Pearson Edexcel IAL (Further) Mathematics

Statistics 1

www.CasperYC.club/wst01

Past Paper Collection

Last updated: January 21, 2025

CC BY NC ND

Comments and suggestions to DrYuFromShanghai@QQ.com

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel
International Advanced Level

Centre Number | Candidate Number

Monday 20 January 2020

Afternoon (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 (Blue), 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 ▶

Leave blank

1. The discrete random variable X has the following probability distribution

x	–2	–1	1	3	4
$P(X = x)$	0.15	a	b	c	0.15

where a, b and c are probabilities.

The mean value of X is 1 and $F(1) = 0.63$

Find the value of a, the value of b and the value of c.

(5)

Leave blank

Question 1 continued

Q1

(Total 5 marks)

Leave blank

2. A group of 40 families was asked whether their family had a dog, a cat or a rabbit as pets.

No family had a dog and a rabbit
2 families had both a dog and a cat
12 families had a dog
14 families had a cat
11 families had a rabbit
9 families did not have any of these animals as pets

A family from this group is selected at random.

D represents the event that the family has a dog
C represents the event that the family has a cat
R represents the event that the family has a rabbit

(a) Draw a Venn diagram to illustrate this information.

(4)

(b) State, giving a reason, a pair of mutually exclusive events from D, C and R.

(1)

(c) Find the probability that the family has exactly 2 of these kinds of animals as pets.

(1)

(d) Showing your working clearly, determine whether or not the events D and C are independent.

(2)

Sarah's family is in the group and her family has a pet cat.

(e) Find the probability that Sarah's family also has a pet rabbit.

(2)

(f) Find the exact value of $P([D \cup R] \mid C')$

(2)

Leave blank

Question 2 continued

Leave blank

Question 2 continued

Leave blank

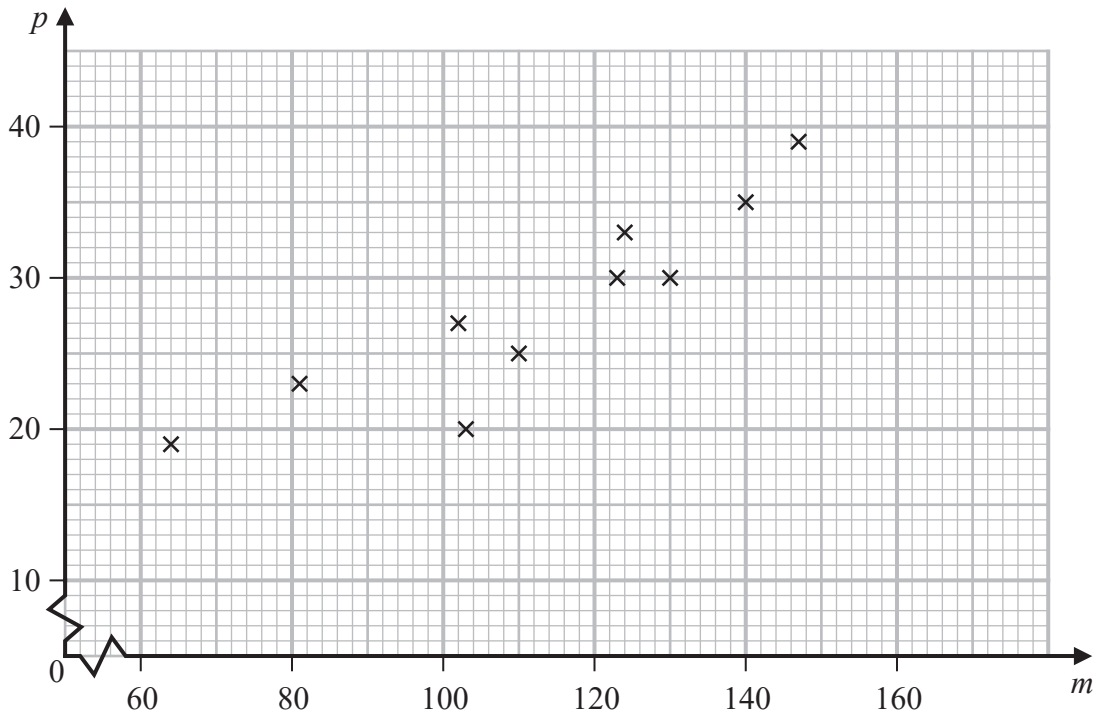
Question 2 continued

(Total 12 marks)

Q2

Leave blank

3. *Soapern* sells washing machines. When a customer buys a washing machine from *Soapern*, the customer is also invited to buy a guarantee policy to cover breakdowns and repairs for the next three years.
The manager of *Soapern* believes that the relationship between the number of washing machines sold (m) and the number of guarantee policies sold (p) can be modelled by a straight line.
She collected data each month for 10 months. The scatter diagram below illustrates these data.



The data are summarised by the following statistics.

$$\sum m = 1124 \quad \sum p = 281 \quad \sum mp = 32\,958 \quad S_{mm} = 6046.4 \quad S_{pp} = 382.9$$

(a) Show that $S_{mp} = 1373.6$

(1)

(b) Find the value of the product moment correlation coefficient for these data.

(2)

(c) State, giving a reason, whether or not the data are consistent with the manager's belief.

(1)

The manager noticed that the total number of washing machines sold was k times the total number of guarantee policies sold and suggests a model of the form $p = \frac{1}{k}m$, where k is an integer.

(d) Find the value of k.

(2)

Leave blank

Jiang works for *Soapern* and thought that this model oversimplified the situation and suggested that a linear regression of p on m may be more appropriate.

(e) Find the equation of the linear regression of p on m, giving your answer in the form $p = a + bm$, where a and b should be given to 3 significant figures.

(4)

(f) Use Jiang's model to estimate the number of guarantee policies sold when 70 washing machines are sold in a month.

(1)

Usually about 70 washing machines are sold in January. *Soapern* decides to offer a bonus to staff during January based on the number of guarantee policies sold. If the number of guarantee policies sold is greater than the number estimated by the model, the bonus will be paid.

(g) State, giving your reasons, whether you would recommend that the staff use the manager's model or Jiang's model.

(2)

Leave blank

Question 3 continued

Leave blank

Question 3 continued

(Total 13 marks)

Q3

Leave blank

4. A researcher is studying the birth weights of babies. A random sample of 98 babies was taken and their birth weights, w kg, are summarised in the table below.

Birth weight (w kg)	Frequency (f)	Birth weight midpoint (x)
$1.50 \leqslant w < 2.50$	16	2.00
$2.50 \leqslant w < 3.00$	24	2.75
$3.00 \leqslant w < 3.50$	32	3.25
$3.50 \leqslant w < 4.00$	14	3.75
$4.00 \leqslant w < 5.50$	12	4.75

$$\left(\text{You may use } \sum fx = 311.5 \text{ and } \sum fx^2 = 1051.125\right)$$

A histogram is drawn to represent these data.

The bar representing the birth weight $1.50 \leqslant w < 2.50$ has a width of 1 cm and a height of 4 cm.

(a) Calculate the width and height of the bar representing birth weight $3.50 \leqslant w < 4.00$

(3)

(b) Use linear interpolation to estimate the lower quartile of the birth weights of the 98 babies.

(2)

The researcher estimated the median to be 3.14 kg and the upper quartile to be 3.55 kg.

(c) Use the median and quartiles to describe the skewness of these data.

(2)

(d) Find an estimate for (i) the mean birth weight
(ii) the standard deviation of the birth weights.

(3)

(e) Use the formula

$$\text{skewness} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

to estimate a value for the skewness of these data. Give your answer to 2 significant figures.

(2)

The researcher read that birth weights should be approximately normally distributed and decides to split the class $3.00 \leqslant w < 3.50$

The frequency for $3.00 \leqslant w < 3.25$ is 9 and the frequency for $3.25 \leqslant w < 3.50$ is 23

(f) (i) State, giving a reason, what the effect would be on the estimate of the median.

(ii) Without carrying out any further calculations state, giving a reason, what the effect of this change would be on the estimate of the mean.

(2)

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Leave blank

Question 4 continued

(Total 14 marks)

Q4

Leave blank

5. The random variable X has a normal distribution with mean 10 and standard deviation 6

(a) Find $P(X < 7)$

(3)

(b) Find the value of k such that

$$P(10 - k < X < 10 + k) = 0.60$$

(3)

A single observation x, of X, is to be taken.
A rectangle is drawn on a centimetre grid with vertices having coordinates $(0, 0)$, $(x, 0)$, $(x, x - 3)$ and $(0, x - 3)$

(c) Find the probability that the area of this rectangle is more than $40\,cm^2$

(8)

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Leave blank

Question 5 continued

(Total 14 marks)

Q5

Leave blank

6. A tennis tournament has 5 rounds. After each round, winners go into the next round and losers are knocked out of the tournament. To enter the tournament players must pay an entry fee of \$10 but only the person who wins all 5 rounds receives the prize of \$260

Serena enters this tennis tournament. The random variable S represents the total number of rounds Serena wins. The probability distribution for S is given in the following table.

s	0	1	2	3	4	5
$P(S = s)$	k	$\frac{k}{2}$	$\frac{k}{3}$	$\frac{k}{4}$	$\frac{k}{5}$	$\frac{k}{6}$

(a) Show that $k = \frac{20}{49}$

(2)

(b) Find $E(S)$

(3)

(c) Find Serena's expected profit if she enters the tennis tournament.

(3)

Roger also enters this tennis tournament. Given that Roger is still in the tournament, the probability that he wins the next round is a constant p.
The random variable R represents the total number of rounds that Roger wins.

(d) Explain why $P(R = 2) = p^2(1 - p)$

(2)

(e) Find, in terms of p, the probability distribution for R.

(3)

(f) Find the smallest value of p such that Roger's expected profit is at least as great as Serena's.

(4)

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Q6

(Total 17 marks)

TOTAL FOR PAPER: 75 MARKS

END

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel
International Advanced Level

Centre Number | Candidate Number

Thursday 22 October 2020

Afternoon (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 (Blue), 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 ▶

Leave blank

1. The discrete random variable X takes the values -1, 2, 3, 4 and 7 only.

Given that

$$P(X = x) = \frac{8 - x}{k} \text{ for } x = -1, 2, 3, 4 \text{ and } 7$$

find the value of $E(X)$

(5)

Leave blank

Question 1 continued

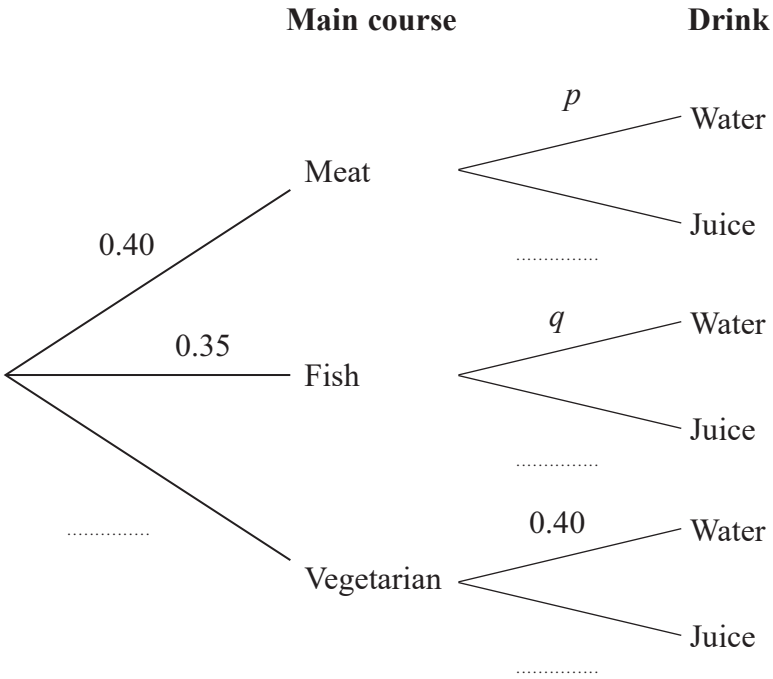
Q1

(Total 5 marks)

Leave blank

2. In a school canteen, students can choose from a main course of meat (M), fish (F) or vegetarian (V). They can then choose a drink of either water (W) or juice (J).

The partially completed tree diagram, where p and q are probabilities, shows the probabilities of these choices for a randomly selected student.



(a) Complete the tree diagram, giving your answers in terms of p and q where appropriate.

(2)

(b) Find an expression, in terms of p and q, for the probability that a randomly selected student chooses water to drink.

(1)

The events "choosing a vegetarian main course" and "choosing water to drink" are independent.

(c) Find a linear equation in terms of p and q.

(2)

A student who has chosen juice to drink is selected at random. The probability that they chose fish for their main course is $\frac{7}{30}$

(d) Find the value of p and the value of q.

(5)

The canteen manager claims that students who choose water to drink are most likely to choose a fish main course.

(e) State, showing your working clearly, whether or not the manager's claim is correct.

(3)

Leave blank

Question 2 continued

Leave blank

Question 2 continued

Leave blank

Question 2 continued

(Total 13 marks)

Q2

Leave blank

3. The distance achieved in a long jump competition by students at a school is normally distributed with mean 3.8 metres and standard deviation 0.9 metres.

Students who achieve a distance greater than 4.3 metres receive a medal.

(a) Find the proportion of students who receive medals.

(3)

The school wishes to give a certificate of achievement or a medal to the 80% of students who achieve a distance of at least d metres.

(b) Find the value of d.

(3)

Of those who received medals, the $\frac{1}{3}$ who jump the furthest will receive gold medals.

(c) Find the shortest distance, g metres, that must be achieved to receive a gold medal.

(4)

A journalist from the local newspaper interviews a randomly selected group of 3 medal winners.

(d) Find the exact probability that there is at least one gold medal winner in the group.

(3)

Leave blank

Question 3 continued

Leave blank

Question 3 continued

Leave blank

Question 3 continued

(Total 13 marks)

Q3

Leave blank

4. A group of students took some tests. A teacher is analysing the average mark for each student. Each student obtained a different average mark.

For these average marks, the lower quartile is 24, the median is 30 and the interquartile range (IQR) is 10
The three lowest average marks are 8, 10 and 15.5 and the three highest average marks are 45, 52.5 and 56

The teacher defines an outlier to be a value that is either

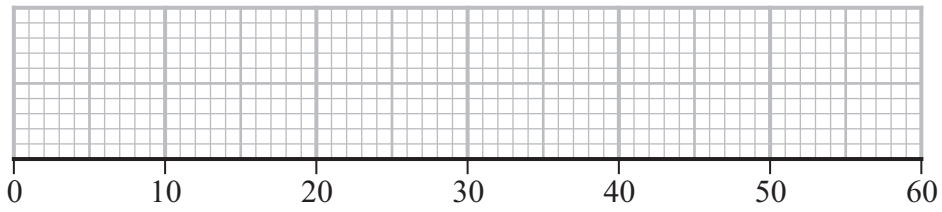
more than 1.5 × IQR below the lower quartile or
more than 1.5 × IQR above the upper quartile

(a) Determine any outliers in these data.

(4)

(b) On the grid below draw a box plot for these data, indicating clearly any outliers.

(3)



(c) Use the quartiles to describe the skewness of these data.
Give a reason for your answer.

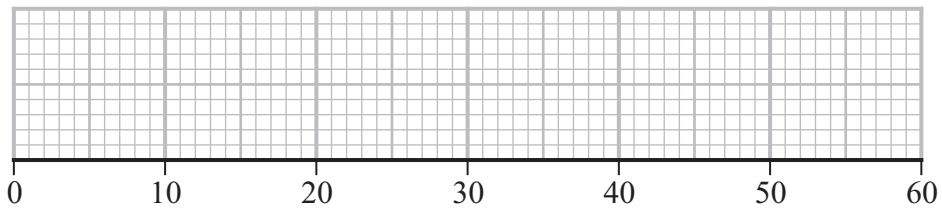
(2)

Two more students also took the tests. Their average marks, which were both less than 45, are added to the data and the box plot redrawn.

The median and the upper quartile are the same but the lower quartile is now 26

(d) Redraw the box plot on the grid below.

(3)



(e) Give ranges of values within which each of these students' average marks must lie.

(2)

Leave blank

Question 4 continued

Turn over for spare grids if you need to redraw your answer for part (b) or part (d).

Leave blank

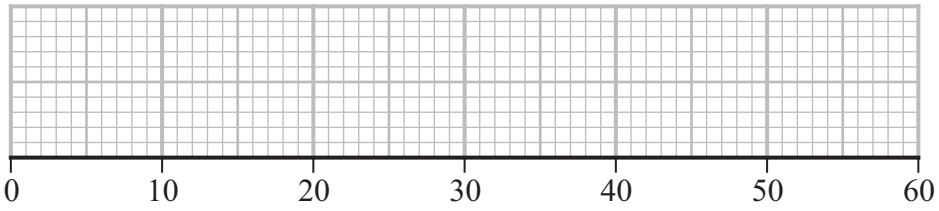
Question 4 continued

Leave blank

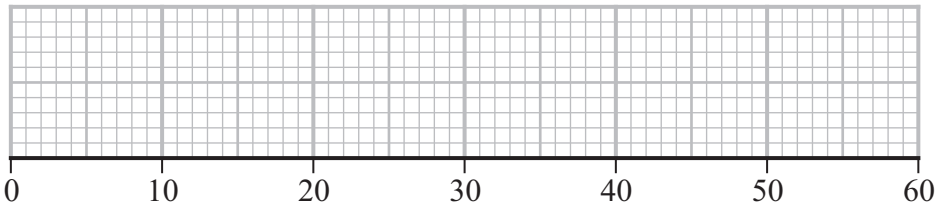
Question 4 continued

Only use these grids if you need to redraw your answer for part (b) or part (d).

Copy of grid for part (b)



Copy of grid for part (d)



Q4

(Total 14 marks)

Leave blank

5. A large company rents shops in different parts of the country. A random sample of 10 shops was taken and the floor area, x in $10\,m^2$, and the annual rent, y in thousands of dollars, were recorded.
The data are summarised by the following statistics

$$\sum x = 900 \quad \sum x^2 = 84\,818 \quad \sum y = 183 \quad \sum y^2 = 3434$$

and the regression line of y on x has equation $y = 6.066 + 0.136x$

(a) Use the regression line to estimate the annual rent in dollars for a shop with a floor area of $800\,m^2$

(2)

(b) Find S_{yy} and S_{xx}

(3)

(c) Find the product moment correlation coefficient between y and x.

(4)

An 11th shop is added to the sample. The floor area is $900\,m^2$ and the annual rent is 15 000 dollars.

(d) Use the formula $S_{xy} = \sum(x - \bar{x})(y - \bar{y})$ to show that the value of S_{xy} for the 11 shops will be the same as it was for the original 10 shops.

(2)

(e) Find the new equation of the regression line of y on x for the 11 shops.

(3)

The company is considering renting a larger shop with area of $3000\,m^2$

(f) Comment on the suitability of using the new regression line to estimate the annual rent. Give a reason for your answer.

(1)

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Leave blank

Question 5 continued

(Total 15 marks)

Q5

Leave blank

6. The random variable A represents the score when a spinner is spun. The probability distribution for A is given in the following table.

a	1	4	5	7
$P(A = a)$	0.40	0.20	0.25	0.15

(a) Show that $E(A) = 3.5$

(2)

(b) Find $\text{Var}(A)$

(3)

The random variable B represents the score on a 4-sided die. The probability distribution for B is given in the following table where k is a positive integer.

b	1	3	4	k
$P(B = b)$	0.25	0.25	0.25	0.25

(c) Write down the name of the probability distribution of B.

(1)

(d) Given that $E(B) = E(A)$ state, giving a reason, the value of k.

(1)

The random variable $X \sim N(\mu, \sigma^2)$

Sam and Tim are playing a game with the spinner and the die.

They each spin the spinner once to obtain their value of A and each roll the die once to obtain their value of B.
Their value of A is taken as their value of μ and their value of B is taken as their value of σ.
The person with the larger value of $P(X > 3.5)$ is the winner.

(e) Given that Sam obtained values of $a = 4$ and $b = 3$ and Tim obtained $b = 4$ find, giving a reason, the probability that Tim wins.

(2)

(f) Find the largest value of $P(X > 3.5)$ achievable in this game.

(4)

(g) Find the probability of achieving this value.

(2)

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Q6

(Total 15 marks)

TOTAL FOR PAPER: 75 MARKS

END

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel
International Advanced Level

Centre Number | Candidate Number

Wednesday 13 January 2021

Afternoon (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 (Blue), 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.*

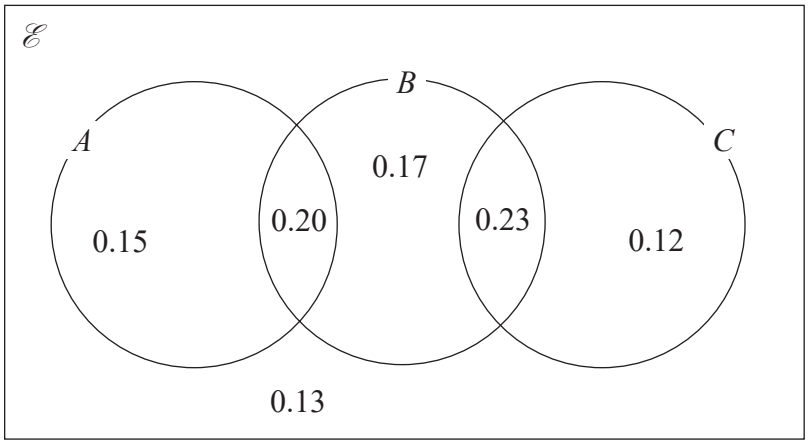
Advice

- 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 ▶

Leave blank

1. The Venn diagram shows the events A, B and C and their associated probabilities.



Find

(a) $P(B')$

(1)

(b) $P(A \cup C)$

(2)

(c) $P(A \mid B')$

(2)

Leave blank

Question 1 continued

Q1

(Total 5 marks)

Leave blank

2. The stem and leaf diagram below shows the ages (in years) of the residents in a care home.

Age												Key: 4|3 is an age of 43
4	3											(1)
5	4											(1)
6	2	3	5	6	8	8	8	9	9			(9)
7	1	1	3	4	4	6	6	6	8	8	9	(11)
8	0	0	2	7	8	8	9					(7)
9	3	7										(2)

(a) Find the median age of the residents.

(1)

(b) Find the interquartile range (IQR) of the ages of the residents.

(2)

An outlier is defined as a value that is either

more than 1.5 × (IQR) below the lower quartile or

more than 1.5 × (IQR) above the upper quartile.

(c) Determine any outliers in these data. Show clearly any calculations that you use.

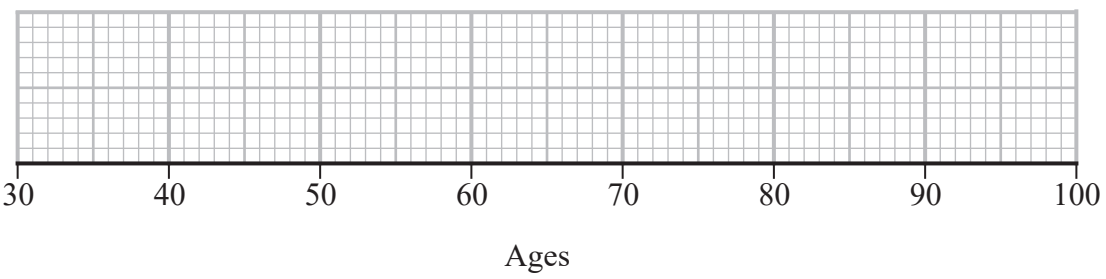
(3)

(d) On the grid on page 5, draw a box plot to summarise these data.

(3)

Leave blank

Question 2 continued



Q2

(Total 9 marks)

Leave blank

3. The weights of packages that arrive at a factory are normally distributed with a mean of 18 kg and a standard deviation of 5.4 kg

(a) Find the probability that a randomly selected package weighs less than 10 kg

(3)

The heaviest 15% of packages are moved around the factory by Jemima using a forklift truck.

(b) Find the weight, in kg, of the lightest of these packages that Jemima will move.

(3)

One of the packages **not** moved by Jemima is selected at random.

(c) Find the probability that it weighs more than 18 kg

(4)

A delivery of 4 packages is made to the factory.
The weights of the packages are independent.

(d) Find the probability that exactly 2 of them will be moved by Jemima.

(3)

Leave blank

Question 3 continued

Leave blank

Question 3 continued

Leave blank

Question 3 continued

(Total 13 marks)

Q3

Leave blank

4. A spinner can land on the numbers 10, 12, 14 and 16 only and the probability of the spinner landing on each number is the same.
The random variable X represents the number that the spinner lands on when it is spun once.

(a) State the name of the probability distribution of X.

(1)

(b) (i) Write down the value of $E(X)$

(1)

(ii) Find $Var(X)$

(2)

A second spinner can land on the numbers 1, 2, 3, 4 and 5 only.
The random variable Y represents the number that this spinner lands on when it is spun once. The probability distribution of Y is given in the table below

y	1	2	3	4	5
$P(Y = y)$	$\frac{4}{30}$	$\frac{9}{30}$	$\frac{6}{30}$	$\frac{5}{30}$	$\frac{6}{30}$

(c) Find (i) $E(Y)$

(2)

(ii) $Var(Y)$

(3)

The random variable $W = aX + b$, where a and b are constants and $a > 0$
Given that $E(W) = E(Y)$ and $Var(W) = Var(Y)$

(d) find the value of a and the value of b.

(5)

Each of the two spinners is spun once.

(e) Find $P(W = Y)$

(2)

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Q4

(Total 16 marks)

Leave blank

5. A company director wants to introduce a performance-related pay structure for her managers. A random sample of 15 managers is taken and the annual salary, y in £1000, was recorded for each manager. The director then calculated a performance score, x, for each of these managers.
The results are shown on the scatter diagram in Figure 1 on the next page.

(a) Describe the correlation between performance score and annual salary.

(1)

The results are also summarised in the following statistics.

$$\sum x = 465 \quad \sum y = 562 \quad S_{xx} = 2492 \quad \sum y^2 = 23140 \quad \sum xy = 19428$$

(b) (i) Show that $S_{xy} = 2006$

(1)

(ii) Find S_{yy}

(2)

(c) Find the product moment correlation coefficient between performance score and annual salary.

(2)

The director believes that there is a linear relationship between performance score and annual salary.

(d) State, giving a reason, whether or not these data are consistent with the director's belief.

(1)

(e) Calculate the equation of the regression line of y on x, in the form $y = a + bx$
Give the value of a and the value of b to 3 significant figures.

(4)

(f) Give an interpretation of the value of b.

(1)

(g) Plot your regression line on the scatter diagram in Figure 1

(2)

The director hears that one of the managers in the sample seems to be underperforming.

(h) On the scatter diagram, circle the point that best identifies this manager.

(1)

The director decides to use this regression line for the new performance related pay structure.

(i) Estimate, to 3 significant figures, the new salary of a manager with a performance score of 30

(2)

Leave blank

Question 5 continued

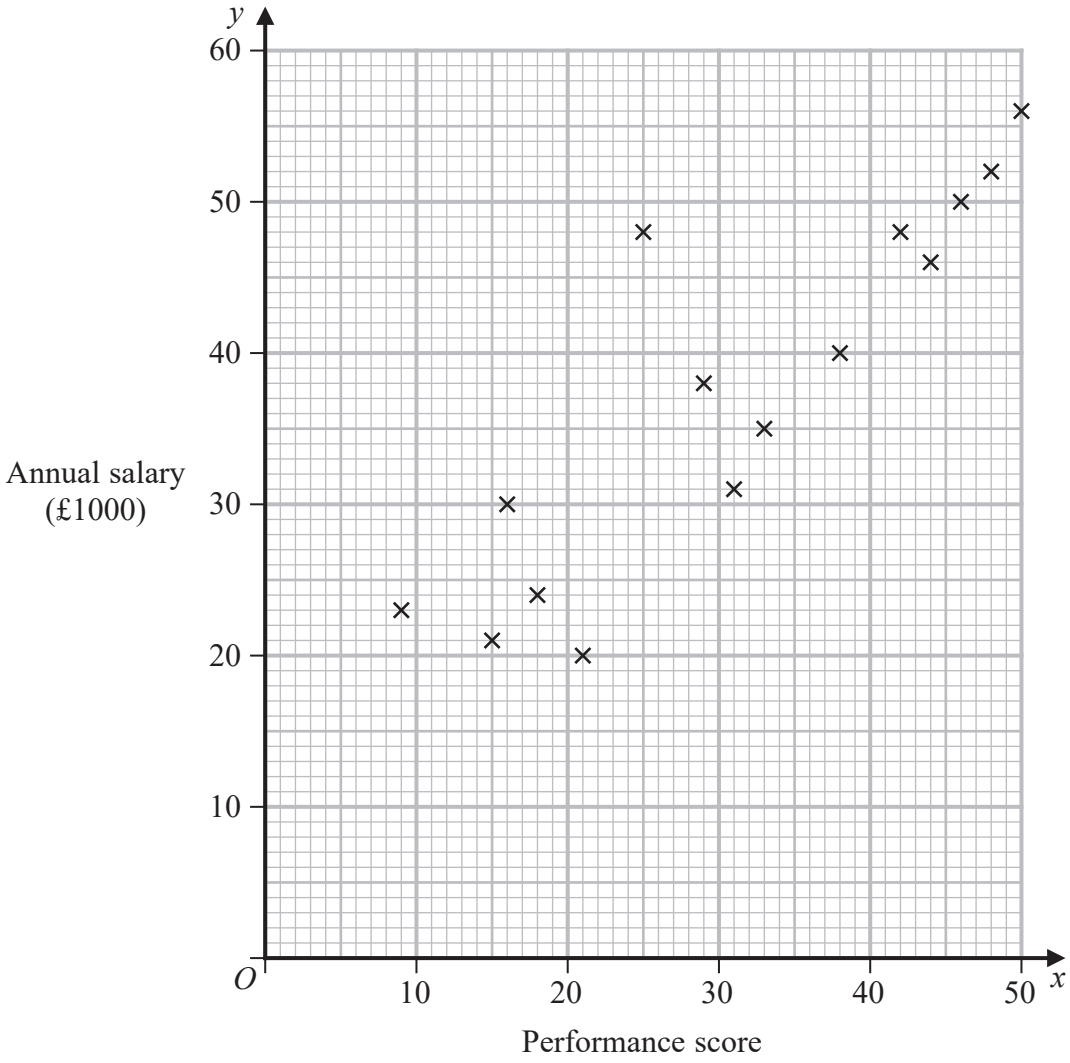


Figure 1

Turn over for a spare copy of the scatter diagram if you need to redraw your line.

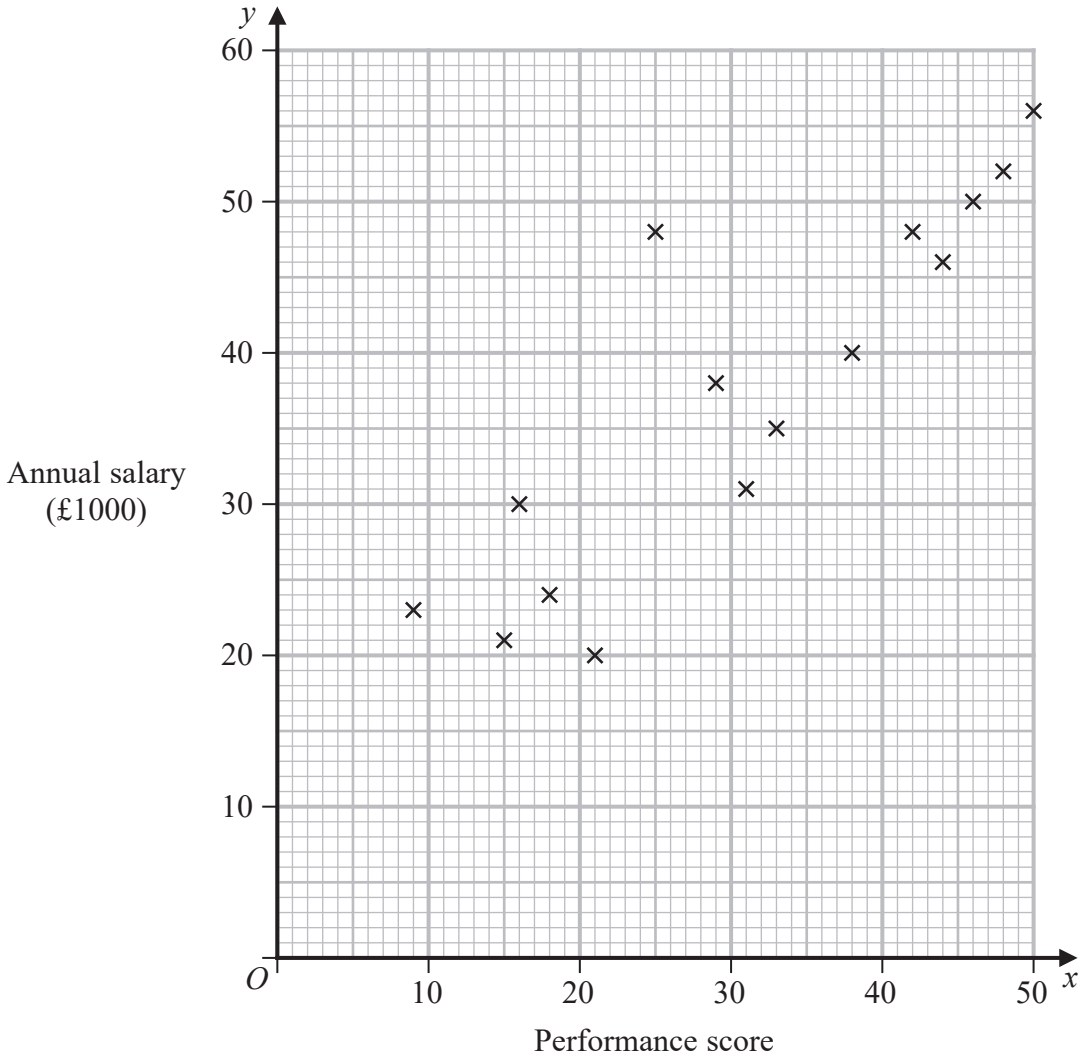
Leave blank

Question 5 continued

Leave blank

Question 5 continued

Only use this scatter diagram if you need to redraw your line.



Q5

(Total 17 marks)

Leave blank

6. A disc of radius 1 cm is rolled onto a horizontal grid of rectangles so that the disc is equally likely to land anywhere on the grid. Each rectangle is 5 cm long and 3 cm wide. There are no gaps between the rectangles and the grid is sufficiently large so that no discs roll off the grid.

If the disc lands inside a rectangle without covering any part of the edges of the rectangle then a prize is won.

By considering the possible positions for the centre of the disc,

(a) show that the probability of winning a prize on any particular roll is $\frac{1}{5}$

(3)

A group of 15 students each roll the disc onto the grid twenty times and record the number of times, x, that each student wins a prize. Their results are summarised as follows

$$\sum x = 61 \qquad \sum x^2 = 295$$

(b) Find the standard deviation of the number of prizes won per student.

(2)

A second group of 12 students each roll the disc onto the grid twenty times and the mean number of prizes won per student is 3.5 with a standard deviation of 2

(c) Find the mean and standard deviation of the number of prizes won per student for the whole group of 27 students.

(7)

The 27 students also recorded the number of times that the disc covered a corner of a rectangle and estimated the probability to be 0.2216 (to 4 decimal places).

(d) Explain how this probability could be used to find an estimate for the value of π and state the value of your estimate.

(3)

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Q6

(Total 15 marks)

TOTAL FOR PAPER IS 75 MARKS

END

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel
International Advanced Level

Centre Number | Candidate Number

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

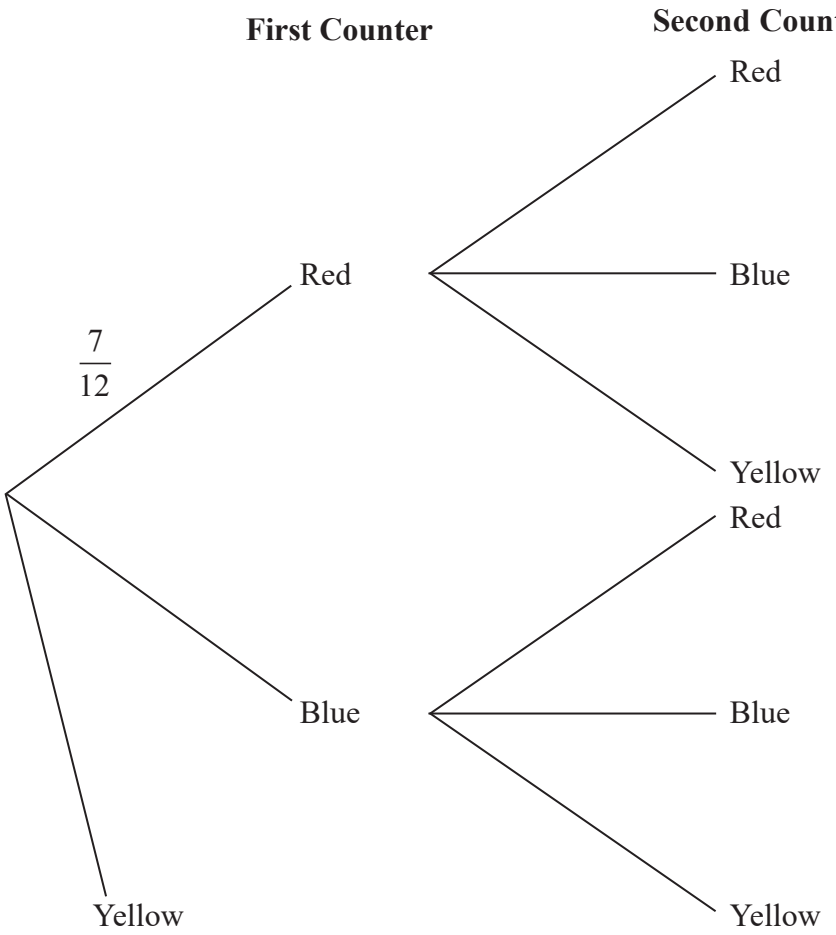
- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.
- If you change your mind about an answer, cross it out and put your new answer and any working underneath.
- Good luck with your examination.

Turn over ▶

Leave blank

1. There are 7 red counters, 3 blue counters and 2 yellow counters in a bag. Gina selects a counter at random from the bag and keeps it. If the counter is yellow she does not select any more counters. If the counter is not yellow she randomly selects a second counter from the bag.

(a) Complete the tree diagram.



(2)

Given that Gina has selected a yellow counter,

(b) find the probability that she has 2 counters.

(3)

Leave blank

Question 1 continued

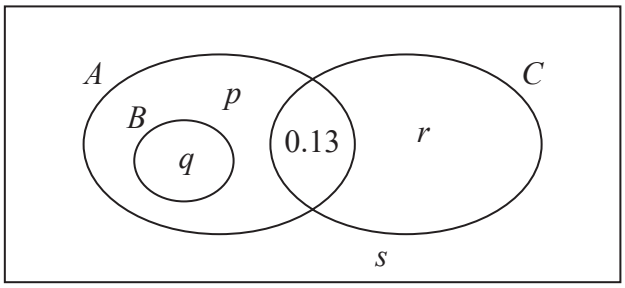
Q1

(Total 5 marks)

Leave blank

2. In the Venn diagram below, A, B and C are events and p, q, r and s are probabilities.

The events A and C are independent and $P(A) = 0.65$



(a) State which two of the events A, B and C are mutually exclusive.

(1)

(b) Find the value of r and the value of s.

(5)

The events $(A \cap C')$ and $(B \cup C)$ are also independent.

(c) Find the exact value of p and the exact value of q. Give your answers as fractions.

(6)

Leave blank

Question 2 continued

Q2

(Total 12 marks)

Leave blank

3. A random sample of 100 carrots is taken from a farm and their lengths, L cm, recorded. The data are summarised in the following table.

Length, L cm	Frequency, f	Class mid point, x cm
$5 \leqslant L < 8$	5	6.5
$8 \leqslant L < 10$	13	9
$10 \leqslant L < 12$	16	11
$12 \leqslant L < 15$	25	13.5
$15 \leqslant L < 20$	30	17.5
$20 \leqslant L < 28$	11	24

A histogram is drawn to represent these data.
The bar representing the class $5 \leqslant L < 8$ is 1.5 cm wide and 1 cm high.

(a) Find the width and height of the bar representing the class $15 \leqslant L < 20$

(3)

(b) Use linear interpolation to estimate the median length of these carrots.

(2)

(c) Estimate

(i) the mean length of these carrots,

(2)

(ii) the standard deviation of the lengths of these carrots.

(3)

A supermarket will only buy carrots with length between 9 cm and 22 cm.

(d) Estimate the proportion of carrots from the farm that the supermarket will buy.

(2)

Any carrots that the supermarket does not buy are sold as animal feed.

The farm makes a profit of 2.2 pence on each carrot sold to the supermarket, a profit of 0.8 pence on each carrot longer than 22 cm and a loss of 1.2 pence on each carrot shorter than 9 cm.

(e) Find an estimate of the mean profit per carrot made by the farm.

(2)

Leave blank

Question 3 continued

Leave blank

Question 3 continued

Leave blank

Question 3 continued

(Total 14 marks)

Q3

Leave blank

4. Kris works in the mailroom of a large company and is responsible for all the letters sent by the company. The weights of letters sent by the company, W grams, have a normal distribution with mean 165 g and standard deviation 35 g.

(a) Estimate the proportion of letters sent by the company that weigh less than 120 g.

(3)

Kris splits the letters to be sent into 3 categories: heavy, medium and light, with $\frac{1}{3}$ of the letters in each category.

(b) Find the weight limits that determine medium letters.

(4)

A heavy letter is chosen at random.

(c) Find the probability that this letter weighs less than 200 g.

(3)

Kris chooses a random sample of 3 letters from those in the mailroom one day.

(d) Find the probability that there is one letter in each of the 3 categories.

(3)

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Leave blank

Question 4 continued

(Total 13 marks)

Q4

Leave blank

5. The discrete random variable X has the following probability distribution

x	-2	-1	0	1	4
$P(X = x)$	a	b	c	b	a

Given that $E(X) = 0.5$

(a) find the value of a.

(2)

Given also that $\text{Var}(X) = 5.01$

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

(5)

The random variable $Y = 5 - 8X$

(c) Find (i) $E(Y)$

(ii) $\text{Var}(Y)$

(3)

(d) Find $P(4X^2 > Y)$

(5)

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Leave blank

Question 5 continued

(Total 15 marks)

Q5

Leave blank

6. Two economics students, Andi and Behrouz, are studying some data relating to unemployment, x%, and increase in wages, y%, for a European country. The least squares regression line of y on x has equation

$$y = 3.684 - 0.3242x$$

and $\sum y = 23.7$ $\sum y^2 = 42.63$ $\sum x^2 = 756.81$ $n = 16$

(a) Show that $S_{yy} = 7.524\,375$

(1)

(b) Find S_{xx}

(4)

(c) Find the product moment correlation coefficient between x and y.

(3)

Behrouz claims that, assuming the model is valid, the data show that when unemployment is 2% wages increase at over 3%

(d) Explain how Behrouz could have come to this conclusion.

(1)

Andi uses the formula

$$\text{range} = \text{mean} \pm 3 \times \text{standard deviation}$$

to estimate the range of values for x.

(e) Find estimates of the minimum value and the maximum value of x in these data using Andi's formula.

(3)

(f) Comment, giving a reason, on the reliability of Behrouz's claim.

(2)

Andi suggests using the regression line with equation $y = 3.684 - 0.3242x$ to estimate unemployment when wages are increasing at 2%

(g) Comment, giving a reason, on Andi's suggestion.

(2)

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Q6

(Total 16 marks)

TOTAL FOR PAPER: 75 MARKS

END

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

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.*

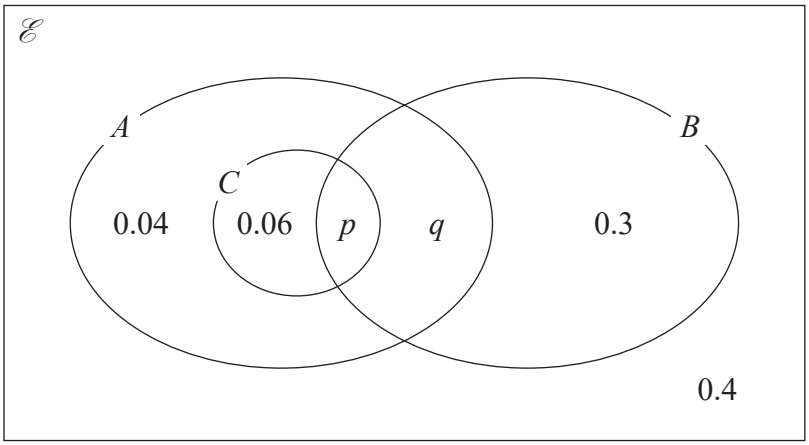
Advice

- 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 ▶

Leave blank

1. The Venn diagram shows the events A, B and C and their associated probabilities, where p and q are probabilities.



(a) Find $\mathrm{P}(B)$

(1)

(b) Determine whether or not A and B are independent.

(2)

Given that $\mathrm{P}(C \mid B) = \mathrm{P}(C)$

(c) find the value of p and the value of q

(3)

The event D is such that

- A and D are mutually exclusive
- $\mathrm{P}(B \cap D) > 0$

(d) On the Venn diagram show a possible position for the event D

(1)

Leave blank

Question 1 continued

Q1

(Total 7 marks)

Leave blank

2. A large company is analysing how much money it spends on paper in its offices each year. The number of employees in the office, x, and the amount spent on paper in a year, p (\$ hundreds), in each of 12 randomly selected offices were recorded.

The results are summarised in the following statistics.

$$\sum x = 93 \qquad S_{xx} = 148.25 \qquad \sum p = 273 \qquad \sum p^2 = 6602.72 \qquad \sum xp = 2347$$

(a) Show that $S_{xp} = 231.25$

(1)

(b) Find the product moment correlation coefficient for these data.

(3)

(c) Find the equation of the regression line of p on x in the form $p = a + bx$

(4)

(d) Give an interpretation of the gradient of your regression line.

(1)

The director of the company wants to reduce the amount spent on paper each year.

He wants each office to aim for a model of the form $p = \frac{4}{5}a + \frac{1}{2}bx$, where a and b are the values found in part (c).

Using the data for the 93 employees from the 12 offices,

(e) estimate the percentage saving in the amount spent on paper each year by the company using the director's model.

(3)

Leave blank

Question 2 continued

Leave blank

Question 2 continued

Leave blank

Question 2 continued

(Total 12 marks)

Q2

Leave blank

3. The stem and leaf diagram shows the ages of the 35 male passengers on a cruise.

Age **Key:** 1 | 3 represents an age of 13 years

1	3	(1)
2	7 9	(2)
3	1 2 8 8	(4)
4	5 5 6 7 8 8 9	(7)
5	2 2 3 3 4 4 5 6 6 8	(10)
6	0 1 1 4 4 4 7	(7)
7	3 6	(2)
8	7 8	(2)

(a) Find the median age of the male passengers.

(1)

(b) Show that the interquartile range (IQR) of these ages is 16

(2)

An outlier is defined as a value that is more than

$1.5 \times$ IQR above the upper quartile

or

$1.5 \times$ IQR below the lower quartile

(c) Show that there are 3 outliers amongst these ages.

(3)

(d) On the grid in Figure 1 on page 9, draw a box plot for the ages of the male passengers on the cruise.

(4)

Figure 1 on page 9 also shows a box plot for the ages of the female passengers on the cruise.

(e) Comment on any difference in the distributions of ages of male and female passengers on the cruise.
State the values of any statistics you have used to support your comment.

(1)

Anja, along with her 2 daughters and a granddaughter, now join the cruise.
Anja's granddaughter is younger than both of Anja's daughters.
Anja had her 23rd birthday on the day her eldest daughter was born.
When their 4 ages are included with the other female passengers on the cruise, the box plot does not change.

(f) State, giving reasons, what you can say about
(i) the granddaughter's age
(ii) Anja's age.

(3)

Leave blank

Question 3 continued

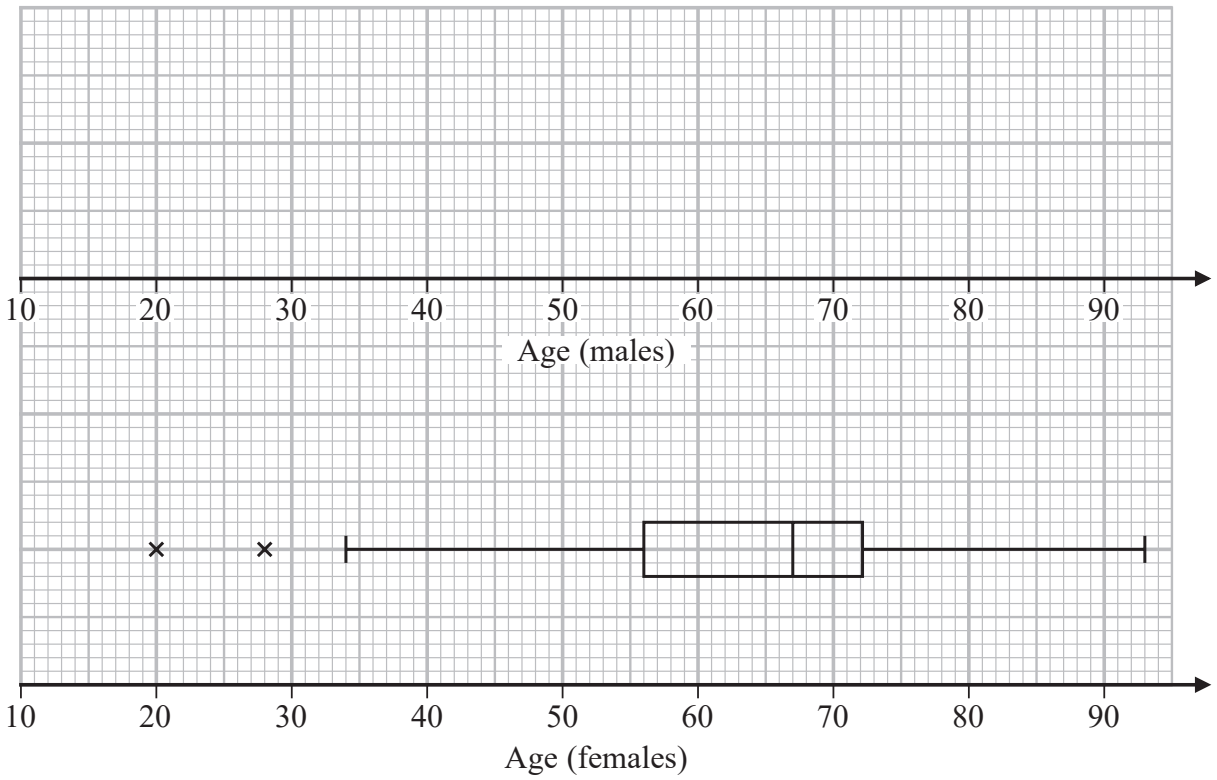


Figure 1

Turn over for a spare grid if you need to redraw your box plot.

Leave blank

Question 3 continued

Leave blank

Question 3 continued

Only use this grid if you need to redraw your box plot.

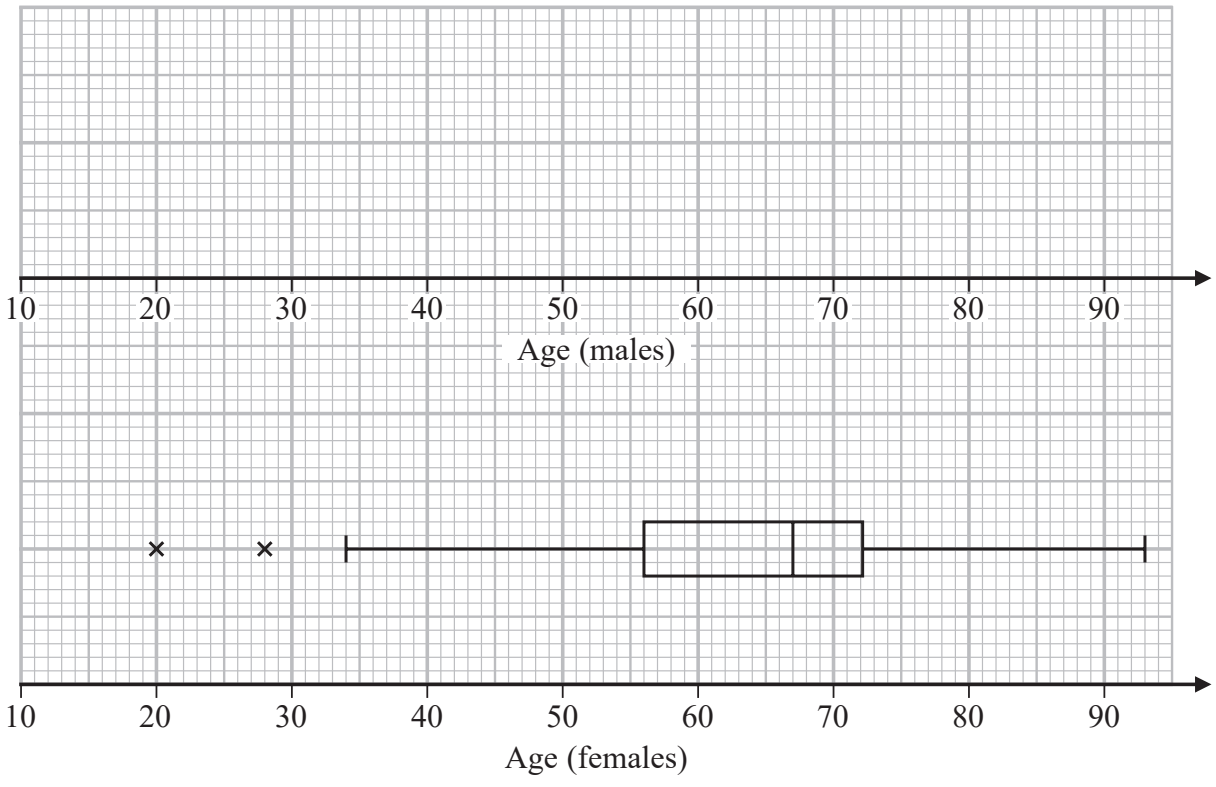


Figure 1

(Total 14 marks)

Q3

Leave blank

4. Three bags **A**, **B** and **C** each contain coloured balls.

Bag **A** contains 4 red balls and 2 yellow balls only.
Bag **B** contains 4 red balls and 1 yellow ball only.
Bag **C** contains 6 red balls only.

In a game
Mike takes a ball at random from bag **A**, records the colour and places it in bag **C**.
He then takes a ball at random from bag **B**, records the colour and places it in bag **C**.
Finally, Mike takes a ball at random from bag **C** and records the colour.

(a) Complete the tree diagram on the page opposite, to illustrate the game by adding the remaining branches and all probabilities.

(3)

(b) Show that the probability that Mike records a yellow ball exactly twice is $\frac{1}{10}$

(3)

Given that Mike records exactly 2 yellow balls,

(c) find the probability that the ball drawn from bag **A** is red.

(2)

Mike plays this game a large number of times, each time starting with the bags containing balls as described above. The random variable X represents the number of yellow balls recorded in a single game.

(d) Find the probability distribution of X

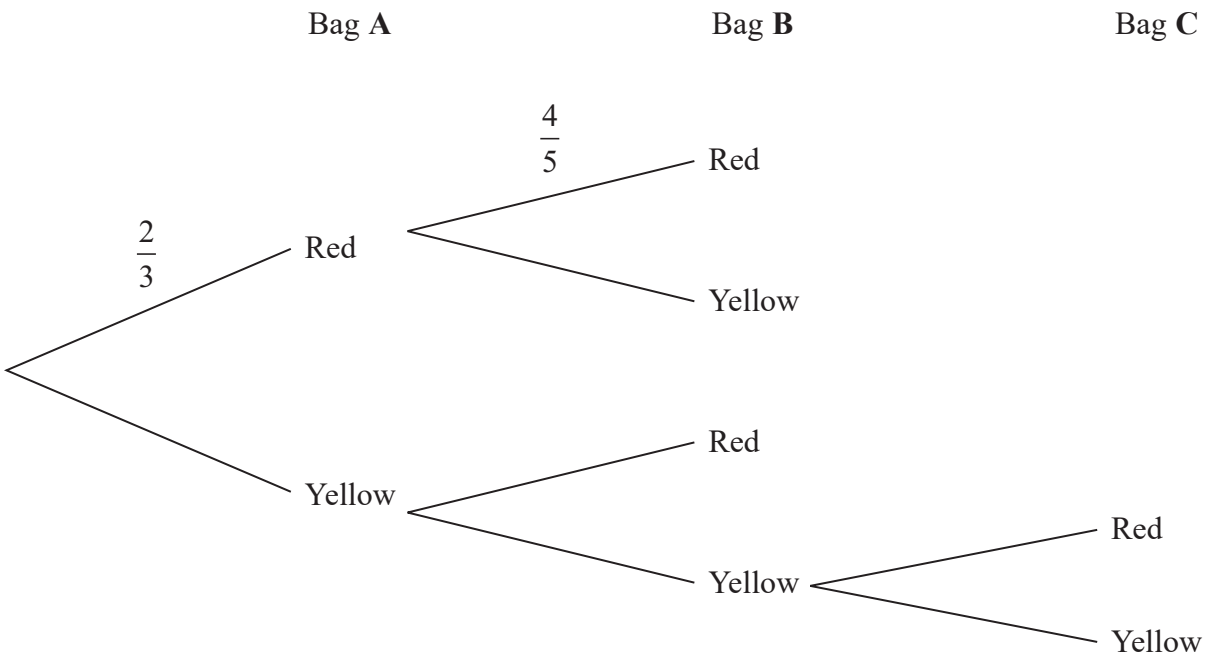
(3)

(e) Find $\mathrm{E}(X)$

(2)

Leave blank

Question 4 continued



Leave blank

Question 4 continued

Leave blank

Question 4 continued

(Total 13 marks)

Q4

Leave blank

5. The discrete random variable Y has the following probability distribution

y	-9	-5	0	5	9
$P(Y = y)$	q	r	u	r	q

where q, r and u are probabilities.

(a) Write down the value of $E(Y)$

(1)

The cumulative distribution function of Y is $F(y)$

Given that $F(0) = \frac{19}{30}$

(b) show that the value of u is $\frac{4}{15}$

(3)

Given also that $\text{Var}(Y) = 37$

(c) find the value of q and the value of r

(4)

The coordinates of a point P are $(12, Y)$

The random variable D represents the length of OP

(d) Find the probability distribution of D

(6)

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Leave blank

Question 5 continued

(Total 14 marks)

Q5

Leave blank

6. Xiang is designing shelves for a bookshop. The height, H cm, of books is modelled by the normal distribution with mean 25.1 cm and standard deviation 5.5 cm

(a) Show that $P(H > 30.8) = 0.15$

(3)

Xiang decided that the smallest 5% of books and books taller than 30.8 cm would not be placed on the shelves. All the other books will be placed on the shelves.

(b) Find the range of heights of books that will be placed on the shelves.

(3)

The books that will be placed on the shelves have heights classified as small, medium or large.
The numbers of small, medium and large books are in the ratios 2 : 3 : 3

(c) The medium books have heights x cm where $m < x < d$

(i) Show that $d = 25.8$ to 1 decimal place.

(3)

(ii) Find the value of m

(4)

Xiang wants 2 shelves for small books, 3 shelves for medium books and 3 shelves for large books.
These shelves will be placed one above another and made of wood that is 1 cm thick.

(d) Work out the minimum total height needed.

(2)

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Q6

(Total 15 marks)

TOTAL FOR PAPER: 75 MARKS

END

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

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 7 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.

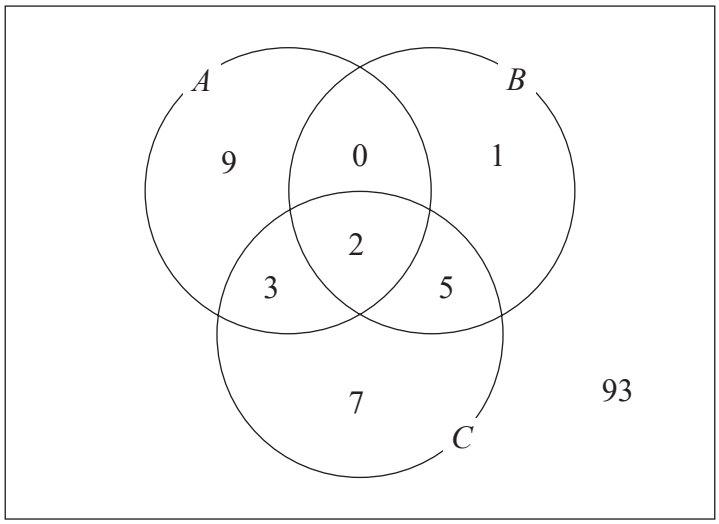
Turn over ▶

Leave blank

1. A factory produces shoes.

A quality control inspector at the factory checks a sample of 120 shoes for each of three types of defect. The Venn diagram represents the inspector's results.

A represents the event that a shoe has defective stitching
B represents the event that a shoe has defective colouring
C represents the event that a shoe has defective soles



One of the shoes in the sample is selected at random.

(a) Find the probability that it does **not** have defective soles.

(1)

(b) Find $P(A \cap B \cap C')$

(1)

(c) Find $P(A \cup B \cup C')$

(2)

(d) Find the probability that the shoe has at most one type of defect.

(2)

(e) Given the selected shoe has at most one type of defect, find the probability it has defective stitching.

(2)

The random variable X is the number of the events A, B, C that occur for a randomly selected shoe.

(f) Find $E(X)$

(3)

Leave blank

Question 1 continued

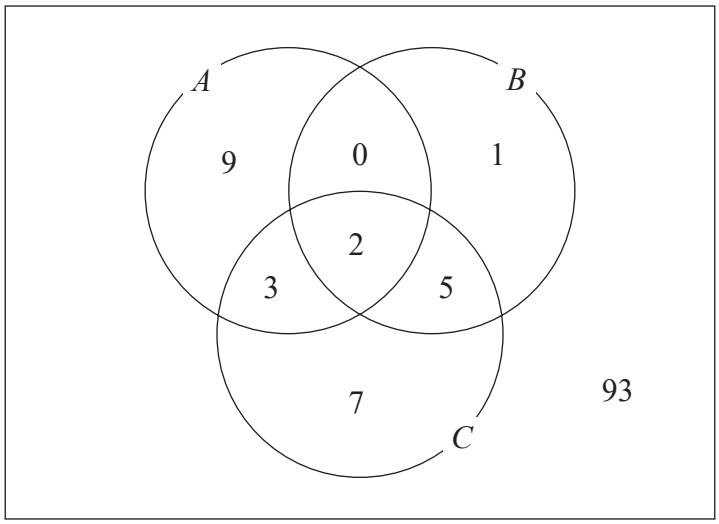
Leave blank

Question 1 continued

Leave blank

Question 1 continued

This is a copy of the Venn diagram for this question.



Q1

(Total 11 marks)

Leave blank

2. Tom's car holds 50 litres of petrol when the fuel tank is full.

For each of 10 journeys, each starting with 50 litres of petrol in the fuel tank, Tom records the distance travelled, d kilometres, and the amount of petrol used, p litres.

The summary statistics for the 10 journeys are given below.

$$\sum d = 1029 \quad \sum p = 50.8 \quad \sum dp = 5240.8 \quad S_{dd} = 344.9 \quad S_{pp} = 0.576$$

(a) Calculate the product moment correlation coefficient between d and p

(3)

The amount of petrol remaining in the fuel tank for each journey, w litres, is recorded.

(b) (i) Write down an equation for w in terms of p

(ii) Hence, write down the value of the product moment correlation coefficient between w and p

(2)

(c) Write down the value of the product moment correlation coefficient between d and w

(1)

Leave blank

Question 2 continued

Leave blank

Question 2 continued

Leave blank

Question 2 continued

Q2

(Total 6 marks)

Leave blank

3. The stem and leaf diagram shows the number of deliveries made by Pat each day for 24 days

Key: 10 | 8 represents 108 deliveries

10	8	9										(2)
11	0	3	6	6	6	8	8	9	9	9	9	(11)
12	4	5	5	5	5	5	5	8				(8)
13	a	b	c									(3)

where a, b and c are positive integers with $a < b < c$

An outlier is defined as any value greater than $1.5 \times$ interquartile range above the upper quartile.

Given that there is only one outlier for these data,

(a) show that $c = 9$

(3)

The number of deliveries made by Pat each day is represented by d

The data in the stem and leaf diagram are coded using

$$x = d - 125$$

and the following summary statistics are obtained

$$\sum x = -96 \quad \text{and} \quad \sum (x - \bar{x})^2 = 1306$$

(b) Find the mean number of deliveries.

(3)

(c) Find the standard deviation of the number of deliveries.

(2)

One of these 24 days is selected at random. The random variable D represents the number of deliveries made by Pat on this day.

The random variable $X = D - 125$

(d) Find $P(D > 118 | X < 0)$

(2)

Leave blank

Question 3 continued

Leave blank

Question 3 continued

Leave blank

Question 3 continued

(Total 10 marks)

Q3

Leave blank

4. The random variable W has a discrete uniform distribution where

$$P(W = w) = \frac{1}{5} \quad \text{for } w = 1, 2, 3, 4, 5$$

(a) Find $P(2 \leqslant W < 3.5)$

(1)

The discrete random variable $X = 5 - 2W$

(b) Find $E(X)$

(3)

(c) Find $P(X < W)$

(2)

The discrete random variable $Y = \frac{1}{W}$

(d) Find

(i) the probability distribution of Y

(ii) $\text{Var}(Y)$, showing your working.

(5)

(e) Find $\text{Var}(2 - 3Y)$

(2)

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Leave blank

Question 4 continued

(Total 13 marks)

Q4

Leave blank

5. Jia writes a computer program that randomly generates values from a normal distribution.

He sets the mean as 40 and the standard deviation as 2.4

(a) Find the probability that a particular value generated by the computer program is less than 37

(3)

Jia changes the mean to m but leaves the standard deviation as 2.4

The computer program then randomly generates 2 independent values from this normal distribution.

The probability that both of these values are greater than 32 is 0.16

(b) Find the value of m, giving your answer to 2 decimal places.

(4)

Jia now changes the mean to 4 and the standard deviation to 8

The computer program then randomly generates 5 independent values from this normal distribution.

(c) Find the probability that at least one of these values is negative.

(4)

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Leave blank

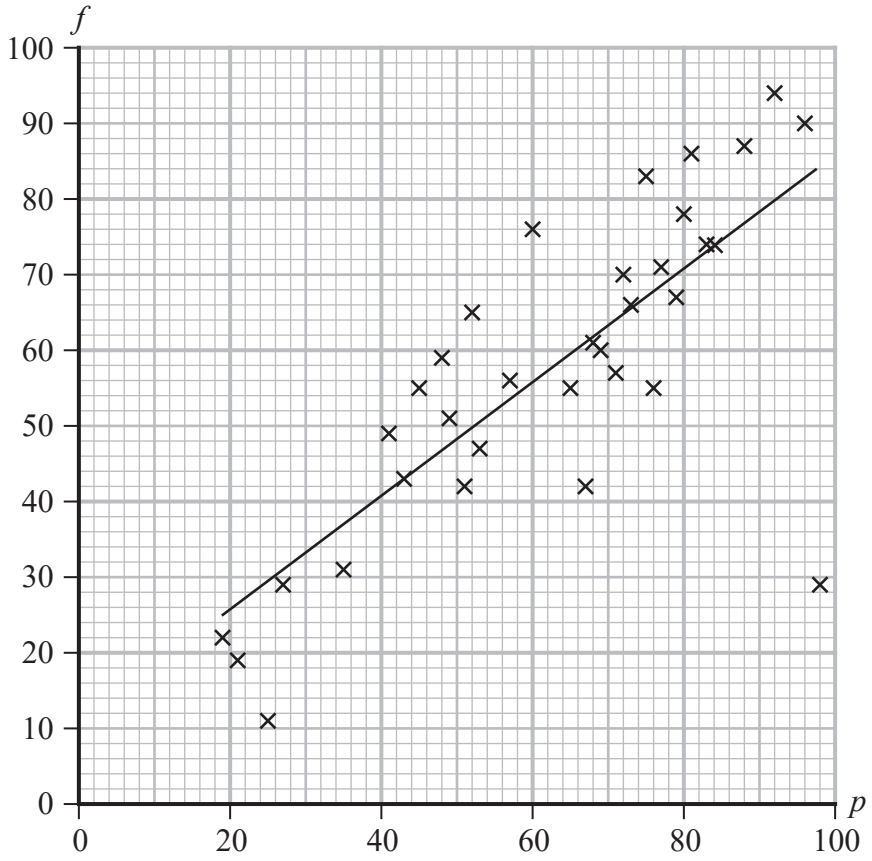
Question 5 continued

(Total 11 marks)

Q5

Leave blank

6. Students on a psychology course were given a pre-test at the start of the course and a final exam at the end of the course. The teacher recorded the number of marks achieved on the pre-test, p, and the number of marks achieved on the final exam, f, for 34 students and displayed them on the scatter diagram.



The equation of the least squares regression line for these data is found to be

$$f = 10.8 + 0.748p$$

For these students, the mean number of marks on the pre-test is 62.4

(a) Use the regression model to find the mean number of marks on the final exam.

(2)

(b) Give an interpretation of the gradient of the regression line.

(1)

Considering the equation of the regression line, Priya says that she would expect someone who scored 0 marks on the pre-test to score 10.8 marks on the final exam.

(c) Comment on the reliability of Priya's statement.

(1)

(d) Write down the number of marks achieved on the final exam for the student who exceeded the expectation of the regression model by the largest number of marks.

(1)

Question 6 continues on page 24.

Leave blank

Question 6 continued

Question 6 continues on page 24.

Leave blank

Question 6 continued

(e) Find the range of values of p for which this regression model, $f = 10.8 + 0.748p$, predicts a greater number of marks on the final exam than on the pre-test.

(3)

Later the teacher discovers an error in the recorded data. The student who achieved a score of 98 on the pre-test, scored 92 not 29 on the final exam.

The summary statistics used for the model $f = 10.8 + 0.748p$ are corrected to include this information and a new least squares regression line is found.

Given the **original** summary statistics were,

$$n = 34 \quad \sum p = 2120 \quad \sum pf = 133\,486 \quad S_{pp} = 15\,573.76 \quad S_{pf} = 11\,648.35$$

(f) calculate the gradient of the new regression line. Show your working clearly.

(5)

Leave blank

Question 6 continued

(Total 13 marks)

Q6

Leave blank

7. A bag contains n marbles of which 7 are green.

From the bag, 3 marbles are selected at random.

The random variable X represents the number of green marbles selected.

The cumulative distribution function of X is given by

x	0	1	2	3
F(x)	a	b	$\frac{37}{38}$	1

(a) Show that $n(n-1)(n-2) = 7980$

(4)

(b) Verify that $n = 21$ satisfies the equation in part (a).

(1)

Given that $n = 21$

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

(6)

Leave blank

Question 7 continued

Leave blank

Question 7 continued

Q7

(Total 11 marks)

TOTAL FOR PAPER: 75 MARKS

END

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

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 ▶

1. The company *Seafield* requires contractors to record the number of hours they work each week. A random sample of 38 weeks is taken and the number of hours worked per week by contractor Kiana is summarised in the stem and leaf diagram below.

Stem	Leaf											
1	4	4	4	5	5	5	6	6	9	9	9	(11)
2	1	2	2	3	3	4	4	4	w	9		(10)
3	2	3	4	4	5	6	7	7	7	9		(10)
4	1	1	2	3								(4)
5	1	9										(2)
6	4											(1)

Key : 3|2 means 32

The quartiles for this distribution are summarised in the table below.

Q_1	Q_2	Q_3
x	26	y

(a) Find the values of w, x and y

(3)

Kiana is looking for outliers in the data. She decides to classify as outliers any observations greater than

$$Q_3 + 1.0 \times (Q_3 - Q_1)$$

(b) Showing your working clearly, identify any outliers that Kiana finds.

(2)

(c) Draw a box plot for these data in the space provided on the grid opposite.

(3)

(d) Use the formula

$$\text{skewness} = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}$$

to find the skewness of these data. Give your answer to 2 significant figures.

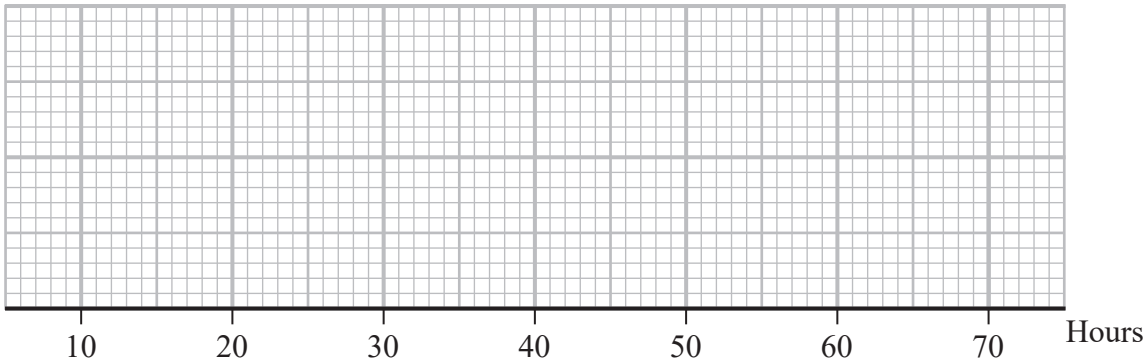
(2)

Kiana's new employer, *Landacre*, wishes to know the average number of hours per week she worked during her employment at *Seafield* to help calculate the cost of employing her.

(e) Explain why *Landacre* might prefer to know Kiana's mean, rather than median, number of hours worked per week.

(1)

Question 1 continued

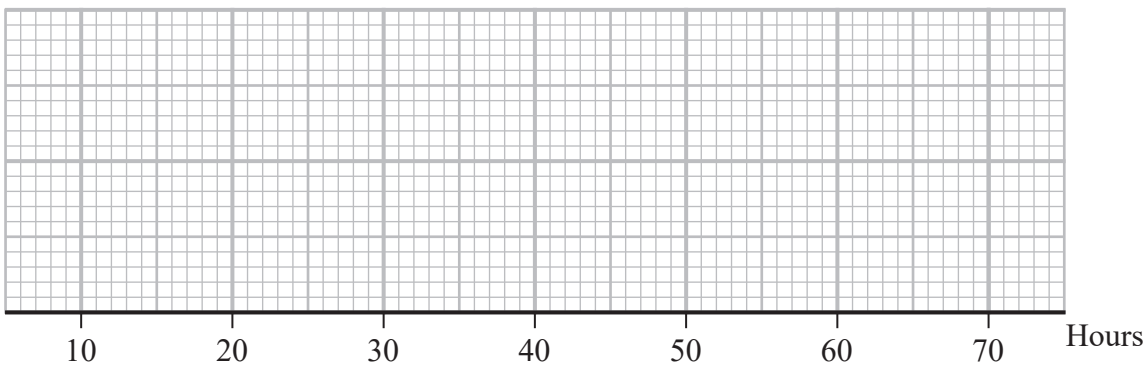


Turn over for a spare grid if you need to redraw your box plot.

Question 1 continued

Question 1 continued

Only use this grid if you need to redraw your box plot.



(Total for Question 1 is 11 marks)

2. Stuart is investigating the relationship between Gross Domestic Product (GDP) and the size of the population for a particular country.
He takes a random sample of 9 years and records the size of the population, t millions, and the GDP, g billion dollars for each of these years.

The data are summarised as

$$n = 9 \quad \sum t = 7.87 \quad \sum g = 144.84 \quad \sum g^2 = 3624.41 \quad S_{tt} = 1.29 \quad S_{tg} = 40.25$$

(a) Calculate the product moment correlation coefficient between t and g

(3)

(b) Give an interpretation of your product moment correlation coefficient.

(1)

(c) Find the equation of the least squares regression line of g on t in the form $g = a + bt$

(4)

(d) Give an interpretation of the value of b in your regression line.

(1)

(e) (i) Use the regression line from part (c) to estimate the GDP, in billions of dollars, for a population of 7 000 000

(2)

(ii) Comment on the reliability of your answer in part (i). Give a reason, in context, for your answer.

(1)

Using the regression line from part (c), Stuart estimates that for a population increase of x million there will be an increase of 0.1 billion dollars in GDP.

(f) Find the value of x

(2)

Question 2 continued

Question 2 continued

Question 2 continued

(Total for Question 2 is 14 marks)

3. Gill buys a bag of logs to use in her stove. The lengths, l cm, of the 88 logs in the bag are summarised in the table below.

Length (l)	Frequency (f)
$15 < l \leqslant 20$	19
$20 < l \leqslant 25$	35
$25 < l \leqslant 27$	16
$27 < l \leqslant 30$	15
$30 < l \leqslant 40$	3

A histogram is drawn to represent these data.
The bar representing logs with length $27 < l \leqslant 30$ has a width of 1.5 cm and a height of 4 cm.

(a) Calculate the width and height of the bar representing log lengths of $20 < l \leqslant 25$

(3)

(b) Use linear interpolation to estimate the median of l

(2)

The maximum length of log Gill can use in her stove is 26 cm.
Gill estimates, using linear interpolation, that x logs from the bag will fit into her stove.

(c) Show that $x = 62$

(1)

Gill randomly selects 4 logs from the bag.

(d) Using $x = 62$, find the probability that all 4 logs will fit into her stove.

(2)

The weights, W grams, of the logs in the bag are coded using $y = 0.5w - 255$ and summarised by

$$n = 88 \qquad \sum y = 924 \qquad \sum y^2 = 12\,862$$

(e) Calculate

(i) the mean of W

(3)

(ii) the variance of W

(3)

Question 3 continued

Question 3 continued

Question 3 continued

(Total for Question 3 is 14 marks)

4. The events H and W are such that

$$P(H) = \frac{3}{8} \qquad P(H \cup W) = \frac{3}{4}$$

Given that H and W are independent,

(a) show that $P(W) = \frac{3}{5}$

(4)

The event N is such that

$$P(N) = \frac{1}{15} \qquad P(H \cap N) = P(N)$$

(b) Find $P(N' \mid H)$

(2)

Given that W and N are mutually exclusive,

(c) draw a Venn diagram to represent the events H, W and N giving the exact probabilities of each region in the Venn diagram.

(5)

Question 4 continued

Question 4 continued

Question 4 continued

(Total for Question 4 is 11 marks)

5. A red spinner is designed so that the score R is given by the following probability distribution.

r	2	3	4	5	6
$P(R = r)$	0.25	0.3	0.15	0.1	0.2

(a) Show that $E(R^2) = 15.8$

(1)

Given also that $E(R) = 3.7$

(b) find the standard deviation of R, giving your answer to 2 decimal places.

(2)

A yellow spinner is designed so that the score Y is given by the probability distribution in the table below. The cumulative distribution function $F(y)$ is also given.

y	2	3	4	5	6
$P(Y = y)$	0.1	0.2	0.1	a	b
$F(y)$	0.1	0.3	0.4	c	d

(c) Write down the value of d

(1)

Given that $E(Y) = 4.55$

(d) find the value of c

(5)

Pabel and Jessie play a game with these two spinners.
Pabel uses the red spinner.
Jessie uses the yellow spinner.
They take turns to spin their spinner.
The winner is the first person whose spinner lands on the number 2 and the game ends.
Jessie spins her spinner first.

(e) Find the probability that Jessie wins on her second spin.

(2)

(f) Calculate the probability that, in a game, the score on Pabel's first spin is the same as the score on Jessie's first spin.

(3)

Question 5 continued

Question 5 continued

Question 5 continued

(Total for Question 5 is 14 marks)

6. A manufacturer fills bottles with oil.
The volume of oil in a bottle, V ml, is normally distributed with $V \sim \text{N}(100, 2.5^2)$

(a) Find $\text{P}(V > 104.9)$

(3)

(b) In a pack of 150 bottles, find the expected number of bottles containing more than 104.9 ml

(2)

(c) Find the value of v, to 2 decimal places, such that $\text{P}(V > v \,|\, V < 104.9) = 0.2801$

(6)

Question 6 continued

Question 6 continued

(Total for Question 6 is 11 marks)

TOTAL FOR PAPER: 75 MARKS

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

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 7 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 ▶

1. The stem lengths of a sample of 120 tulips are recorded in the grouped frequency table below.

Stem length (cm)	Frequency
$40 \leqslant x < 42$	12
$42 \leqslant x < 45$	18
$45 \leqslant x < 50$	23
$50 \leqslant x < 55$	35
$55 \leqslant x < 58$	24
$58 \leqslant x < 60$	8

A histogram is drawn to represent these data.

The area of the bar representing the $40 \leqslant x < 42$ class is $16.5\,\text{cm}^2$

(a) Calculate the exact area of the bar representing the $42 \leqslant x < 45$ class.

(2)

The height of the tallest bar in the histogram is 10 cm.

(b) Find the exact height of the second tallest bar.

(3)

Q_1 for these data is 45 cm.

(c) Use linear interpolation to find an estimate for

(i) Q_2

(ii) the interquartile range.

(4)

One measure of skewness is given by

$$\frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

(d) By calculating this measure, describe the skewness of these data.

(2)

Question 1 continued

Question 1 continued

Question 1 continued

(Total for Question 1 is 11 marks)

2. The production cost, £c million, of a film and the total ticket sales, £t million, earned by the film are recorded for a sample of 40 films.

Some summary statistics are given below.

$$\sum c = 1634 \qquad \sum t = 1361 \qquad \sum t^2 = 82\,873 \qquad \sum ct = 83\,634 \qquad S_{cc} = 28\,732.1$$

(a) Find the exact value of S_{tt} and the exact value of S_{ct}

(3)

(b) Calculate the value of the product moment correlation coefficient for these data.

(2)

(c) Give an interpretation of your answer to part (b)

(1)

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

$$t = -5.84 + 0.976c$$

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

(3)

(e) Find the expected total ticket sales for a film with a production cost of £90 million.

(2)

Using the regression line in part (d)

(f) find the range of values of the production cost of a film for which the total ticket sales are less than 80% of its production cost.

(2)

Question 2 continued

Question 2 continued

Question 2 continued

(Total for Question 2 is 13 marks)

3. Morgan is investigating the body length, b centimetres, of squirrels.

A random sample of 8 squirrels is taken and the data for each squirrel is coded using

$$x = \frac{b - 21}{2}$$

The results for the coded data are summarised below

$$\sum x = -1.2 \qquad \sum x^2 = 5.1$$

(a) Find the mean of b

(3)

(b) Find the standard deviation of b

(3)

A 9th squirrel is added to the sample. Given that for all 9 squirrels $\sum x = 0$

(c) find

(i) the body length of the 9th squirrel,

(2)

(ii) the standard deviation of x for all 9 squirrels.

(2)

Question 3 continued

Question 3 continued

Question 3 continued

(Total for Question 3 is 10 marks)

4. The cumulative distribution function of the discrete random variable W, which takes only the values 6, 7 and 8, is given by

$$\mathrm{F}(W) = \frac{(w+3)(w-1)}{77} \quad \text{for } w = 6, 7, 8$$

Find $\mathrm{E}(W)$

(4)

Question 4 continued

(Total for Question 4 is 4 marks)

5. The weights, W grams, of kiwi fruit grown on a farm are normally distributed with mean 80 grams and standard deviation 8 grams.

The table shows the classifications of the kiwi fruit by their weight, where k is a positive constant.

Small		**Large**		
Tiny	**Petite**	**Extra**	**Jumbo**	**Mega**
$w < 66$	$66 \leqslant w < 70$	$70 \leqslant w < 80$	$80 \leqslant w < k$	$w \geqslant k$

One kiwi fruit is selected at random from those grown on the farm.

(a) Find the probability that this kiwi fruit is Large.

(3)

35% of the kiwi fruit are Jumbo.

(b) Find the value of k to one decimal place.

(4)

75% of Tiny kiwi fruit weigh more than y grams.

(c) Find the value of y giving your answer to one decimal place.

(5)

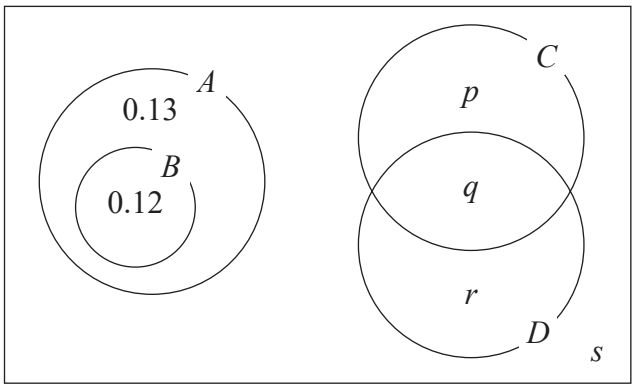
Question 5 continued

Question 5 continued

Question 5 continued

(Total for Question 5 is 12 marks)

6. The Venn diagram shows the events A, B, C and D, where p, q, r and s are probabilities.



(a) Write down the value of

(i) $P(A)$

(ii) $P(A|B)$

(iii) $P(A|C)$

(3)

Given that $P(B' \cap D') = \frac{7}{10}$ and $P(C|D) = \frac{3}{5}$

(b) find the exact value of q and the exact value of r

(6)

Given also that $P(B \cup C') = \frac{5}{8}$

(c) find the exact value of s

(2)

Question 6 continued

Question 6 continued

Question 6 continued

(Total for Question 6 is 11 marks)

7. Adana selects one number at random from the distribution of X which has the following probability distribution.

x	0	5	10
$P(X = x)$	0.1	0.2	0.7

(a) Given that the number selected by Adana is not 5, write down the probability it is 0

(1)

(b) Show that $E(X^2) = 75$

(1)

(c) Find $\text{Var}(X)$

(3)

(d) Find $\text{Var}(4 - 3X)$

(2)

Bruno and Charlie each independently select one number at random from the distribution of X

(e) Find the probability that the number Bruno selects is greater than the number Charlie selects.

(3)

Devika multiplies Bruno's number by Charlie's number to obtain a product, D

(f) Determine the probability distribution of D

(4)

Question 7 continued

Question 7 continued

Question 7 continued

Question 7 continued

(Total for Question 7 is 14 marks)

TOTAL FOR PAPER: 75 MARKS

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

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.

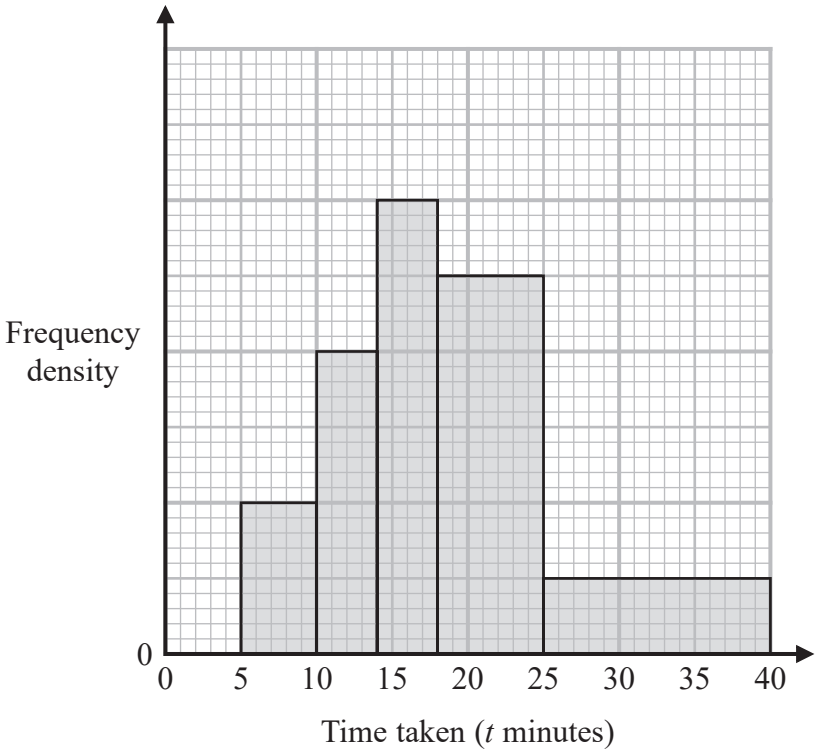
Information

- 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.

1. The histogram shows the times taken, t minutes, by each of 100 people to swim 500 metres.



(a) Use the histogram to complete the frequency table for the times taken by the 100 people to swim 500 metres.

Time taken (t minutes)	5 – 10	10 – 14	14 – 18	18 – 25	25 – 40
Frequency (f)	10	16	24		

(1)

(b) Estimate the number of people who took less than 16 minutes to swim 500 metres.

(2)

(c) Find an estimate for the mean time taken to swim 500 metres.

(2)

Given that $\sum ft^2 = 41\,033$

(d) find an estimate for the standard deviation of the times taken to swim 500 metres.

(2)

Given that $Q_3 = 23$

(e) use linear interpolation to estimate the interquartile range of the times taken to swim 500 metres.

(3)

Question 1 continued

Question 1 continued

Question 1 continued

(Total for Question 1 is 10 marks)

2. Two bags, ***X*** and ***Y***, each contain green marbles (G) and blue marbles (B) only.

- Bag ***X*** contains 5 green marbles and 4 blue marbles
- Bag ***Y*** contains 6 green marbles and 5 blue marbles

A marble is selected at random from bag ***X*** and placed in bag ***Y***
A second marble is selected at random from bag ***X*** and placed in bag ***Y***
A third marble is then selected, this time from bag ***Y***

(a) Use this information to complete the tree diagram shown on page 7

(3)

(b) Find the probability that the 2 marbles selected from bag ***X*** are of different colours.

(2)

(c) Find the probability that all 3 marbles selected are the same colour.

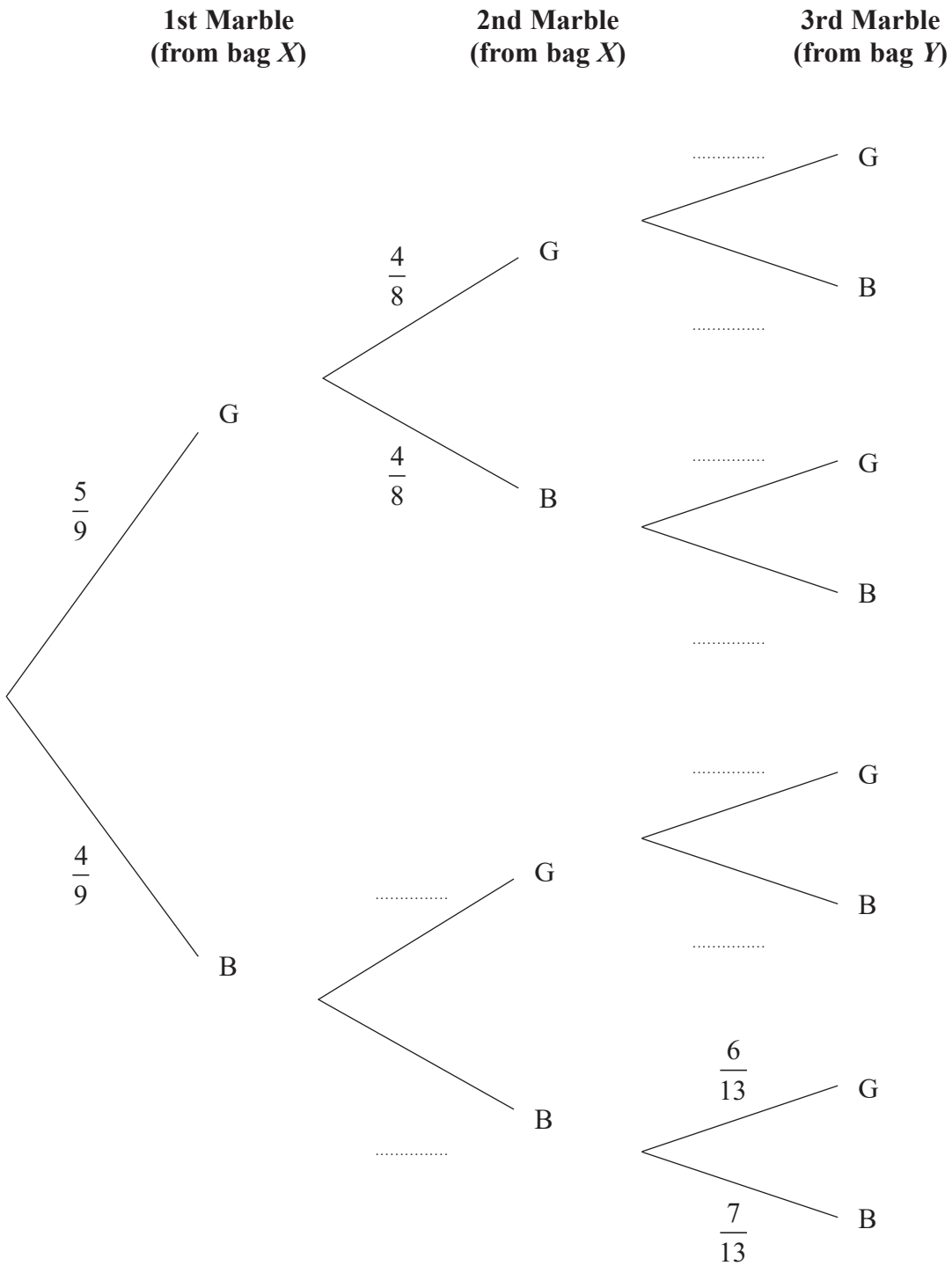
(2)

Given that all three marbles selected are the same colour,

(d) find the probability that they are all green.

(3)

Question 2 continued



Question 2 continued

Question 2 continued

(Total for Question 2 is 10 marks)

3. The probability distribution of the discrete random variable X is given by

x	2	3	4
$P(X = x)$	a	0.4	$0.6 - a$

where a is a constant.

(a) Find, in terms of a, $E(X)$

(2)

(b) Find the range of the possible values of $E(X)$

(3)

Given that $Var(X) = 0.56$

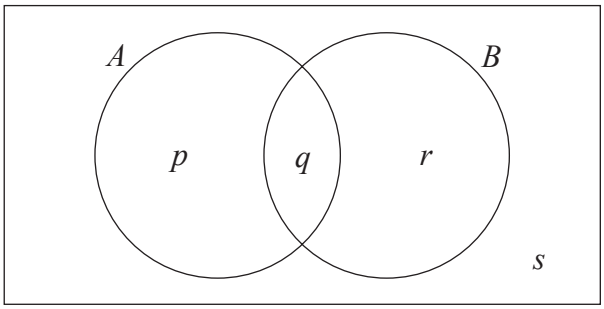
(c) find the possible values of a

(6)

Question 3 continued

(Total for Question 3 is 11 marks)

4. (i) In the Venn diagram below, A and B represent events and p, q, r and s are probabilities.



$$P(A) = \frac{7}{25} \qquad P(B) = \frac{1}{5} \qquad P\left[\left(A \cap B'\right) \cup \left(A' \cap B\right)\right] = \frac{8}{25}$$

(a) Use algebra to show that $2p + 2q + 2r = \frac{4}{5}$

(4)

(b) Find the value of p, the value of q, the value of r and the value of s

(5)

(ii) Two events, C and D, are such that

$$P(C) = \frac{x}{x+5} \qquad P(D) = \frac{5}{x}$$

where x is a positive constant.

By considering $P(C) + P(D)$ show that C and D **cannot** be mutually exclusive.

(4)

Question 4 continued

Question 4 continued

Question 4 continued

(Total for Question 4 is 13 marks)

5. The lengths, L mm, of housefly wings are normally distributed with $L \sim N(4.5, 0.4^2)$

(a) Find the probability that a randomly selected housefly has a wing length of less than 3.86 mm.

(3)

(b) Find

(i) the upper quartile (Q_3) of L

(ii) the lower quartile (Q_1) of L

(4)

A value that is greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$ or smaller than $Q_1 - 1.5 \times (Q_3 - Q_1)$ is defined as an outlier.

(c) Find these two outlier limits.

(3)

A housefly is selected at random.

(d) Using standardisation, show that the probability that this housefly is **not** an outlier is 0.993 to 3 decimal places.

(3)

Given that this housefly is **not** an outlier,

(e) showing your working, find the probability that the wing length of this housefly is greater than 5 mm.

(4)

Question 5 continued

Question 5 continued

Question 5 continued

(Total for Question 5 is 17 marks)

6. A research student is investigating the maximum weight, y grams, of sugar that will dissolve in 100 grams of water at various temperatures, x °C, where $10 \leqslant x \leqslant 80$

The research student calculated the regression line of y on x and found it to be

$$y = 151.2 + 2.72x$$

(a) Give an interpretation of the gradient of the regression line.

(1)

(b) Use the regression line to estimate the maximum weight of sugar that will dissolve in 100 grams of water when the temperature is 90 °C.

(2)

(c) Comment on the reliability of your estimate, giving a reason for your answer.

(2)

Using the regression line of y on x and the following summary statistics

$$\sum y = 3119 \qquad \sum y^2 = 851\,093 \qquad \sum x^2 = 24\,500 \qquad n = 12$$

(d) show that the product moment correlation coefficient for these data is 0.988 to 3 decimal places.

(7)

The research student's supervisor plotted the original data on a scatter diagram, shown on page 23

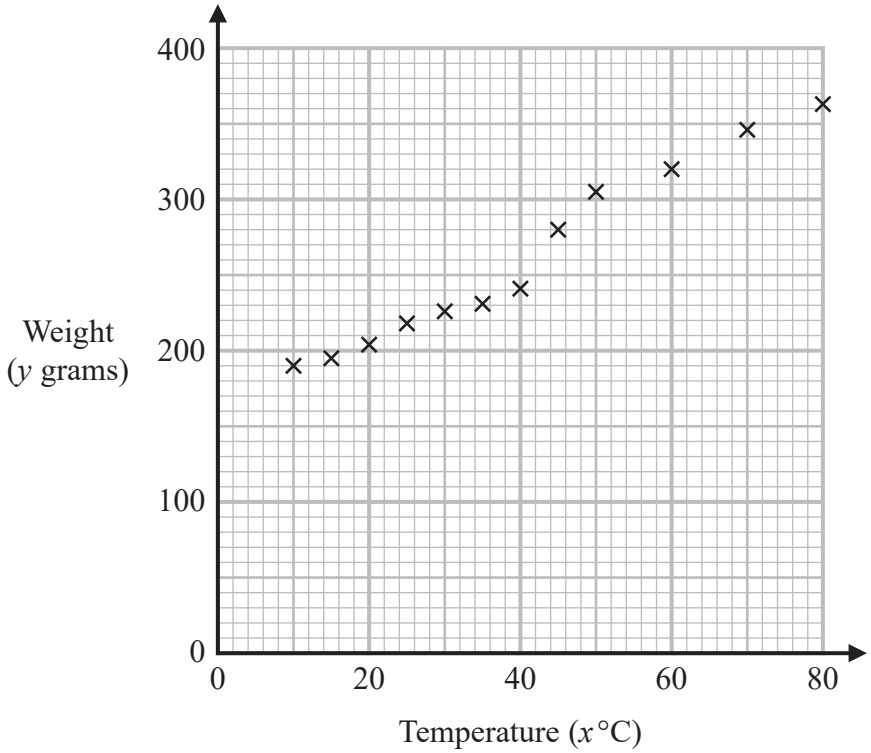
With reference to both the scatter diagram and the correlation coefficient,

(e) discuss the suitability of a linear regression model to describe the relationship between x and y.

(2)

Question 6 continued

Question 6 continued


400
300
200
100
0
Weight
(y grams)
0
20
40
60
80
Temperature (x °C)

Question 6 continued

(Total for Question 6 is 14 marks)

TOTAL FOR PAPER IS 75 MARKS

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

Thursday 11 May 2023

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 your 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

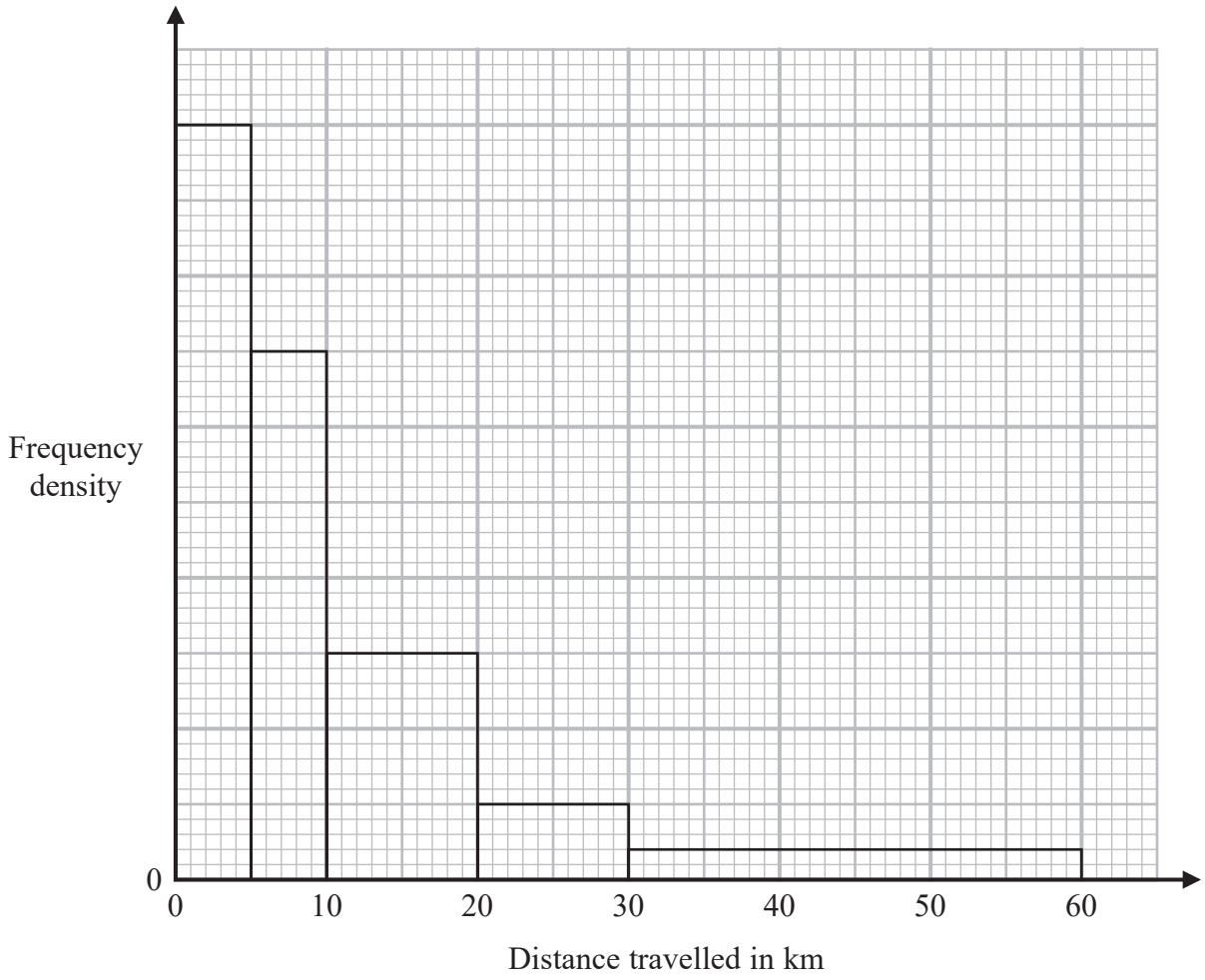
- A booklet 'Mathematical Formulae and Statistical Tables' is provided.
- There are 7 questions in this question paper. The total mark for this paper is 75.
- The marks for **each** question are shown in brackets
  – *use this as a guide as to how much time to spend on each question.*

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 ▶

1. The histogram shows the distances, in km, that 274 people travel to work.



Given that 60 of these people travel between 10 km and 20 km to work, estimate

(a) the number of people who travel between 22 km and 45 km to work,

(3)

(b) the median distance travelled to work by these 274 people,

(2)

(c) the mean distance travelled to work by these 274 people.

(3)

Question 1 continued

Question 1 continued

Question 1 continued

(Total for Question 1 is 8 marks)

2. Two students, Olive and Shan, collect data on the weight, w grams, and the tail length, t cm, of 15 mice.

Olive summarised the data as follows

$S_{tt} = 5.3173$ $\quad \sum w^2 = 6089.12$ $\quad \sum tw = 2304.53$ $\quad \sum w = 297.8$ $\quad \sum t = 114.8$

(a) Calculate the value of S_{tw} and the value of S_{ww}

(3)

(b) Calculate the value of the product moment correlation coefficient between w and t

(2)

(c) Show that the equation of the regression line of w on t can be written as

$$w = -16.7 + 4.77t$$

(3)

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

(1)

(e) Explain why it would not be appropriate to use the regression line in part (c) to estimate the weight of a mouse with a tail length of 2 cm.

(2)

Shan decided to code the data using $x = t - 6$ and $y = \frac{w}{2} - 5$

(f) Write down the value of the product moment correlation coefficient between x and y

(1)

(g) Write down an equation of the regression line of y on x
You do not need to simplify your equation.

(1)

Question 2 continued

Question 2 continued

Question 2 continued

(Total for Question 2 is 13 marks)

3. Jim records the length, l mm, of 81 salmon. The data are coded using $x = l - 600$ and the following summary statistics are obtained.

$$n = 81 \qquad \sum x = 3711 \qquad \sum x^2 = 475\,181$$

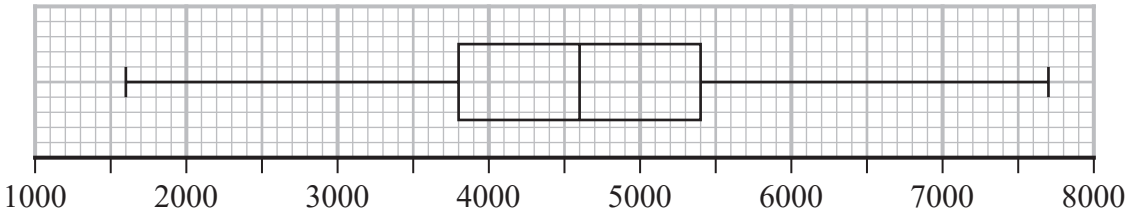
(a) Find the mean length of these salmon.

(3)

(b) Find the variance of the lengths of these salmon.

(2)

The weight, w grams, of each of the 81 salmon is recorded to the nearest gram.
The recorded results for the 81 salmon are summarised in the box plot below.



Weight of salmon (w grams)

(c) Find the maximum number of salmon that have weights in the interval

$$4600 < w \leqslant 7700$$

(1)

Raj says that the box plot is incorrect as Jim has not included outliers.

For these data an outlier is defined as a value that is more than

$$1.5 \times \text{IQR above the upper quartile} \quad \text{or} \quad 1.5 \times \text{IQR below the lower quartile}$$

(d) Show that there are no outliers.

(3)

Question 3 continued

Question 3 continued

Question 3 continued

(Total for Question 3 is 9 marks)

4. A bag contains a large number of coloured counters. Each counter is labelled A, B or C

30% of the counters are labelled A
45% of the counters are labelled B
The rest of the counters are labelled C

It is known that

2% of the counters labelled A are red
4% of the counters labelled B are red
6% of the counters labelled C are red

One counter is selected at random from the bag.

(a) Complete the tree diagram on the opposite page to illustrate this information.

(2)

(b) Calculate the probability that the counter is labelled A and is not red.

(2)

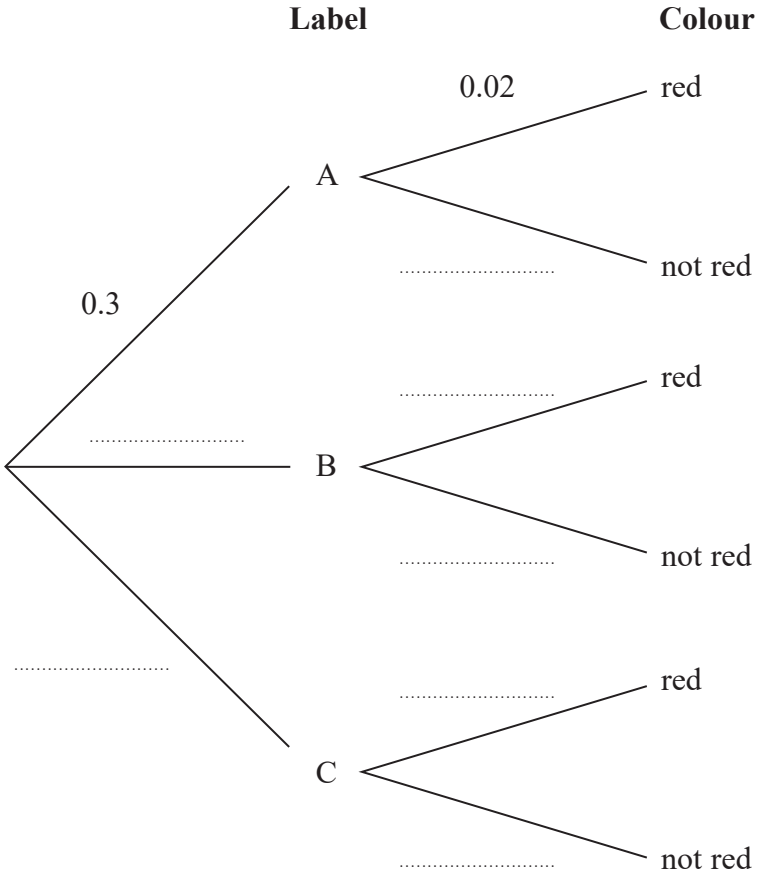
(c) Calculate the probability that the counter is red.

(2)

(d) Given that the counter is red, find the probability that it is labelled C

(3)

Question 4 continued



(Total for Question 4 is 9 marks)

5. A discrete random variable Y has probability function

$$\mathrm{P}(Y = y) = \begin{cases} k(3 - y) & y = 1, 2 \\ k(y^2 - 8) & y = 3, 4, 5 \\ k & y = 6 \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant.

(a) Show that $k = \frac{1}{30}$

(2)

Find the exact value of

(b) $\mathrm{P}(1 < Y \leqslant 4)$

(2)

(c) $\mathrm{E}(Y)$

(2)

The random variable $X = 15 - 2Y$

(d) Calculate $\mathrm{P}(Y \geqslant X)$

(3)

(e) Calculate $\mathrm{Var}(X)$

(4)

Question 5 continued

Question 5 continued

Question 5 continued

(Total for Question 5 is 13 marks)

6. Three events A, B and C are such that

$$\mathrm{P}(A) = 0.1 \qquad \mathrm{P}(B \mid A) = 0.3 \qquad \mathrm{P}(A \cup B) = 0.25 \qquad \mathrm{P}(C) = 0.5$$

Given that A and C are mutually exclusive

(a) find $\mathrm{P}(A \cup C)$

(1)

(b) Show that $\mathrm{P}(B) = 0.18$

(3)

Given also that B and C are independent,

(c) draw a Venn diagram to represent the events A, B and C and the probabilities associated with each region.

(5)

Question 6 continued

Question 6 continued

Question 6 continued

(Total for Question 6 is 9 marks)

7. A machine squeezes apples to extract their juice. The volume of juice, J ml, extracted from 1 kg of apples is modelled by a normal distribution with mean μ and standard deviation σ

Given that $\mu = 500$ and $\sigma = 25$ use standardisation to

(a) (i) show that $P(J > 510) = 0.3446$

(2)

(ii) calculate the value of d such that $P(J > d) = 0.9192$

(3)

Zen randomly selects 5 bags each containing 1 kg of apples and records the volume of juice extracted from each bag of apples.

(b) Calculate the probability that each of the 5 bags of apples produce less than 510 ml of juice.

(2)

Following adjustments to the machine, the volume of juice, R ml, extracted from 1 kg of apples is such that $\mu = 520$ and $\sigma = k$

Given that $P(R < r) = 0.15$ and $P(R > 3r - 800) = 0.005$

(c) find the value of r and the value of k

(7)

Question 7 continued

Question 7 continued

Question 7 continued

Question 7 continued

(Total for Question 7 is 14 marks)

TOTAL FOR PAPER IS 75 MARKS

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

Thursday 12 October 2023

Afternoon (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 your 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 ▶

1. Sally plays a game in which she can either win or lose.

A turn consists of up to 3 games. On each turn Sally plays the game up to 3 times. If she wins the first 2 games or loses the first 2 games, then she will not play the 3rd game.

- The probability that Sally wins the first game in a turn is 0.7
- If Sally wins a game the probability that she wins the next game is 0.6
- If Sally loses a game the probability that she wins the next game is 0.2

(a) Use this information to complete the tree diagram on page 3

(3)

(b) Find the probability that Sally wins the first 2 games in a turn.

(2)

(c) Find the probability that Sally wins exactly 2 games in a turn.

(2)

Given that Sally wins 2 games in a turn,

(d) find the probability that she won the first 2 games.

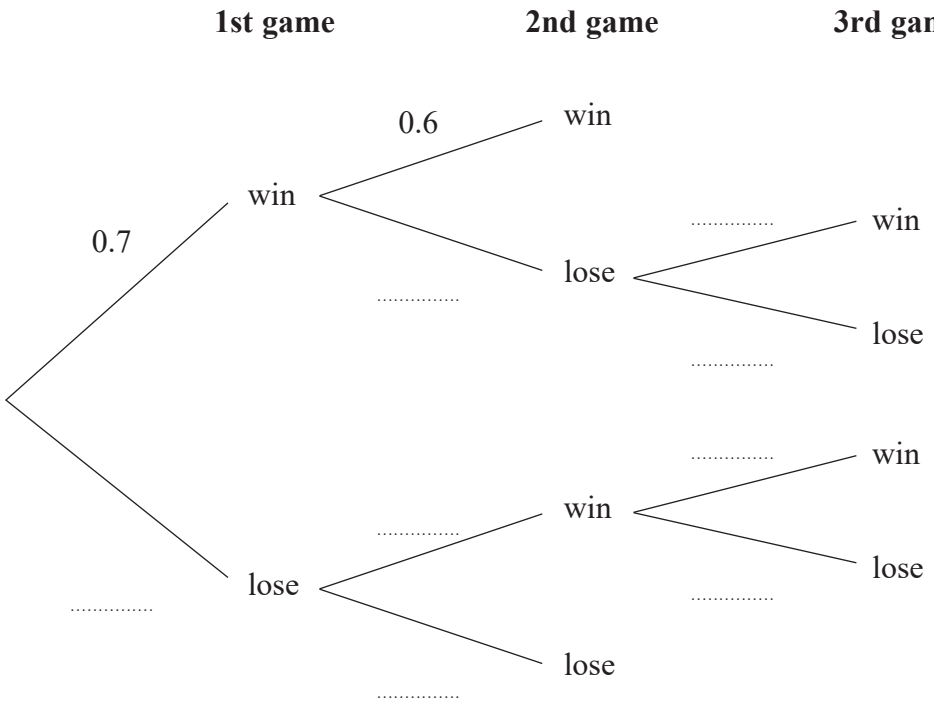
(2)

Given that Sally won the first game in a turn,

(e) find the probability that she won 2 games.

(2)

Question 1 continued



Question 1 continued

Question 1 continued

(Total for Question 1 is 11 marks)

2. The weights, to the nearest kilogram, of a sample of 33 red kangaroos taken in December are summarised in the stem and leaf diagram below.

	Weight (kg)	Totals
1	6	(1)
2	3 6	(2)
3	2 4 6	(3)
4	2 5 5 6 6 7 8	(7)
5	3 4 7 7 7 8 9 9	(8)
6	0 2 2 3 3 7 8	(7)
7	2 8	(2)
8	2 6	(2)
9	4	(1)

Key: 3 | 2 represents 32 kg

(a) Find

(i) the value of the median

(ii) the value of Q_1 and the value of Q_3

for the weights of these red kangaroos.

(3)

For these data an outlier is defined as a value that is

greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$

or smaller than $Q_1 - 1.5 \times (Q_3 - Q_1)$

(b) Show that there are 2 outliers for these data.

(3)

Figure 1 on page 7 shows a box plot for the weights of the same 33 red kangaroos taken in February, earlier in the year.

(c) In the space on Figure 1, draw a box plot to represent the weights of these red kangaroos in December.

(4)

(d) Compare the distribution of the weights of red kangaroos taken in February with the distribution of the weights of red kangaroos taken in December of the same year. You should interpret your comparisons in the context of the question.

(3)

Question 2 continued

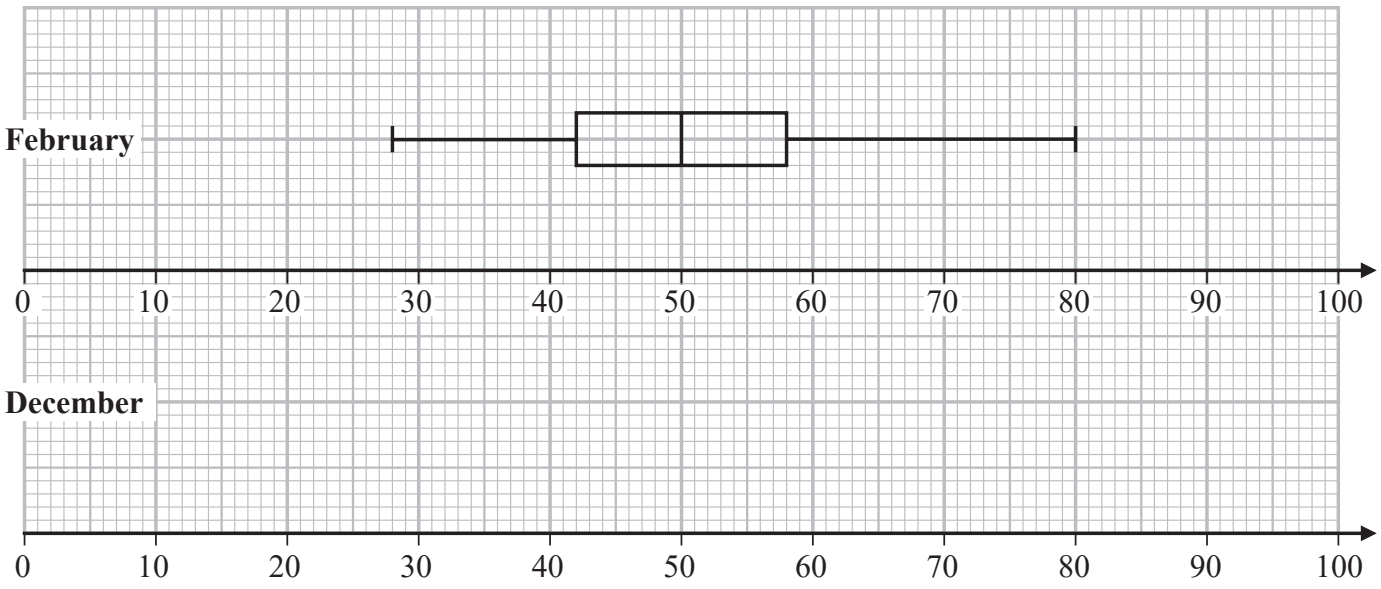


Figure 1

Turn over for a spare grid if you need to redraw your box plot.

Question 2 continued

Question 2 continued

Only use this grid if you need to redraw your box plot.

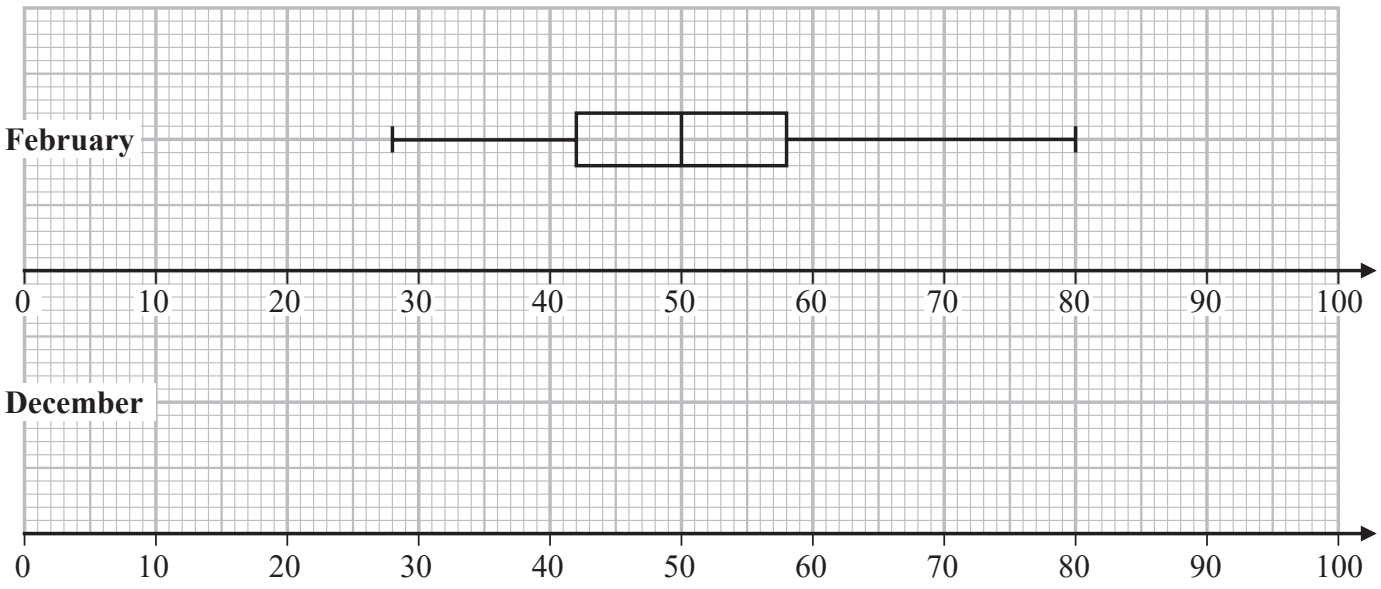


Figure 1

(Total for Question 2 is 13 marks)

3. (i) Bob shops at a market each week. The event that

Bob buys carrots is denoted by C

Bob buys onions is denoted by O

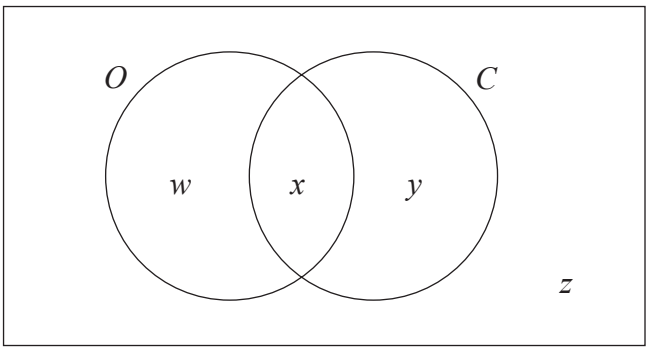
At each visit, Bob may buy neither, or one, or both of these items. The probability that

Bob buys carrots is 0.65

Bob does **not** buy onions is 0.3

Bob buys onions but not carrots is 0.15

The Venn diagram below represents the events C and O



where w, x, y and z are probabilities.

(a) Find the value of w, the value of x, the value of y and the value of z

(4)

For one visit to the market,

(b) find the probability that Bob buys either carrots or onions but not both.

(1)

(c) Show that the events C and O are **not** independent.

(2)

(ii) F, G and H are 3 events. F and H are mutually exclusive. F and G are independent.

Given that

$$\mathrm{P}(F) = \frac{2}{7} \qquad \mathrm{P}(H) = \frac{1}{4} \qquad \mathrm{P}(F \cup G) = \frac{5}{8}$$

(a) find $\mathrm{P}(F \cup H)$

(1)

(b) find $\mathrm{P}(G)$

(3)

(c) find $\mathrm{P}(F \cap G)$

(1)

Question 3 continued

Question 3 continued

Question 3 continued

(Total for Question 3 is 12 marks)

4. The discrete random variable X has the following probability distribution.

x	1	2	3	4
$P(X = x)$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{2}{5}$

(a) Show that $E\left(\frac{1}{X}\right) = \frac{2}{5}$

(1)

(b) Find $\text{Var}\left(\frac{1}{X}\right)$

(3)

The random variable $Y = \frac{30}{X}$

(c) Find

(i) $E(Y)$

(ii) $\text{Var}(Y)$

(3)

(d) Find $P(X < 3 \mid Y < 20)$

(5)

Question 4 continued

Question 4 continued

Question 4 continued

(Total for Question 4 is 12 marks)

5. The weights, X grams, of a particular variety of fruit are normally distributed with

$$X \sim \mathrm{N}(210, 25^2)$$

A fruit of this variety is selected at random.

(a) Show that the probability that the weight of this fruit is less than 240 grams is 0.8849

(2)

(b) Find the probability that the weight of this fruit is between 190 grams and 240 grams.

(2)

(c) Find the value of k such that $\mathrm{P}(210 - k < X < 210 + k) = 0.95$

(3)

A wholesaler buys large numbers of this variety of fruit and classifies the lightest 15% as small.

(d) Find the maximum weight of a fruit that is classified as small.
You must show your working clearly.

(3)

The weights, Y grams, of a second variety of this fruit are normally distributed with

$$Y \sim \mathrm{N}(\mu, \sigma^2)$$

Given that 5% of these fruit weigh less than 152 grams and 40% weigh more than 180 grams,

(e) calculate the mean and standard deviation of the weights of this variety of fruit.

(5)

Question 5 continued

Question 5 continued

Question 5 continued

(Total for Question 5 is 15 marks)

6. The variables x and y have the following regression equations based on the same 12 observations.

	Regression equation
y on x	$y = 1.4x + 1.5$
x on y	$x = 1.2 + 0.2y$

(a) (i) Find the point of intersection of these lines.

(ii) Hence show that $\sum x = 25$

(4)

Given that

$$\sum xy = \frac{6961}{60}$$

(b) Find S_{xy}

(4)

(c) Find the product moment correlation coefficient between x and y

(4)

Question 6 continued

Question 6 continued

(Total for Question 6 is 12 marks)

TOTAL FOR PAPER IS 75 MARKS

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

Wednesday 10 January 2024

Afternoon (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.

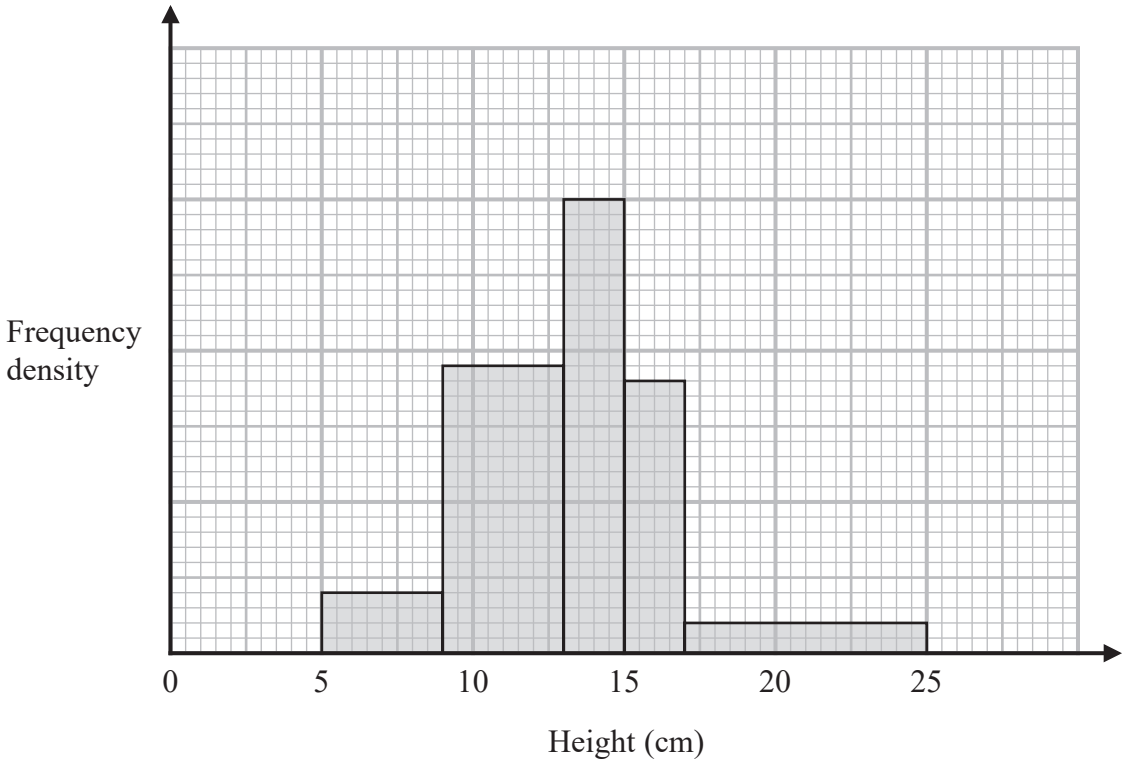
Information

- A booklet 'Mathematical Formulae and Statistical Tables' is provided.
- There are 8 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.

1. The histogram below shows the distribution of the heights, to the nearest cm, of 408 plants.



(a) Use the histogram to complete the following table.

Height (h cm)	$5 \leqslant h < 9$	$9 \leqslant h < 13$	$13 \leqslant h < 15$	$15 \leqslant h < 17$	$17 \leqslant h < 25$
Frequency	32	152	120		

(2)

(b) Use interpolation to estimate the median.

(2)

The mean height of these plants is 13.2 cm correct to one decimal place.

(c) Describe the skew of these data. Give a reason for your answer.

(1)

Two of these plants are chosen at random.

(d) Estimate the probability that both of their heights are between 8 cm and 14 cm

(3)

Question 1 continued

Question 1 continued

Question 1 continued

(Total for Question 1 is 8 marks)

2. The average minimum monthly temperature, x degrees Fahrenheit (°F), and the average maximum monthly temperature, y degrees Fahrenheit (°F), in Kolkata were recorded for 12 months.

Some of the summary statistics are given below.

$$\sum x = 862 \quad \sum x^2 = 62\,802 \quad S_{yy} = 413.67 \quad S_{xy} = 512.67 \quad n = 12$$

(a) (i) Calculate the mean of the 12 values of the average **minimum** monthly temperature.

(ii) Show that the standard deviation of the 12 values of the average **minimum** monthly temperature is 8.57°F to 3 significant figures.

(3)

(b) Calculate the product moment correlation coefficient between x and y

(3)

For comparative purposes with a UK city, it was necessary to convert the temperatures from degrees Fahrenheit (°F) to degrees Celsius (°C).

The formula used was

$$c = \frac{5}{9}(f - 32)$$

where f is the temperature in °F and c is the temperature in °C

(c) Use this formula and the values from part (a) to calculate, in °C, the mean and the standard deviation of the 12 values of the average **minimum** monthly temperature in Kolkata.
Give your answers to 3 significant figures.

(4)

Given that

- u is the equivalent temperature in °C of x
- v is the equivalent temperature in °C of y

(d) state, giving a reason, the product moment correlation coefficient between u and v

(2)

Question 2 continued

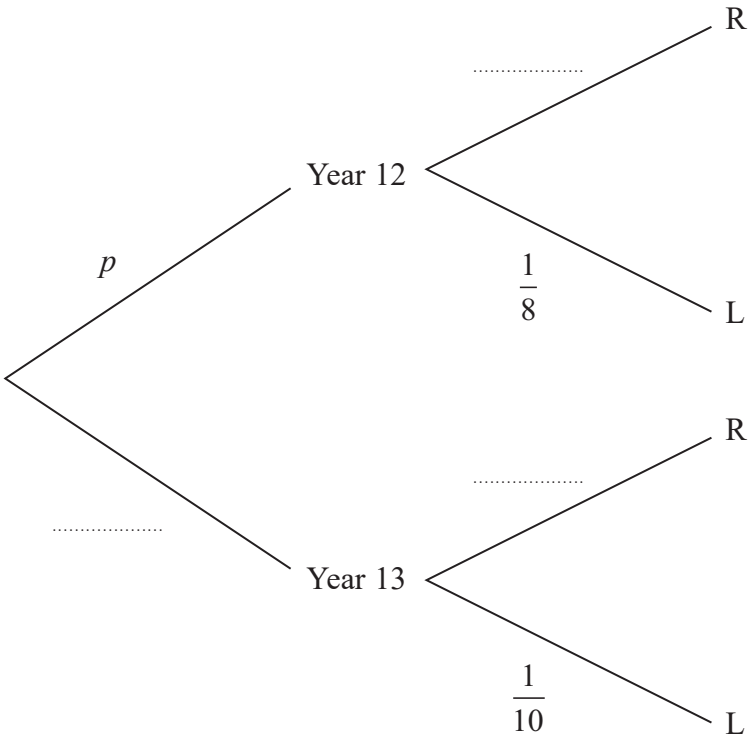
Question 2 continued

Question 2 continued

(Total for Question 2 is 12 marks)

3. In a sixth form college each student in Year 12 and Year 13 is either left-handed (L) or right-handed (R).

The partially completed tree diagram, where p is a probability, gives information about these students.



(a) Complete the tree diagram, in terms of p where necessary.

(1)

The probability that a student is left-handed is 0.11

(b) Find the value of p

(3)

(c) Find the probability that a student selected at random is in Year 12 and left-handed.

(2)

Given that a student is right-handed,

(d) find the probability that the student is in Year 12

(2)

Question 3 continued

(Total for Question 3 is 8 marks)

4. A French test and a Spanish test were sat by 11 students.

The table below shows their marks.

Student	A	B	C	D	E	F	G	H	I	J	K
French mark (f)	24	30	32	32	36	36	40	44	50	60	68
Spanish mark (s)	16	90	24	28	32	36	38	44	48	48	68

Greg says that if these points were plotted on a scatter diagram, then the point (30, 90) would be an outlier because 90 is an outlier for the Spanish marks.

An outlier is defined as a value that is

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

(a) Show that 90 is an outlier for the Spanish marks.

(3)

Ignoring the point (30, 90), Greg calculated the following summary statistics.

$$\sum f = 422 \quad \sum s = 382 \quad S_{ff} = 1667.6 \quad S_{fs} = 1735.6$$

(b) Use these summary statistics to show that the equation of the least squares regression line of s on f for the remaining 10 students is

$$s = -5.72 + 1.04f$$

where the values of the intercept and gradient are given to 3 significant figures. You must show your working.

(3)

(c) Give an interpretation of the gradient of the regression line.

(1)

Two further students sat the French test but missed the Spanish test.

(d) Using the equation given in part (b), estimate

(i) a Spanish mark for the student who scored 55 marks in their French test,

(ii) a Spanish mark for the student who scored 18 marks in their French test.

(3)

(e) State, giving a reason, which of the two estimates found in part (d) would be the more reliable estimate.

(2)

Question 4 continued

Question 4 continued

Question 4 continued

(Total for Question 4 is 12 marks)

5. The distance an athlete can throw a discus is normally distributed with mean 40 m and standard deviation 4 m

(a) Using standardisation, show that the probability that this athlete throws the discus less than 38.8 m is 0.3821

(2)

This athlete enters a discus competition.
To qualify for the final, they have 3 attempts to throw the discus a distance of more than 38.8 m
Once they qualify, they do not use any of their remaining attempts.

Given that they qualified for the final and that throws are independent,

(b) find the probability that this athlete qualified for the final on their second throw with a distance of more than 44 m

(5)

Question 5 continued

(Total for Question 5 is 7 marks)

6. The events A and B satisfy

$$P(A) = x \qquad P(B) = y \qquad P(A \cup B) = 0.65 \qquad P(B \mid A) = 0.3$$

(a) Show that

$$14x + 20y = 13$$

(3)

The events B and C are mutually exclusive such that

$$P(B \cup C) = 0.85 \qquad P(C) = \frac{1}{2}x + y$$

(b) (i) Find a second equation in x and y

(ii) Hence find the value of x and the value of y

(4)

(c) Determine whether or not A and B are statistically independent.
You must show your working clearly.

(2)

Question 6 continued

Question 6 continued

Question 6 continued

(Total for Question 6 is 9 marks)

7. The cumulative distribution of a discrete random variable X is given by

x	1	2	3	4
$\mathrm{F}(x)$	$\frac{1}{13}$	$\frac{2k-1}{26}$	$\frac{3(k+1)}{26}$	$\frac{k+4}{8}$

where k is a positive constant.

(a) Show that $k = 4$

(1)

(b) Find the probability distribution of the discrete random variable X

(3)

(c) Using your answer to part (b), write down the mode of X

(1)

(d) Calculate $\mathrm{Var}(13X - 6)$

(5)

Question 7 continued

Question 7 continued

Question 7 continued

(Total for Question 7 is 10 marks)

8. The random variable X is normally distributed with mean μ and variance 36

Given that

$$P(\mu - 2k < X < \mu + 2k) = 0.6$$

(a) find the value of k

(4)

The random variable Y is normally distributed with mean μ and standard deviation σ

Given that

$$2\mu = 3\sigma^2 \quad \text{and} \quad P\left(Y > \frac{3}{2}\mu\right) = 0.0668$$

(b) find the value of μ and the value of σ

(5)

Question 8 continued

Question 8 continued

(Total for Question 8 is 9 marks)

TOTAL FOR PAPER IS 75 MARKS

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 ▶

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)

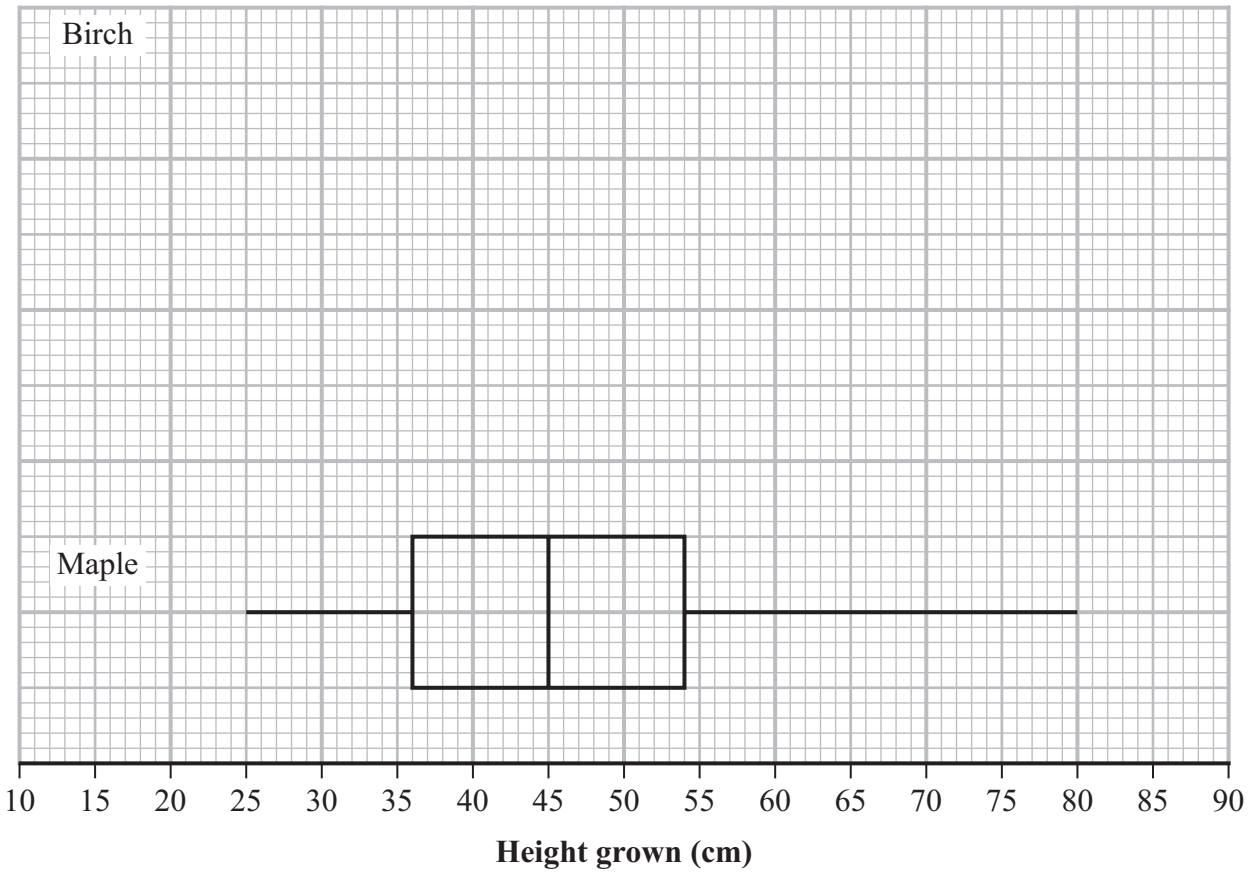
The researcher realises he has missed out 4 pieces of data for the **Maple** trees. The heights each seedling grows in one year, to the nearest cm, in ascending order, for these 4 Maple trees are 27 cm, a cm, 48 cm, $2a$ cm.

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

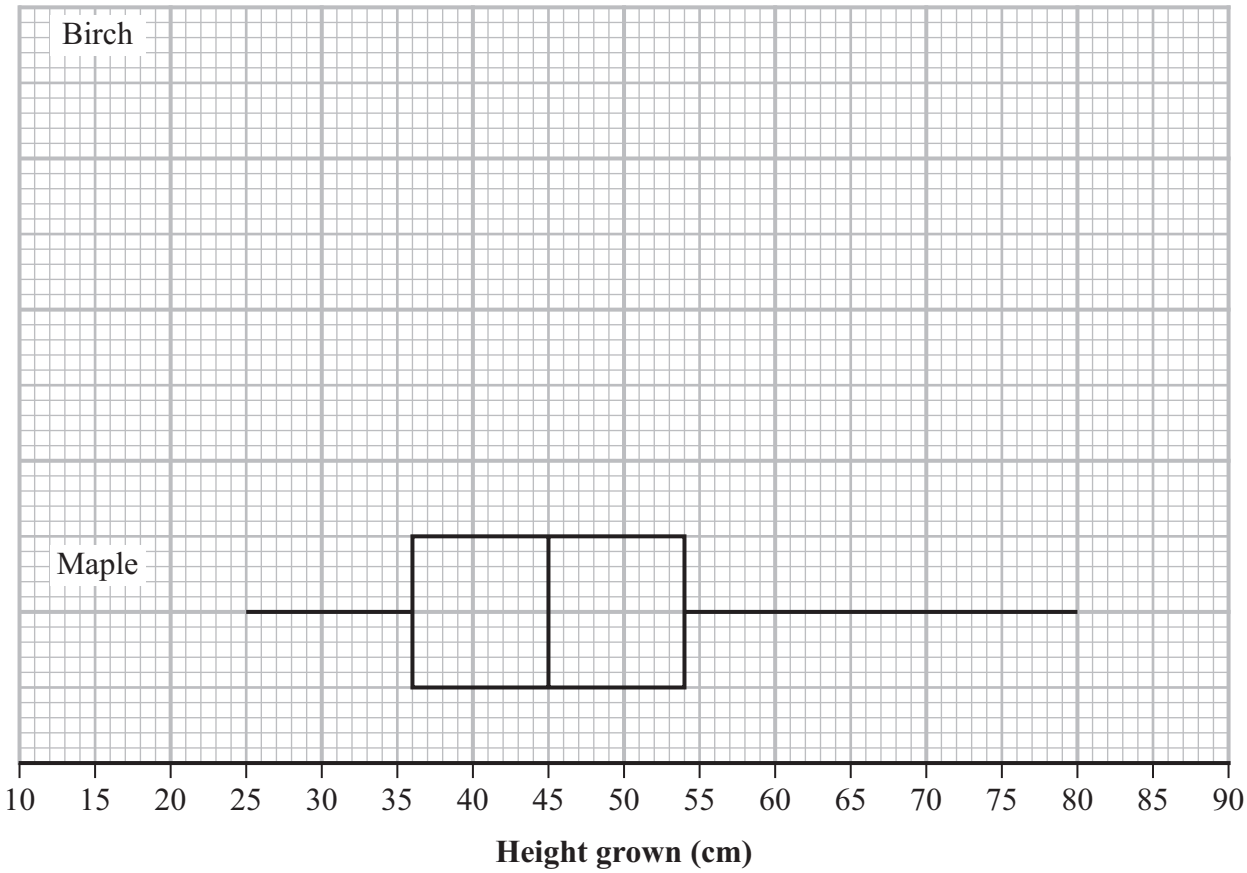


Turn over for spare grid if you need to redraw your answer for part (d)

Question 1 continued

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 $\text{Var}(X) = 4.14$

(3)

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

Given that

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

(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

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)

Question 3 continued

Question 3 continued

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

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

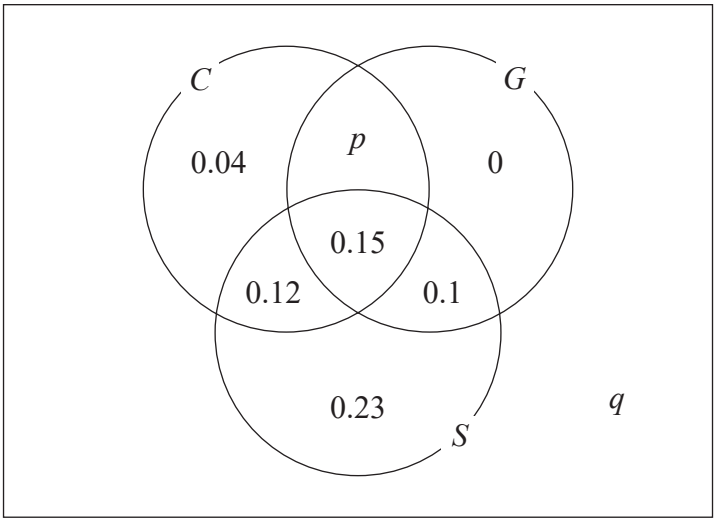
Question 5 continued

(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)

Question 6 continued

Question 6 continued

(Total for Question 6 is 13 marks)

TOTAL FOR PAPER IS 75 MARKS

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel International Advanced Level

Thursday 17 October 2024

Afternoon (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 8 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 ▶

1. The back-to-back stem and leaf diagram on page 3 shows information about the running times of 31 Action films and 31 Comedy films.
The running times are given to the nearest minute.

(a) Write down the modal running time for these Action films.

(1)

Some of the quartiles for these two distributions are shown in the table below.

	Action films	**Comedy films**
Lower quartile	121	a
Median	b	117
Upper quartile	138	c

(b) Find the value of a, the value of b and the value of c

(3)

(c) For these Action films find, to one decimal place,

(i) the mean running time,

(ii) the standard deviation of the running times.

(You may use $\sum x = 4016$ and $\sum x^2 = 525056$ where x is the running time, in minutes, of an Action film.)

(3)

One measure of skewness is found using

$$\frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

(d) Evaluate this measure and describe the skewness for the running times of these Action films.

(2)

(e) Comment on one difference between the distribution of the running times of these Action films and the distribution of the running times of these Comedy films.
State the values of any statistics you have used to support your comment.

(1)

Question 1 continued

Totals	Action films		Comedy films	Totals
(1)	0	9	2 2 3 5 8	(5)
(0)		10	3 5 6 6 8 9	(6)
(5)	9 8 6 4 2	11	0 2 4 6 7 9 9 9	(8)
(10)	9 9 8 7 6 5 4 3 1 0	12	1 2 4 6 6 7 7 7 7 8 9	(11)
(8)	8 7 7 7 5 4 2 1	13	1	(1)
(7)	7 7 6 6 4 3 1	14		(0)

Key: 0 | 9 | 2 means 90 minutes for an Action film and 92 minutes for a Comedy film

Question 1 continued

Question 1 continued

(Total for Question 1 is 10 marks)

2. A biologist records the length, y cm, and the weight, w kg, of 50 rabbits.
The following summary statistics are calculated from these data.

$$\sum y = 2015 \quad \sum y^2 = 81938.5 \quad \sum w = 125 \quad S_{ww} = 72.25 \quad S_{yw} = 219.55$$

(a) (i) Show that $S_{yy} = 734$

(ii) Calculate the product moment correlation coefficient for these data.
Give your answer to 3 decimal places.

(3)

(b) Interpret your value of the product moment correlation coefficient.

(1)

The biologist believes that a linear regression model may be appropriate to describe these data.

(c) State, with a reason, whether or not your value of the product moment correlation coefficient is consistent with the biologist's belief.

(1)

(d) Find the equation of the regression line of w on y, giving your answer in the form $w = a + by$

(4)

Jeff has a pet rabbit of length 45 cm.

(e) Use your regression equation to estimate the weight of Jeff's rabbit.

(2)

Question 2 continued

Question 2 continued

Question 2 continued

(Total for Question 2 is 11 marks)

3. A group of 200 adults were asked whether they read cooking magazines, travel magazines or sport magazines.
Their replies showed that

- 29 read only cooking magazines
- 33 read only travel magazines
- 42 read only sport magazines
- 17 read cooking magazines and sport magazines but **not** travel magazines
- 11 read travel magazines and sport magazines but **not** cooking magazines
- 22 read cooking magazines and travel magazines but **not** sport magazines
- 32 **do not** read cooking magazines, travel magazines or sport magazines

(a) Using this information, complete the Venn diagram on page 11

(3)

One of these adults was chosen at random.

(b) Find the probability that this adult,

(i) reads cooking magazines and travel magazines and sport magazines,

(1)

(ii) does **not** read cooking magazines.

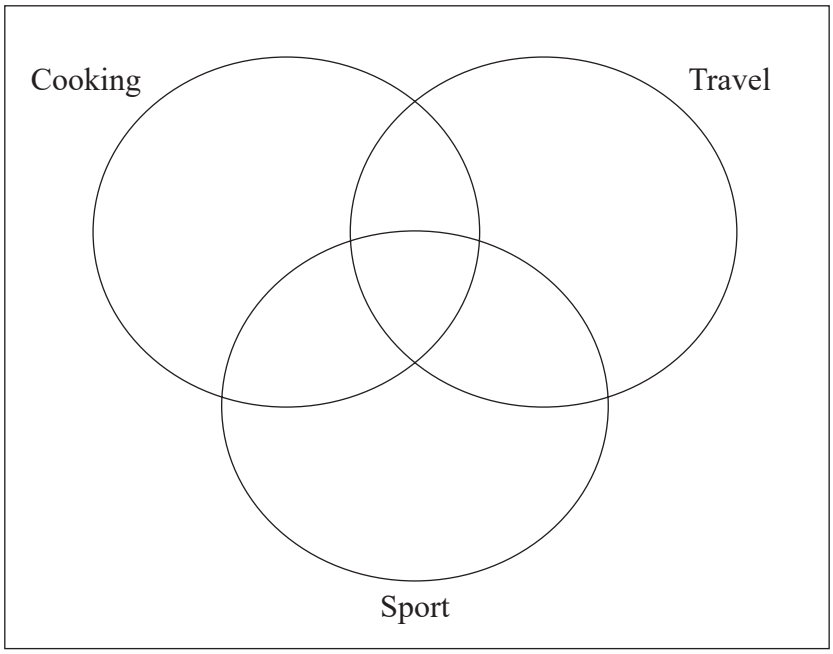
(2)

Given that this adult reads travel magazines,

(c) find the probability that this adult also reads sport magazines.

(2)

Question 3 continued



Question 3 continued

Question 3 continued

(Total for Question 3 is 8 marks)

4. **In this question you must show all stages of your working.**
Solutions relying entirely on calculator technology are not acceptable.

The distances, m miles, a motorbike travels on a full tank of petrol can be modelled by a normal distribution with mean 170 miles and standard deviation 16 miles.

(a) Find the probability that, on a randomly selected journey, the motorbike could travel at least 190 miles on a full tank of petrol.

(2)

The probability that, on a randomly selected journey, the motorbike could travel at least d miles on a full tank of petrol is 0.9

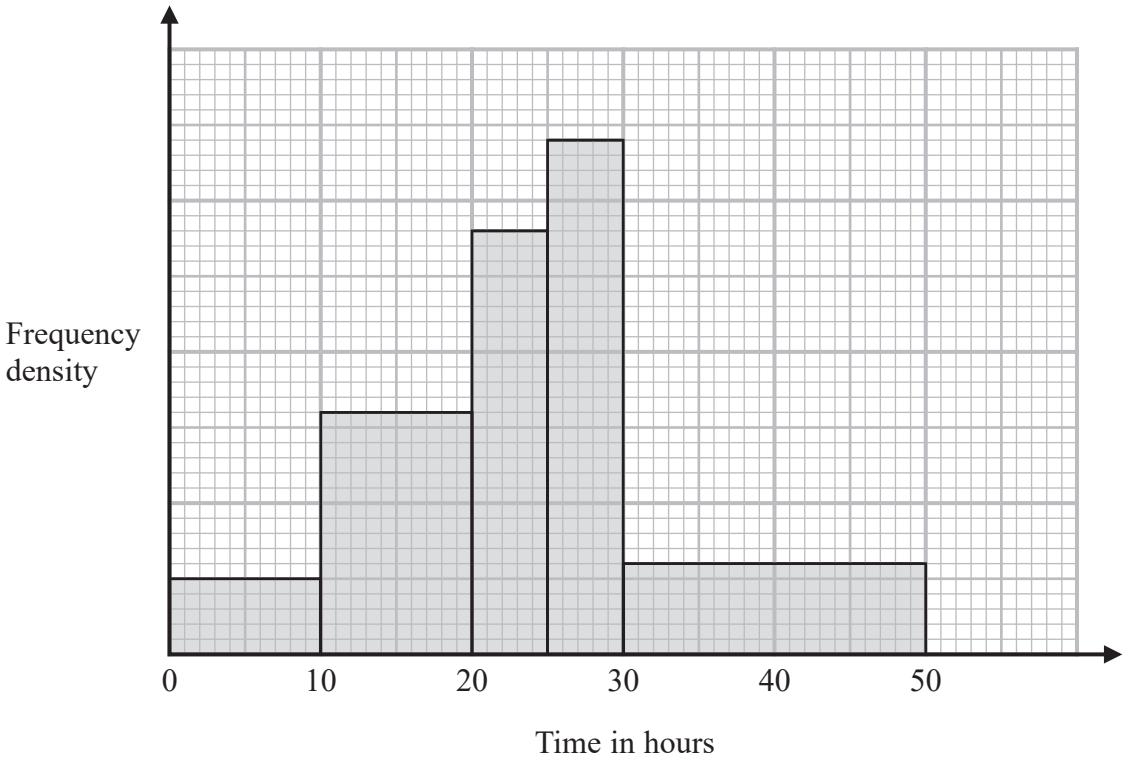
(b) Find the value of d

(3)

Question 4 continued

(Total for Question 4 is 5 marks)

5.



The histogram shows the number of hours worked in a given week by a group of 64 freelance photographers.

(a) Give a reason to justify the use of a histogram to represent these data.

(1)

Given that 16 of these freelance photographers spent between 10 and 20 hours working in this week,

(b) estimate the number that spent between 12 and 24 hours working in this week.

(3)

(c) Find an estimate for the median time spent working in this week by these 64 freelance photographers.

(2)

Charlie decides to model these data using a normal distribution.
Charlie calculates an estimate of the mean to be 23.9 hours to one decimal place.

(d) Comment on Charlie's decision to use a normal distribution.
Give a justification for your answer.

(2)

Question 5 continued

Question 5 continued

Question 5 continued

(Total for Question 5 is 8 marks)

6. A biased die with six faces is rolled. The discrete random variable X represents the score which is uppermost. The **cumulative** distribution function of X is shown in the table below.

x	1	2	3	4	5	6
F(x)	0.1	0.2	$3k$	$5k$	$7k$	$10k$

(a) Find the value of the constant k

(1)

(b) Find the probability distribution of X

(3)

A biased die with eight faces is rolled. The discrete random variable Y represents the score which is uppermost. The probability distribution of Y is shown in the table below, where a and b are constants.

y	1	2	3	4	5	6	7	8
P($Y = y$)	a	a	a	b	b	b	0.11	0.05

Given that $\mathrm{E}(Y) = 4.02$

(c) form and solve two equations in a and b to show that $a = 0.15$
You must show your working.

(*Solutions relying on calculator technology are not acceptable.*)

(3)

(d) Show that $\mathrm{E}(Y^2) = 20.7$

(2)

(e) Find $\mathrm{Var}(5 - 2Y)$

(3)

These dice are each rolled once. The scores on the two dice are independent.

(f) Find the probability that the sum of these two scores is 3

(2)

Question 6 continued

Question 6 continued

Question 6 continued

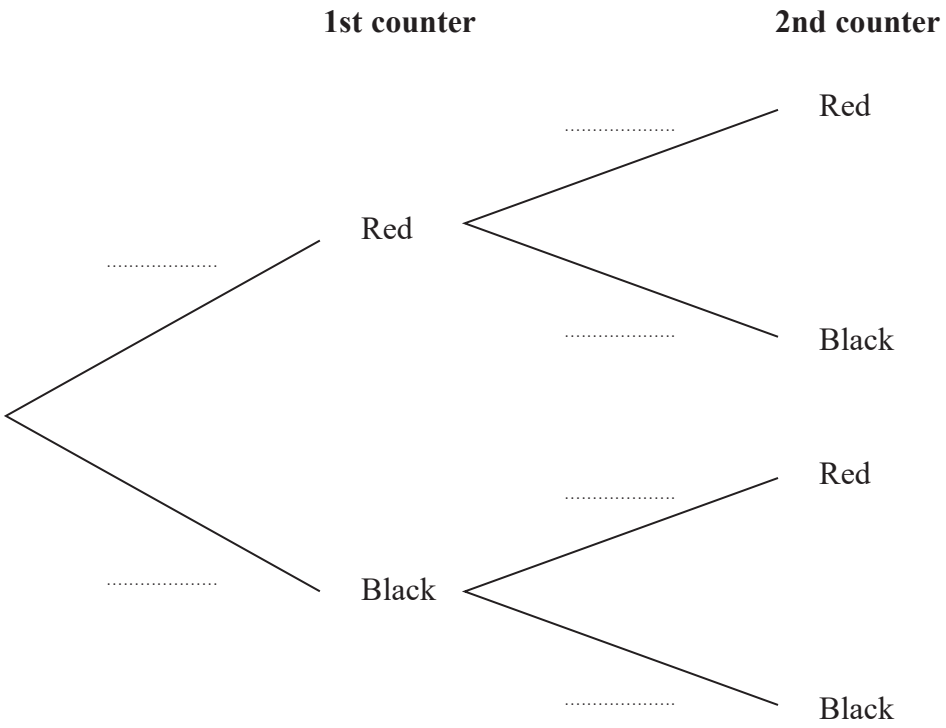
(Total for Question 6 is 14 marks)

7. A box contains only red counters and black counters.

There are n red counters and $n + 1$ black counters.

Two counters are selected at random, one at a time **without** replacement, from the box.

(a) Complete the tree diagram for this information. Give your probabilities in terms of n where necessary.



(3)

(b) Show that the probability that the two counters selected are different colours is

$$\frac{n+1}{2n+1}$$

(2)

The probability that the two counters selected are different colours is $\frac{25}{49}$

(c) Find the total number of counters in the box before any counters were selected.

(2)

Given that the two counters selected are different colours,

(d) find the probability that the 1st counter is black.
You must show your working.

(2)

Question 7 continued

Question 7 continued

Question 7 continued

(Total for Question 7 is 9 marks)

8. An orchard produces apples.

The weights, A grams, of its apples are normally distributed with mean μ grams and standard deviation σ grams.

It is known that

$$P(A < 162) = 0.1 \text{ and } P(162 < A < 175) = 0.7508$$

(a) Calculate the value of μ and the value of σ

(5)

A second orchard also produces apples.

The weights, B grams, of its apples have distribution $B \sim N(215, 10^2)$

An outlier is a value that is

greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$ or smaller than $Q_1 - 1.5 \times (Q_3 - Q_1)$

An apple is selected at random from this second orchard.

Using $Q_3 = 221.74$ grams,

(b) find the probability that this apple is an outlier.

(5)

Question 8 continued

Question 8 continued

Question 8 continued

Question 8 continued

(Total for Question 8 is 10 marks)

TOTAL FOR PAPER IS 75 MARKS