Please check the examination details below before entering your candidate information

Candidate surname StemUp Model Solution | 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 ▶

P72072A

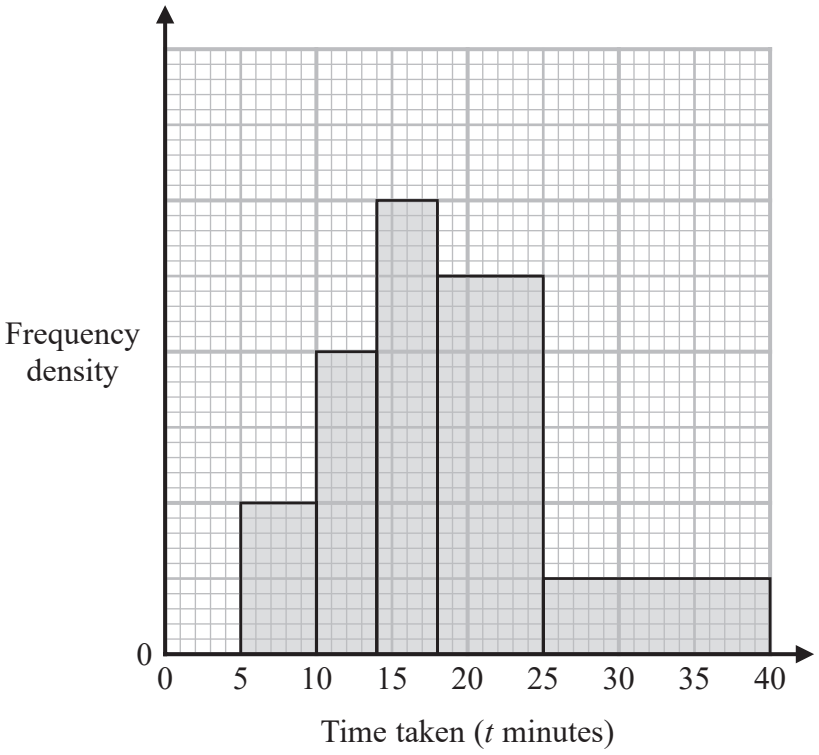
©2023 Pearson Education Ltd.

J:1/1/1/

P 7 2 0 7 2 A 0 1 2 4

Pearson

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	35	15

(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)

(a) 18 – 25: 5 × 7 = 35 25 – 40: 1 × 15 = 15 (see table)

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

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

Question 1 continued

(b) $10 + 16 + \frac{16 - 14}{18 - 14} \times 24 = 10 + 16 + 12$

$= 38$ people

(c) $\Sigma ft = 7.5 \times 10 + 12 \times 16 + 16 \times 24 + 21.5 \times 35 + 32.5 \times 15$

$= 75 + 192 + 384 + 752.5 + 487.5 = 1891$

Mean $= \frac{1891}{100} = 18.91$ minutes

(d) s.d. $= \sqrt{\left[\frac{41033}{100} - \left(\frac{1891}{100}\right)^2\right]}$

$= \sqrt{(410.33 - 357.5881)} = \sqrt{52.7419}$

$= 7.26236\ldots = 7.26$ minutes

(e) Q_1: $100 \div 4 = 25$th value, in the class $10 - 14$

$Q_1 = 10 + \frac{25 - 10}{16} \times (14 - 10) = 10 + \frac{15}{16} \times 4 = 13.75$

$IQR = Q_3 - Q_1 = 23 - 13.75$

$IQR = 9.25$ minutes

Question 1 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 1 continued

(Total for Question 1 is 10 marks)

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

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)

(a) Tree diagram completed on page 7

(b) $P(GB \text{ or } BG) = \frac{5}{9} \times \frac{4}{8} + \frac{4}{9} \times \frac{5}{8} = \frac{20}{72} + \frac{20}{72}$

$= \frac{40}{72} = \frac{5}{9}$

(c) $P(GGG) + P(BBB) = \frac{5}{9} \times \frac{4}{8} \times \frac{8}{13} + \frac{4}{9} \times \frac{3}{8} \times \frac{7}{13}$

$= \frac{160}{936} + \frac{84}{936} = \frac{244}{936} = \frac{61}{234}$ $(= 0.261)$

(d) $P(GGG \mid \text{all same}) = \frac{P(GGG)}{P(\text{all same})} = \frac{160/936}{244/936}$

$= \frac{20/117}{61/234} = \frac{40}{61}$ $(= 0.656)$

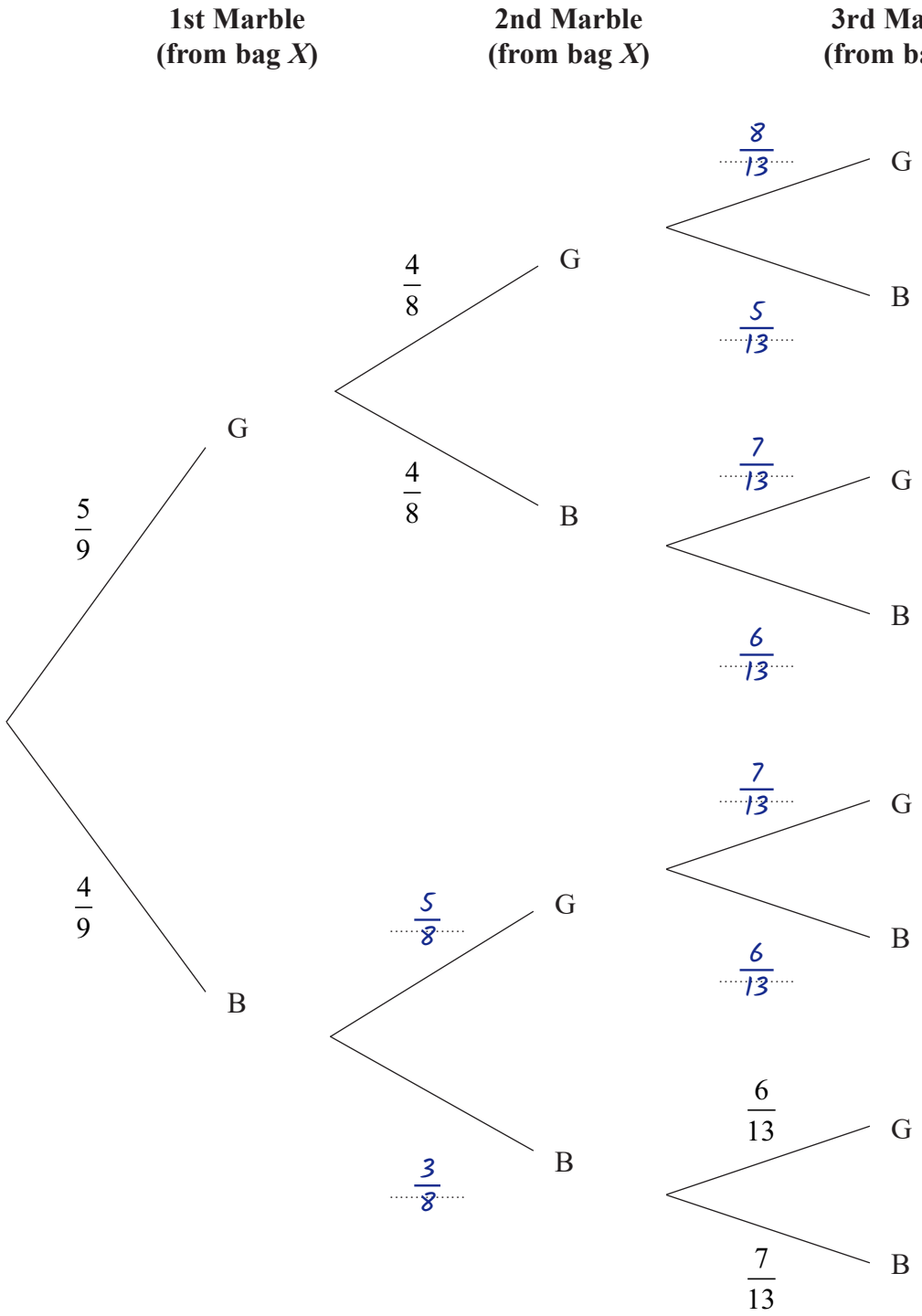
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

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

Question 2 continued



Question 2 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

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

Question 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 $\text{Var}(X) = 0.56$

(c) find the possible values of a

(6)

(a) $E(X) = 2a + 3 \times 0.4 + 4(0.6 - a)$

$= 2a + 1.2 + 2.4 - 4a$

$E(X) = 3.6 - 2a$

(b) Probabilities: $a > 0$ and $0.6 - a > 0$, so $0 < a < 0.6$

$a = 0$: $E(X) = 3.6 - 2 \times 0 = 3.6$

$a = 0.6$: $E(X) = 3.6 - 2 \times 0.6 = 2.4$

$2.4 < E(X) < 3.6$

(c) $E(X^2) = 4a + 9 \times 0.4 + 16(0.6 - a)$

$= 4a + 3.6 + 9.6 - 16a = 13.2 - 12a$

$\text{Var}(X) = E(X^2) - [E(X)]^2 = 13.2 - 12a - (3.6 - 2a)^2$

$0.56 = 13.2 - 12a - (12.96 - 14.4a + 4a^2)$

$0.56 = 0.24 + 2.4a - 4a^2$

$-4a^2 + 2.4a - 0.32 = 0 \Rightarrow 25a^2 - 15a + 2 = 0$

$(5a - 1)(5a - 2) = 0$

$a = \frac{1}{5}$ or $a = \frac{2}{5}$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

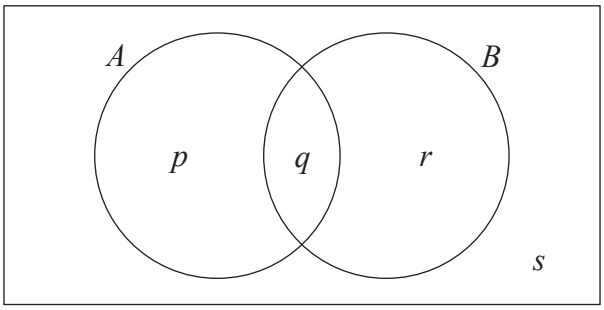
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 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)

(i)(a) $P(A) = p + q = \frac{7}{25}$ $\quad P(B) = q + r = \frac{1}{5}$

$P[(A \cap B') \cup (A' \cap B)] = p + r = \frac{8}{25}$

Adding: $2p + 2q + 2r = \frac{7}{25} + \frac{1}{5} + \frac{8}{25} = \frac{7+5+8}{25}$

$= \frac{20}{25} = \frac{4}{5}$ as required

(b) $p + q + r = \frac{2}{5}$

$p = (p + q + r) - (q + r) = \frac{2}{5} - \frac{1}{5}$ $\quad p = \frac{1}{5}$

$q = (p + q + r) - (p + r) = \frac{10}{25} - \frac{8}{25}$ $\quad q = \frac{2}{25}$

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

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

Question 4 continued

$r = (p + q + r) - (p + q) = \frac{10}{25} - \frac{7}{25}$ $\quad r = \frac{3}{25}$

$p + q + r + s = 1 \quad \Rightarrow \quad s = 1 - \frac{2}{5} \quad s = \frac{3}{5}$

(ii) $P(C) + P(D) = \frac{x}{x+5} + \frac{5}{x} = \frac{x^2 + 5(x+5)}{x(x+5)}$

$= \frac{x^2 + 5x + 25}{x^2 + 5x} = 1 + \frac{25}{x^2 + 5x}$

$x > 0 \Rightarrow x^2 + 5x > 0$, so $\frac{25}{x^2 + 5x} > 0 \Rightarrow P(C) + P(D) > 1$

If C and D were mutually exclusive, $P(C \cup D) = P(C) + P(D) > 1$,
which is impossible since a probability cannot exceed 1.
So $P(C \cap D) > 0$, hence C and D cannot be mutually exclusive.

Question 4 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 4 continued

(Total for Question 4 is 13 marks)

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

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)

(a) $P(L < 3.86) = P\left(Z < \frac{3.86 - 4.5}{0.4}\right) = P(Z < -1.6)$

$= 1 - 0.9452 = 0.0548$

(b)(i) $P(L < Q_3) = 0.75 \Rightarrow \frac{Q_3 - 4.5}{0.4} = 0.6745$

$Q_3 = 4.5 + 0.6745 \times 0.4 = 4.7698 = 4.77$ mm

(ii) By symmetry, $Q_1 = 4.5 - 0.6745 \times 0.4 = 4.2302$

$Q_1 = 4.23$ mm

(c) $1.5 \times (Q_3 - Q_1) = 1.5 \times (4.7698 - 4.2302) = 1.5 \times 0.5396 = 0.8094$

Upper limit $= 4.7698 + 0.8094 = 5.5792 = 5.58$ mm

Lower limit $= 4.2302 - 0.8094 = 3.4208 = 3.42$ mm

(d) $P(3.42 < L < 5.58) = P\left(\frac{3.42 - 4.5}{0.4} < Z < \frac{5.58 - 4.5}{0.4}\right)$

$= P(-2.7 < Z < 2.7) = 2 \times 0.9965 - 1$

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

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

Question 5 continued

$= 0.9930 \quad = 0.993$ (3 d.p.) as required

(e) $P(5 < L < 5.58) = P\left(\frac{5 - 4.5}{0.4} < Z < \frac{5.58 - 4.5}{0.4}\right)$

$= P(1.25 < Z < 2.7) = 0.9965 - 0.8944 = 0.1021$

$P(L > 5 \mid \text{not outlier}) = \frac{P(5 < L < 5.58)}{P(3.42 < L < 5.58)} = \frac{0.1021}{0.9930}$

$= 0.10282\ldots \quad = 0.103$

Question 5 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 5 continued

(Total for Question 5 is 17 marks)

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

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)

(a) For every 1 °C increase in temperature, an extra 2.72 grams of sugar will dissolve in the 100 grams of water.

(b) $y = 151.2 + 2.72 \times 90 = 151.2 + 244.8$
$= 396$ grams

(c) Unreliable: 90 °C is outside the range of the data ($10 \leq x \leq 80$), so this is extrapolation and the estimate may be unreliable.

(d) $\bar{y} = \frac{3119}{12} = 259.9166...$ $\quad \bar{y} = 151.2 + 2.72\bar{x}$

$\bar{x} = \frac{259.9166... - 151.2}{2.72} = 39.96936... \Rightarrow \Sigma x = 12\bar{x} = 479.63235...$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

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

Question 6 continued

$S_{yy} = 851093 - \frac{3119^2}{12} = 40412.9166...$

$S_{xx} = 24500 - \frac{479.63235...^2}{12} = 5329.4005...$

$b = S_{xy} \div S_{xx} \Rightarrow S_{xy} = 2.72 \times 5329.4005... = 14495.9693...$

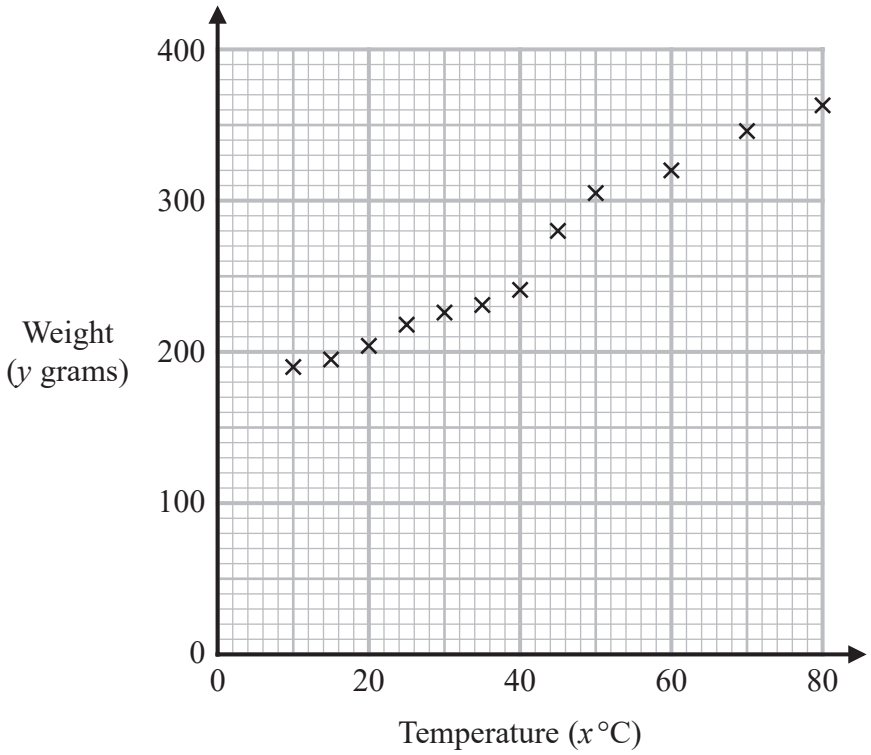
$r = \frac{S_{xy}}{\sqrt{(S_{xx}\, S_{yy})}} = \frac{14495.9693...}{\sqrt{(5329.4005... \times 40412.9166...)}}$

$= 0.98775...$ $= 0.988$ (3 d.p.) as required

(e) The points on the scatter diagram lie reasonably close to a straight line (positive linear correlation) and $r = 0.988$ is close to 1, so a linear regression model is suitable to describe the relationship.

Question 6 continued

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



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

Question 6 continued

(Total for Question 6 is 14 marks)

TOTAL FOR PAPER IS 75 MARKS

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA