Please check the examination details below before entering your candidate information

Candidate surname StemUp Model Solution | 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 ▶

P63150A

©2021 Pearson Education Ltd.

1/1/1/1/1/1/1/

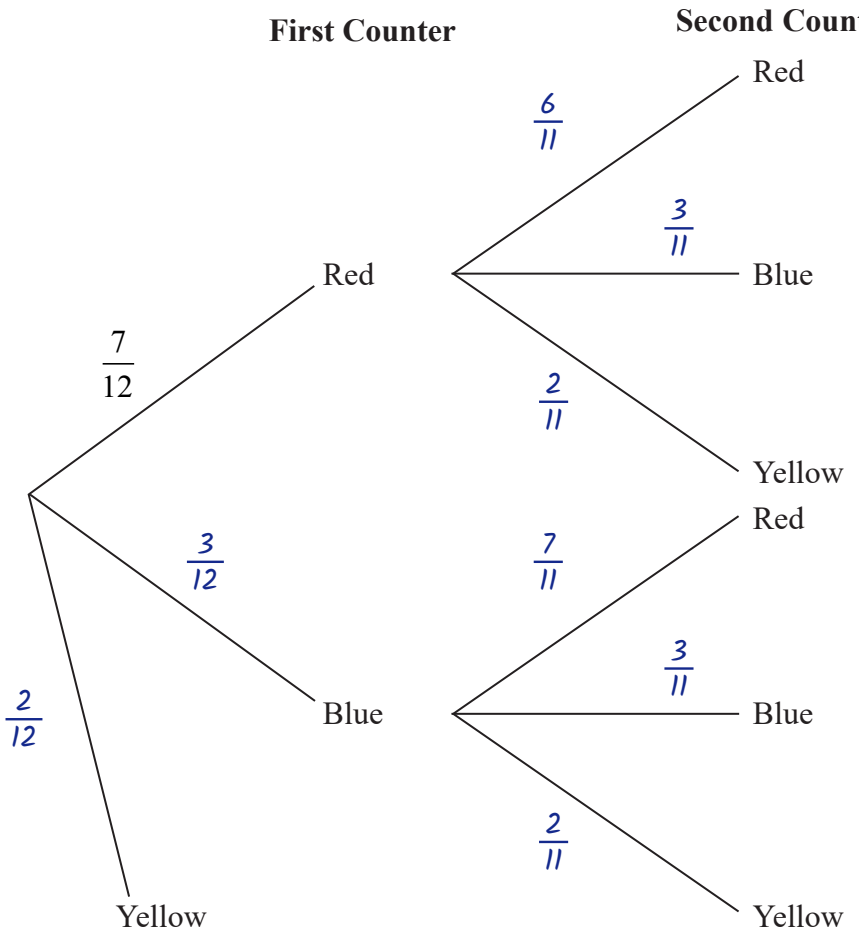
P 6 3 1 5 0 A 0 1 2 0

Pearson

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)

(b) $P(Y) = \frac{7}{12} \times \frac{2}{11} + \frac{3}{12} \times \frac{2}{11} + \frac{2}{12} = \frac{42}{132}$

$P(Y \text{ and 2 counters}) = \frac{7}{12} \times \frac{2}{11} + \frac{3}{12} \times \frac{2}{11} = \frac{20}{132}$

$P(2 \text{ counters} \mid Y) = \frac{P(Y \text{ and 2 counters})}{P(Y)} = \frac{20/132}{42/132} = \frac{20}{42}$

$= \frac{10}{21}$

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

Leave blank

Question 1 continued

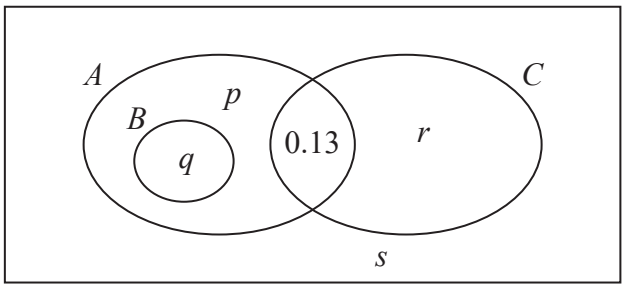
Q1

(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)

(a) B and C

(b) A and C independent: $P(A \cap C) = P(A) \times P(C)$

$0.13 = 0.65 \times P(C) \Rightarrow P(C) = 0.2$

$r = 0.2 - 0.13 = 0.07$

$P(A) + r + s = 1 \Rightarrow 0.65 + 0.07 + s = 1$

$s = 1 - 0.72 = 0.28$

$r = 0.07, \quad s = 0.28$

(c) $P(A \cap C') = p + q = 0.65 - 0.13 = 0.52$

$P(B \cup C) = q + 0.13 + 0.07 = 0.2 + q$

$(A \cap C') \cap (B \cup C) = B$, so $P[(A \cap C') \cap (B \cup C)] = q$

Independent: $P(A \cap C') \times P(B \cup C) = q$

$0.52(0.2 + q) = q$

$0.104 + 0.52q = q \Rightarrow 0.48q = 0.104$

$q = \frac{0.104}{0.48} = \frac{13}{60}$

$p = 0.52 - \frac{13}{60} = \frac{156}{300} - \frac{65}{300} = \frac{91}{300}$

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

Leave blank

Question 2 continued

$p = \frac{91}{300}$, $q = \frac{13}{60}$

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)

(a) Class width 3 cm is 1.5 cm wide, so 1 cm of length ↔ 0.5 cm

Width = 5 × 0.5 = 2.5 cm

Area 1.5 × 1 = 1.5 cm^2 for 5 carrots, so 30 carrots ↔ 6 × 1.5 = 9 cm^2

Height = 9 ÷ 2.5 = 3.6 cm

width = 2.5 cm, height = 3.6 cm

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

Leave blank

Question 3 continued

(b) Median is the 50th value: 34 carrots below 12, 59 below 15

$$Q_2 = 12 + \frac{50 - 34}{25} \times 3 = 12 + \frac{16}{25} \times 3$$

$$= 13.92 \approx 13.9 \text{ cm}$$

(c)(i) $\Sigma fx = 5 \times 6.5 + 13 \times 9 + 16 \times 11 + 25 \times 13.5 + 30 \times 17.5 + 11 \times 24$

$$= 32.5 + 117 + 176 + 337.5 + 525 + 264 = 1452$$

$$\bar{x} = \frac{1452}{100} = 14.52 \approx 14.5 \text{ cm}$$

(ii) $\Sigma fx^2 = 5 \times 6.5^2 + 13 \times 9^2 + 16 \times 11^2 + 25 \times 13.5^2 + 30 \times 17.5^2 + 11 \times 24^2$

$$= 211.25 + 1053 + 1936 + 4556.25 + 9187.5 + 6336 = 23280$$

$$\sigma^2 = \frac{23280}{100} - 14.52^2 = 232.8 - 210.8304 = 21.9696$$

$$\sigma = \sqrt{21.9696} = 4.68717\ldots \approx 4.69 \text{ cm}$$

(d) Number between 9 cm and 22 cm:

$$\frac{1}{2} \times 13 + 16 + 25 + 30 + \frac{2}{8} \times 11 = 6.5 + 71 + 2.75 = 80.25$$

$$\text{Proportion} = \frac{80.25}{100} = 0.8025 \approx 0.803$$

(e) Longer than 22 cm: $\frac{6}{8} \times 11 = 8.25$ carrots

Shorter than 9 cm: $100 - 80.25 - 8.25 = 11.5$ carrots

$$\text{Mean profit} = \frac{2.2 \times 80.25 + 0.8 \times 8.25 - 1.2 \times 11.5}{100}$$

$$= \frac{176.55 + 6.6 - 13.8}{100} = \frac{169.35}{100} = 1.6935$$

$\approx$ 1.7 pence per carrot

Leave blank

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

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)

(a) $W \sim N(165, 35^2)$

$P(W < 120) = P\left(Z < \frac{120 - 165}{35}\right) = P(Z < -1.2857...)$

$= 1 - 0.9007285... = 0.099271... \approx 0.0993$

(b) Upper limit u: $P(W > u) = \frac{1}{3}$, so $P(W < u) = \frac{2}{3}$

$\frac{u - 165}{35} = 0.4307$ (calculator 0.430727...)

$u = 165 + 35 \times 0.430727... = 180.0754...$

Lower limit l: by symmetry $\frac{l - 165}{35} = -0.430727...$

$l = 165 - 35 \times 0.430727... = 149.9245...$

Medium letters weigh between 150 g and 180 g

(c) $P(W < 200 \mid W > 180) = \frac{P(180 < W < 200)}{P(W > 180)}$

$P(W < 200) = P\left(Z < \frac{200 - 165}{35}\right) = P(Z < 1) = 0.8413$

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

Leave blank

Question 4 continued

$P(180 < W < 200) = 0.84134... - \frac{2}{3} = 0.17468...$

$P(W < 200 \mid \text{heavy}) = \frac{0.17468...}{1/3} = 0.52403... \approx 0.524$

(d) Each letter is heavy, medium or light with probability 1/3

One of each in any order (3! orders):

$P = \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times 3! = \frac{6}{27} = \frac{2}{9}$

Leave blank

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

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)

(a) $E(X) = -2a - b + 0 \times c + b + 4a = 2a$

$2a = 0.5 \Rightarrow a = 0.25$

(b) $E(X^2) = (-2)^2a + (-1)^2b + 0 + 1^2b + 4^2a = 20a + 2b$

$= 20(0.25) + 2b = 5 + 2b$

$\text{Var}(X) = E(X^2) - [E(X)]^2 \Rightarrow 5 + 2b - 0.5^2 = 5.01$

$4.75 + 2b = 5.01 \Rightarrow 2b = 0.26 \Rightarrow b = 0.13$

$\Sigma P(X = x) = 1$: $2a + 2b + c = 1$

$c = 1 - 2(0.25) - 2(0.13) = 1 - 0.5 - 0.26 = 0.24$

$b = 0.13, \quad c = 0.24$

(c)(i) $E(Y) = 5 - 8E(X) = 5 - 8 \times 0.5 = 1$

(ii) $\text{Var}(Y) = (-8)^2 \text{Var}(X) = 64 \times 5.01$

$= 320.64 \approx 321$

(d) $4X^2 > Y \Rightarrow 4X^2 > 5 - 8X$

$4X^2 + 8X - 5 > 0$

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

Leave blank

Question 5 continued

$(2X - 1)(2X + 5) > 0$

Critical values $X = 0.5$ and $X = -2.5$, so $X > 0.5$ or $X < -2.5$

X takes values $-2, -1, 0, 1, 4$, so only $X = 1$ or $X = 4$

$P(4X^2 > Y) = P(X = 1) + P(X = 4) = b + a$

$= 0.13 + 0.25 = 0.38$

Leave blank

Question 5 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

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

Leave blank

Question 5 continued

Q5

(Total 15 marks)

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)

(a) $S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 42.63 - \frac{23.7^2}{16}$

$= 42.63 - 35.105625 = 7.524375$ as required

(b) The regression line passes through $(\bar{x}, \bar{y})$:

$\bar{y} = \frac{23.7}{16} = 1.48125$, $1.48125 = 3.684 - 0.3242\bar{x}$

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

Leave blank

Question 6 continued

$\bar{x} = \dfrac{3.684 - 1.48125}{0.3242} = 6.79441...$

$\Sigma x = 16 \times 6.79441... = 108.71067...$

$S_{xx} = \Sigma x^2 - \dfrac{(\Sigma x)^2}{n} = 756.81 - \dfrac{(108.71067...)^2}{16}$

$= 18.18435... \approx 18.2$

(c) $b = \dfrac{S_{xy}}{S_{xx}} \Rightarrow S_{xy} = -0.3242 \times 18.18435... = -5.89536...$

$r = \dfrac{S_{xy}}{\sqrt{(S_{xx}\, S_{yy})}} = \dfrac{-5.89536...}{\sqrt{(18.18435... \times 7.524375)}}$

$= -0.50399... \approx -0.504$

(d) Substituting $x = 2$ into the regression line:

$y = 3.684 - 0.3242 \times 2 = 3.0356$, which is over 3%

(e) Standard deviation of $x = \sqrt{\left(\dfrac{S_{xx}}{n}\right)} = \sqrt{\left(\dfrac{18.18435...}{16}\right)} = 1.06607...$

Mean of $x = 6.79441...$

Minimum $= 6.79441... - 3 \times 1.06607... = 3.5965... \approx 3.6$

Maximum $= 6.79441... + 3 \times 1.06607... = 9.9929... \approx 10$

(f) $x = 2$ lies outside the range of the data (about 3.6 to 10),
so the probability of $x = 2$ being in the range is very small;
this is extrapolation, so Behrouz's claim is unreliable.

(g) The regression line of y on x should only be used to estimate
y (wages) from x (unemployment). To estimate unemployment
he would need the regression line of x on y,
so Andi's suggestion is not suitable.

Leave blank

Question 6 continued

Q6

(Total 16 marks)

TOTAL FOR PAPER: 75 MARKS

END

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA