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 ▶

P71200A

©2022 Pearson Education Ltd.

Q:1/1/1/1/

P 7 1 2 0 0 A 0 1 2 4

Pearson

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)

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

Question 1 continued

(a) $n = 38$: Q_1 is the 10th value, Q_3 is the 29th value

Q_2 = mean of 19th and 20th values = 26

$$\frac{24 + (20 + w)}{2} = 26 \quad \Rightarrow \quad 20 + w = 28 \quad \Rightarrow \quad w = 8$$

$w = 8$, $x = 19$, $y = 37$

(b) $Q_3 + 1.0 \times (Q_3 - Q_1) = 37 + 1.0 \times (37 - 19) = 37 + 18 = 55$

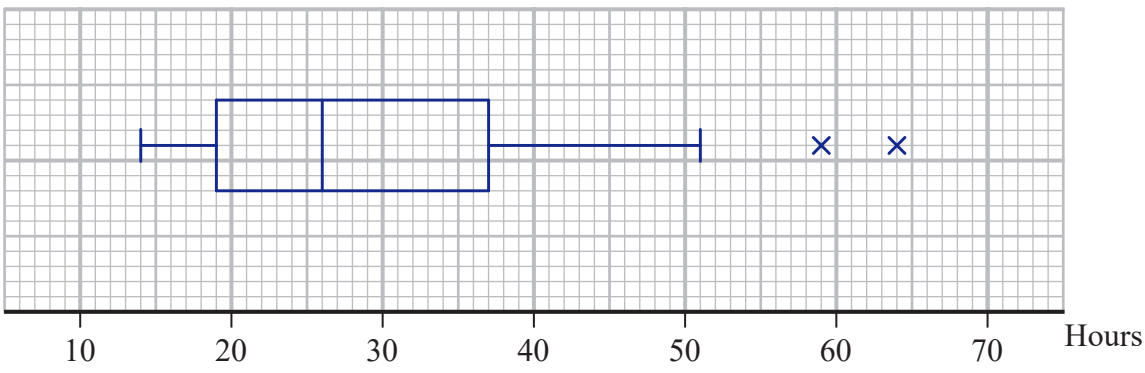
Outliers are values greater than 55: 59 and 64

(c) Box plot drawn on the grid below (outliers 59 and 64 marked ×).

(d) Skewness $= \frac{(37 - 26) - (26 - 19)}{(37 - 19)} = \frac{11 - 7}{18} = \frac{4}{18}$

$= 0.2222\ldots = 0.22$ (2 s.f.)

(e) The mean uses all of the data values, including the outliers (the long weeks of 59 and 64 hours), so it gives a better idea of the total hours Landacre will have to pay for.



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

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

Question 1 continued

DO NOT WRITE IN THIS AREA

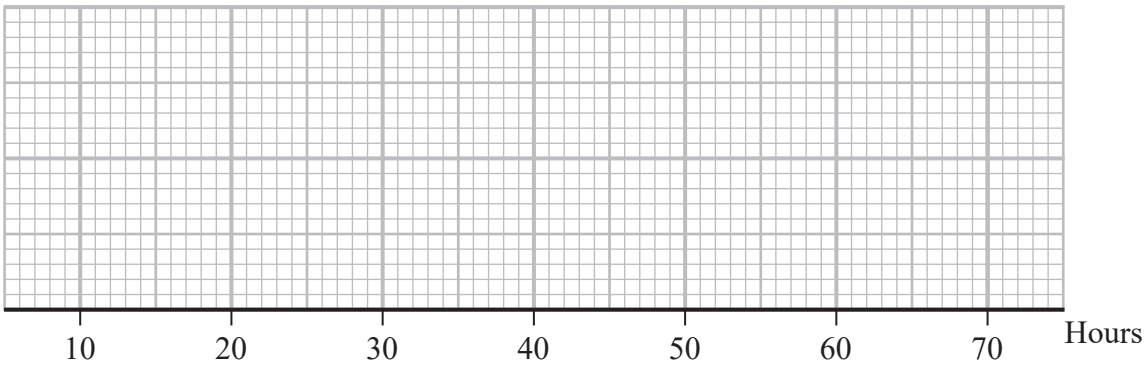
DO NOT WRITE IN THIS AREA

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 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)

(a) $S_{gg} = \Sigma g^2 - \frac{(\Sigma g)^2}{n} = 3624.41 - \frac{144.84^2}{9} = 1293.4516$

$r = \frac{S_{tg}}{\sqrt{(S_{tt}\, S_{gg})}} = \frac{40.25}{\sqrt{(1.29 \times 1293.4516)}} = \frac{40.25}{40.8479...}$

$r = 0.98536... = 0.985$

(b) As the population increases, the GDP increases.

(c) $b = \frac{S_{tg}}{S_{tt}} = \frac{40.25}{1.29} = 31.20155...$

$\bar{t} = \frac{7.87}{9} = 0.87444..., \quad \bar{g} = \frac{144.84}{9} = 16.09333...$

$a = \bar{g} - b\bar{t} = 16.09333... - 31.20155... \times 0.87444...$

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

Question 2 continued

$a = -11.19068...$

$g = -11.2 + 31.2t$

(d) For every increase of 1 million in the population, the GDP increases by 31.2 billion dollars (on average).

(e)(i) $t = 7$: $g = -11.19068... + 31.20155... \times 7$

$g = 207.22... \approx 207$ billion dollars

(ii) Unreliable: a population of 7 million is far above the populations in the sample (mean $t = 0.874$ million), so this is extrapolation outside the range of the data.

(f) Increase in g = b × increase in t

$0.1 = 31.20155... \times x$

$$x = \frac{0.1}{31.20155...} = 0.0032049... \approx 0.0032 \text{ (million)}$$

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

Question 2 continued

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

Question 2 continued

(Total for Question 2 is 14 marks)

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

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)

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

Question 3 continued

(a) $27 < l \leqslant 30$ has class width 3 and bar width 1.5 cm, so 1 cm ↔ 0.5 cm

Width = $5 \times 0.5 = 2.5$ cm

Area = $1.5 \times 4 = 6$ cm^2 represents 15 logs, so 1 log ↔ 0.4 cm^2

35 logs ↔ $35 \times 0.4 = 14$ cm^2

Height = $14 \div 2.5 = 5.6$ cm

width = 2.5 cm, height = 5.6 cm

(b) Median is the 44th value ($88 \div 2$); 19 logs are $\leqslant 20$

$$Q_2 = 20 + \frac{44 - 19}{35} \times 5 = 20 + \frac{25}{35} \times 5$$

$= 23.571\ldots \approx 23.6$ cm

(c) Logs in $25 < l \leqslant 26$: $\frac{26 - 25}{27 - 25} \times 16 = 8$

$x = 19 + 35 + 8 = 62$ as required

(d) P(all 4 fit) $= \frac{62}{88} \times \frac{61}{87} \times \frac{60}{86} \times \frac{59}{85}$

$$= \frac{13388280}{55965360} = 0.23922\ldots \approx 0.239$$

(e)(i) $\bar{y} = \frac{\Sigma y}{n} = \frac{924}{88} = 10.5$

$y = 0.5w - 255 \Rightarrow w = 2(y + 255)$

$\bar{w} = 2(10.5 + 255) = 2 \times 265.5 = 531$

Mean of W = 531 g

(ii) $\text{Var}(Y) = \frac{\Sigma y^2}{n} - \bar{y}^2 = \frac{12862}{88} - 10.5^2$

$= 146.1590\ldots - 110.25 = 35.9090\ldots$

$\text{Var}(W) = 2^2 \times \text{Var}(Y) = 4 \times 35.9090\ldots = 143.636\ldots$

Variance of W ≈ 144 g^2

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

Question 3 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 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 $P(N' | H)$

(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)

(a) $P(H \cup W) = P(H) + P(W) - P(H \cap W)$

Independent: $P(H \cap W) = P(H) \times P(W) = \frac{3}{8} P(W)$

$\frac{3}{4} = \frac{3}{8} + P(W) - \frac{3}{8} P(W)$

$\frac{3}{8} = \frac{5}{8} P(W)$

$P(W) = \frac{3}{8} \times \frac{8}{5} = \frac{3}{5}$ as required

(b) $P(H \cap N) = P(N)$, so N lies entirely inside H

$P(N' | H) = \frac{P(N' \cap H)}{P(H)} = \frac{3/8 - 1/15}{3/8} = \frac{37/120}{45/120}$

$= \frac{37}{45}$ $(= 0.8222...)$

(c) $P(H \cap W) = \frac{3}{8} \times \frac{3}{5} = \frac{9}{40}$

W only: $\frac{3}{5} - \frac{9}{40} = \frac{15}{40} = \frac{3}{8}$

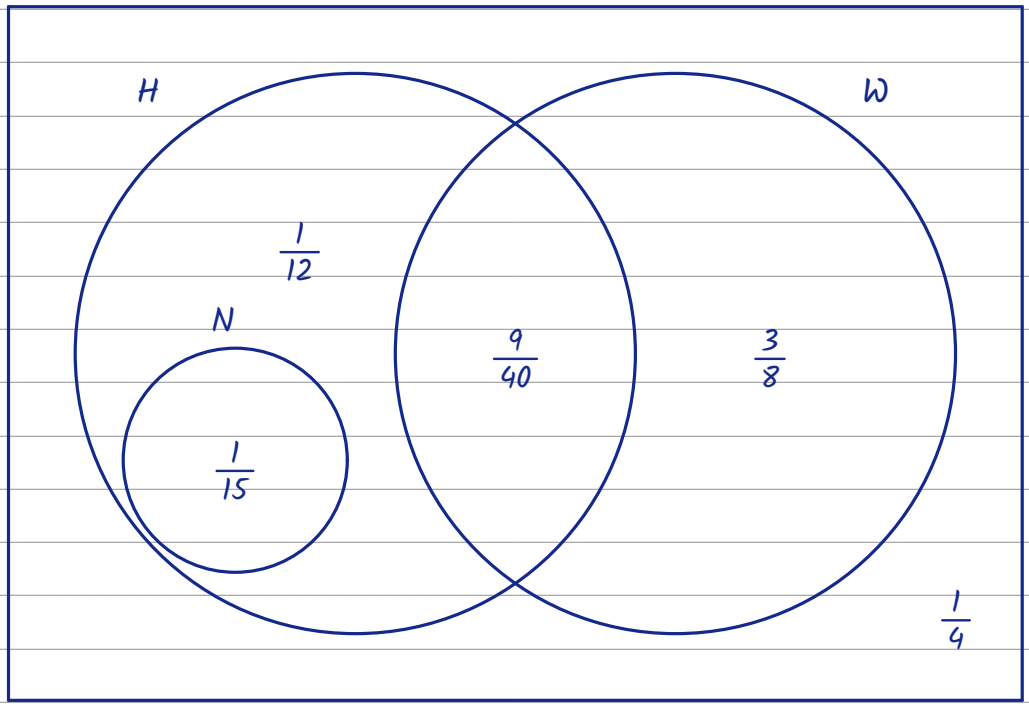
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

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

Question 4 continued

H only (not N): $\frac{3}{8} - \frac{9}{40} - \frac{1}{15} = \frac{45 - 27 - 8}{120} = \frac{1}{12}$

Outside: $1 - P(H \cup W) = 1 - \frac{3}{4} = \frac{1}{4}$



Question 4 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 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)

(a) $E(R^2) = \Sigma r^2 P(R = r)$

$= 2^2 \times 0.25 + 3^2 \times 0.3 + 4^2 \times 0.15 + 5^2 \times 0.1 + 6^2 \times 0.2$

$= 1 + 2.7 + 2.4 + 2.5 + 7.2 = 15.8$ as required

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 5 continued

(b) $Var(R) = E(R^2) - [E(R)]^2 = 15.8 - 3.7^2 = 15.8 - 13.69 = 2.11$

$sd(R) = \sqrt{2.11} = 1.45258... = 1.45$ (2 d.p.)

(c) $d = 1$

(d) Sum of probabilities: $0.1 + 0.2 + 0.1 + a + b = 1 \Rightarrow a + b = 0.6$

$E(Y) = 2\times0.1 + 3\times0.2 + 4\times0.1 + 5a + 6b = 4.55$

$1.2 + 5a + 6b = 4.55 \Rightarrow 5a + 6b = 3.35$

$5(0.6 - b) + 6b = 3.35 \Rightarrow 3 + b = 3.35 \Rightarrow b = 0.35$

$a = 0.6 - 0.35 = 0.25$

$c = F(5) = 0.4 + a = 0.4 + 0.25 = 0.65$

(e) Jessie not 2, Pabel not 2, then Jessie 2:

$P = 0.9 \times 0.75 \times 0.1 = 0.0675$

(f) If Jessie scores 2 she wins, so Pabel has no spin; same score must be 3, 4, 5 or 6:

$P = 0.2\times0.3 + 0.1\times0.15 + 0.25\times0.1 + 0.35\times0.2$

$= 0.06 + 0.015 + 0.025 + 0.07 = 0.17$

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

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 14 marks)

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

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

(a) Find $P(V > 104.9)$

(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 $P(V > v \mid V < 104.9) = 0.2801$

(6)

(a) $P(V > 104.9) = P\left(Z > \frac{104.9 - 100}{2.5}\right) = P(Z > 1.96)$

$= 1 - 0.9750 = 0.025$

(b) Expected number $= 150 \times 0.025 = 3.75$

(c) $P(V > v \mid V < 104.9) = \frac{P(v < V < 104.9)}{P(V < 104.9)} = 0.2801$

$P(V < 104.9) = 1 - 0.025 = 0.975$

$P(v < V < 104.9) = 0.975 \times 0.2801 = 0.2730975$

$P(v < V < 104.9) = P(V < 104.9) - P(V < v) = 0.975 - P(V < v)$

$P(V < v) = 0.975 - 0.2730975 = 0.7019025$

$\frac{v - 100}{2.5} = 0.53$

$v = 100 + 2.5 \times 0.53 = 101.325$

$v = 101.33$ (2 d.p.)

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

Question 6 continued

(Total for Question 6 is 11 marks)

TOTAL FOR PAPER: 75 MARKS

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA