4)

(c) Hence find the value of q

(2)

(d) Find (i) $P((C \cup S) \cap G')$

(1)

(ii) $P(C \mid (S \cap G))$

(2)

A youth club consists of a large number of teenagers.
In this youth club 76 teenagers play Chess and Go.

(e) Use the information in the Venn diagram to estimate how many of the teenagers in the youth club do not play Scrabble.

(3)

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

Question 6 continued

Question 6 continued

(Total for Question 6 is 13 marks)

TOTAL FOR PAPER IS 75 MARKS

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA