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

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 ▶

P60578A

©2020 Pearson Education Ltd.

1/1/1/1/1/

P 6 0 5 7 8 A 0 1 2 4

Pearson

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)

$\Sigma P(X = x) = 1 : \quad 0.15 + a + b + c + 0.15 = 1$

$\Rightarrow \quad a + b + c = 0.7$

$F(1) = 0.63 : \quad 0.15 + a + b = 0.63$

$\Rightarrow \quad a + b = 0.48$

$\therefore \quad c = 0.7 - 0.48 = 0.22$

$E(X) = 1 : \quad -2(0.15) - a + b + 3c + 4(0.15) = 1$

$\Rightarrow \quad -a + b + 3c = 0.7$

subtracting $a + b + c = 0.7 : \quad 2c - 2a = 0 \quad \Rightarrow \quad a = c$

$\therefore \quad a = 0.22$

$b = 0.48 - 0.22 = 0.26$

$a = 0.22, \quad b = 0.26, \quad c = 0.22$

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. 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.

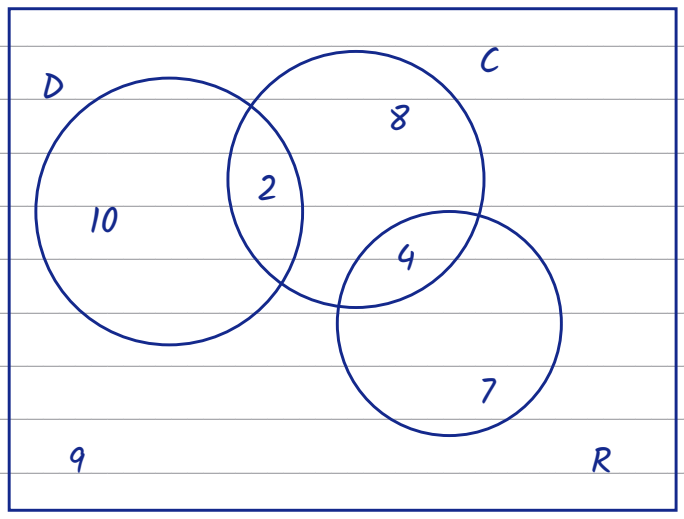
(e) Find the probability that Sarah's family also has a pet rabbit.

(2)

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

(2)

(a)



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

(b) D and R are mutually exclusive, since no family has both a dog and a rabbit, i.e. $P(D \cap R) = 0$

(c) $P(\text{exactly } 2) = \frac{2 + 4}{40} = \frac{3}{20}$

(d) $P(D \cap C) = \frac{2}{40} = \frac{1}{20}$

$P(D) = \frac{12}{40} = \frac{3}{10}$, $P(C) = \frac{14}{40} = \frac{7}{20}$

$P(D) \times P(C) = \frac{3}{10} \times \frac{7}{20} = \frac{21}{200}$

$\frac{1}{20} \neq \frac{21}{200}$ so $P(D \cap C) \neq P(D) \times P(C)$

$\therefore$ D and C are not independent

(e) $P(R \mid C) = \frac{P(R \cap C)}{P(C)} = \frac{4/40}{14/40} = \frac{4}{14}$

$= \frac{2}{7}$

(f) $P([D \cup R] \mid C') = \frac{P([D \cup R] \cap C')}{P(C')} = \frac{10 + 7}{10 + 7 + 9}$

$= \frac{17}{26}$

Leave blank

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

Leave blank

Question 2 continued

(Total 12 marks)

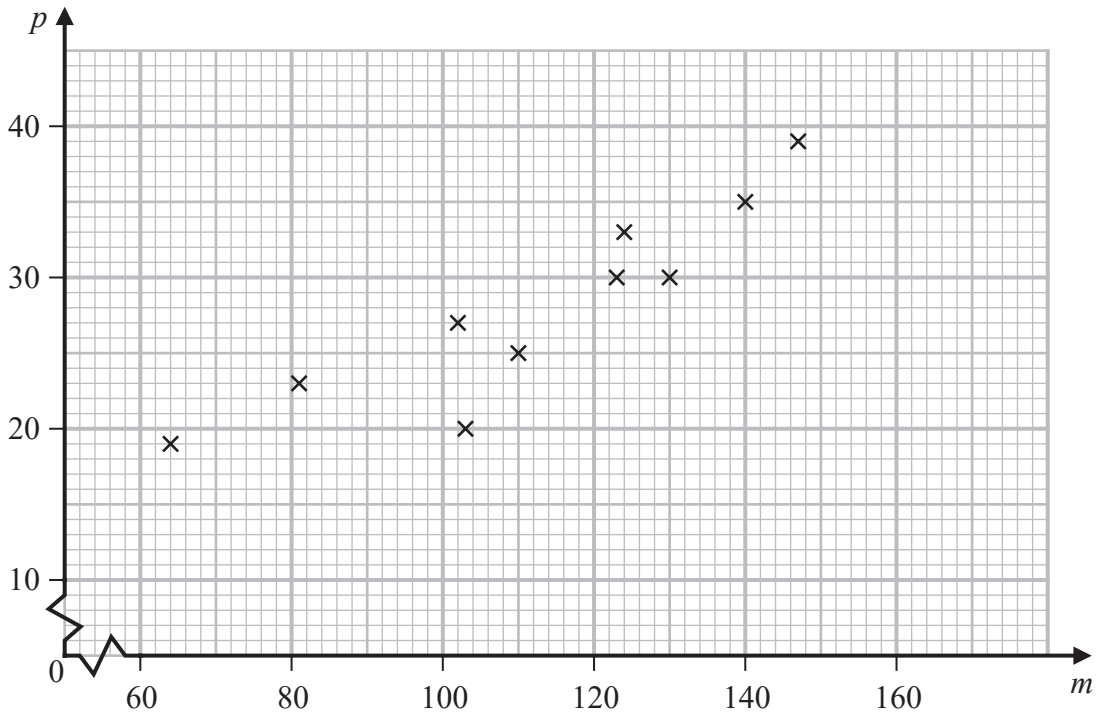
Q2

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)

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

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)

(a) $S_{mp} = \Sigma mp - \frac{\Sigma m \Sigma p}{n} = 32958 - \frac{1124 \times 281}{10}$

$= 32958 - 31584.4 = 1373.6$

(b) $r = \frac{S_{mp}}{\sqrt{(S_{mm}\, S_{pp})}} = \frac{1373.6}{\sqrt{(6046.4 \times 382.9)}}$

$= 0.90275\ldots = 0.903$

(c) r is close to 1 (the points lie close to a straight line),

so the data are consistent with the manager's belief.

(d) $\Sigma m = k\Sigma p \Rightarrow k = \frac{\Sigma m}{\Sigma p} = \frac{1124}{281} = 4$

$k = 4$

(e) $b = \frac{S_{mp}}{S_{mm}} = \frac{1373.6}{6046.4} = 0.227176\ldots$

$\bar{m} = \frac{1124}{10} = 112.4, \quad \bar{p} = \frac{281}{10} = 28.1$

$a = \bar{p} - b\bar{m} = 28.1 - 0.227176\ldots \times 112.4$

$= 2.56536\ldots$

$p = 2.57 + 0.227m$

Leave blank

Question 3 continued

(f) $p = 2.56536... + 0.227176... \times 70$

$= 18.467... \approx 18.5$ (about 18 policies)

(g) Manager's model: $p = 70 \div 4 = 17.5$

Jiang's model gives 18.5, which is higher.

Staff should use the manager's model, because it gives the lower estimate, so they are more likely to sell more policies than the estimate and be paid the bonus.

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 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

(You may use $\sum fx = 311.5$ and $\sum fx^2 = 1051.125$)

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)

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

(a) Width: class width 0.5 is half of 1.0, so width = 0.5 cm

Area: 16 babies ↔ 1 × 4 = 4 cm², so 1 cm² ↔ 4 babies

14 babies ↔ 14 ÷ 4 = 3.5 cm²

Height = $\frac{3.5}{0.5}$ = 7 cm

width = 0.5 cm, height = 7 cm

(b) Q_1 : $\frac{98}{4}$ = 24.5th value, in class $2.50 \leqslant w < 3.00$

$Q_1 = 2.50 + \frac{24.5 - 16}{24} \times 0.5 = 2.50 + 0.177...$

$= 2.677...$ = 2.68 kg

(c) $Q_2 - Q_1 = 3.14 - 2.68 = 0.46$

$Q_3 - Q_2 = 3.55 - 3.14 = 0.41$

$Q_2 - Q_1 > Q_3 - Q_2$ so the data have negative skew

(d)(i) mean = $\frac{\Sigma fx}{\Sigma f} = \frac{311.5}{98}$ = 3.178571... = 3.18 kg

(ii) $\sigma = \sqrt{\frac{1051.125}{98} - 3.178571...^2}$

$= \sqrt{0.622448...}$ = 0.788954... = 0.789 kg

(e) skewness = $\frac{3(3.178571... - 3.14)}{0.788954...}$

= 0.14666... = 0.15

(f)(i) Median is the 49th value. 40 babies are below 3.00 and 9 are in $3.00 \leqslant w < 3.25$, so the 49th value is at 3.25 (> 3.14). So the estimate of the median increases.

(ii) More of the 32 values (23) are now in the upper half $3.25 \leqslant w < 3.50$ than the lower half (9), so their midpoints are higher on average, Σfx increases and the mean increases.

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 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)

(a) $P(X < 7) = P\left(Z < \frac{7 - 10}{6}\right) = P(Z < -0.5)$

$= 1 - 0.6915$

$= 0.30853... = 0.309$

(b) $P(10 - k < X < 10 + k) = 0.60 \Rightarrow P(X < 10 + k) = 0.80$

$\frac{10 + k - 10}{6} = 0.8416$

$k = 6 \times 0.8416 = 5.0496 = 5.05$

(c) Area of rectangle $= X(X - 3)$

Need $X(X - 3) > 40$

$X^2 - 3X - 40 > 0$

$(X - 8)(X + 5) > 0$

critical values 8 and -5

so $X > 8$ or $X < -5$

$P(X > 8) + P(X < -5)$

$= P\left(Z > \frac{8 - 10}{6}\right) + P\left(Z < \frac{-5 - 10}{6}\right)$

$= P(Z > -0.333...) + P(Z < -2.5)$

$= 0.63055... + 0.00620...$

$= 0.63676... = 0.637$

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

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

(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)

(a) $\Sigma P(S = s) = 1$: $k + \frac{k}{2} + \frac{k}{3} + \frac{k}{4} + \frac{k}{5} + \frac{k}{6} = 1$

$k\left(\frac{60 + 30 + 20 + 15 + 12 + 10}{60}\right) = \frac{147k}{60} = 1$

$\therefore \quad k = \frac{60}{147} = \frac{20}{49}$

(b) $E(S) = 0 \times k + 1 \times \frac{k}{2} + 2 \times \frac{k}{3} + 3 \times \frac{k}{4} + 4 \times \frac{k}{5} + 5 \times \frac{k}{6}$

$= 3.55k = 3.55 \times \frac{20}{49}$

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

$= \frac{71}{49}$ $(= 1.4489...)$

(c) $P(S = 5) = \frac{k}{6} = \frac{20}{294} = \frac{10}{147}$

Expected profit $= 260 \times P(S = 5) - 10 = 260 \times \frac{10}{147} - 10$

$= 7.68707... = \$7.69$

(d) $R = 2$ means Roger wins exactly 2 rounds: he must win the 1st and the 2nd rounds, each with probability p, giving p^2, then he must lose the 3rd round, with probability $(1 - p)$, so he is knocked out. $\therefore P(R = 2) = p \times p \times (1 - p) = p^2(1 - p)$

(e)

r	0	1	2	3	4	5
$P(R = r)$	$1 - p$	$p(1 - p)$	$p^2(1 - p)$	$p^3(1 - p)$	$p^4(1 - p)$	p^5

(f) $E(\text{Roger's profit}) = 260 \times P(R = 5) - 10 = 260p^5 - 10$

$260p^5 - 10 \geqslant 7.68707...$

$p^5 \geqslant \frac{17.68707...}{260} = \frac{10}{147} = 0.068027...$

$p \geqslant \left(\frac{10}{147}\right)^{1/5} = 0.58416...$

smallest value of $p = 0.584$

Leave blank

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

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Q6

(Total 17 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