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

Thursday 4 June 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 ▶

P65761A

©2020 Pearson Education Ltd.

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

P 6 5 7 6 1 A 0 1 2 4

Pearson

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)

$P(X = x)$: $\frac{9}{k}, \frac{6}{k}, \frac{5}{k}, \frac{4}{k}, \frac{1}{k}$ for $x = -1, 2, 3, 4, 7$

$\Sigma P(X = x) = 1 \Rightarrow \frac{9 + 6 + 5 + 4 + 1}{k} = 1 \Rightarrow \frac{25}{k} = 1$

$k = 25$

$E(X) = \frac{1}{25}[(-1 \times 9) + (2 \times 6) + (3 \times 5) + (4 \times 4) + (7 \times 1)]$

$= \frac{-9 + 12 + 15 + 16 + 7}{25}$

$E(X) = \frac{41}{25}$ $(= 1.64)$

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

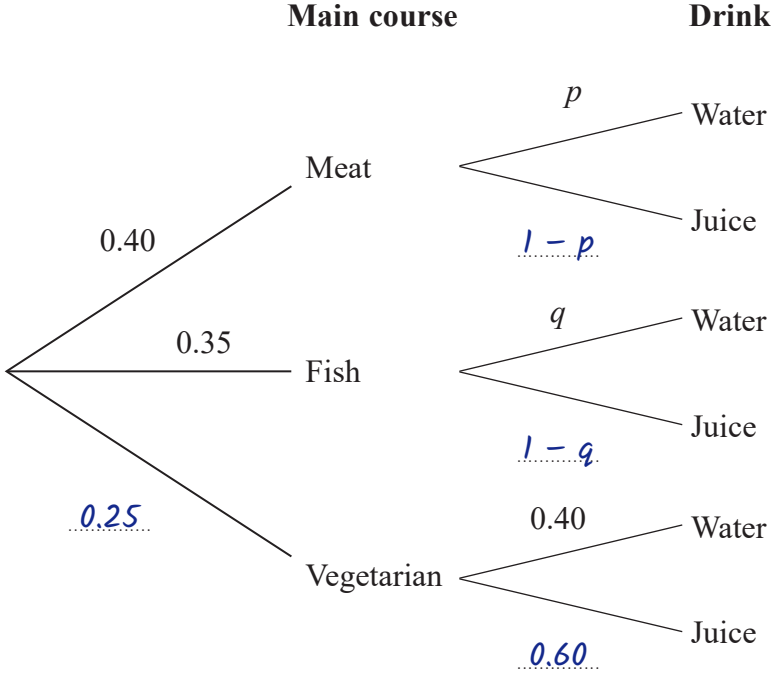
Q1

(Total 5 marks)

Leave blank

2. In 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)

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

(a) $P(V) = 1 - 0.40 - 0.35 = 0.25$

$P(J \mid M) = 1 - p, \quad P(J \mid F) = 1 - q, \quad P(J \mid V) = 1 - 0.40 = 0.60$

(b) $P(W) = 0.40p + 0.35q + 0.25 \times 0.40 = 0.4p + 0.35q + 0.1$

(c) V and W independent $\Rightarrow P(W) = P(W \mid V) = 0.4$

$0.4p + 0.35q + 0.1 = 0.4$

$0.4p + 0.35q = 0.3 \quad$ i.e. $\quad 8p + 7q = 6$

(d) V and W independent, so V and W′ = J are independent:

$P(J) = P(J \mid V) = 0.6$

$$P(F \mid J) = \frac{P(F \cap J)}{P(J)} = \frac{0.35(1-q)}{0.6} = \frac{7}{30}$$

$$1 - q = \frac{7 \times 0.6}{30 \times 0.35} = \frac{4.2}{10.5} = 0.4$$

$q = 0.6$

$8p + 7 \times 0.6 = 6 \Rightarrow 8p = 6 - 4.2 = 1.8$

$p = 0.225 \quad \left(= \frac{9}{40}\right)$

(e) $P(W) = 0.4 \times 0.225 + 0.35 \times 0.6 + 0.1 = 0.09 + 0.21 + 0.1 = 0.4$

$$P(M \mid W) = \frac{0.4 \times 0.225}{0.4} = \frac{0.09}{0.4} = 0.225$$

$$P(F \mid W) = \frac{0.35 \times 0.6}{0.4} = \frac{0.21}{0.4} = \frac{21}{40} = 0.525$$

$P(V \mid W) = P(V) = 0.25 \quad$ (independent)

$0.525 > 0.25 > 0.225$, so $P(F \mid W)$ is the largest (it is > 0.5).

Students choosing water are most likely to choose fish,

so the manager's claim is correct.

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 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)

(a) D = distance achieved, $D \sim N(3.8, 0.9^2)$

$P(D > 4.3) = P\left(Z > \frac{4.3 - 3.8}{0.9}\right) = P(Z > 0.5556)$

$= 1 - 0.7107... = 0.289257...$

Proportion receiving medals = 0.289 (3 s.f.)

(b) $P(D \geq d) = 0.8 \Rightarrow P(D < d) = 0.2 \Rightarrow z = -0.8416$

$\frac{d - 3.8}{0.9} = -0.8416$

$d = 3.8 - 0.8416 \times 0.9 = 3.04256...$

$d = 3.04$

(c) $P(D > g \mid D > 4.3) = \frac{P(D > g)}{P(D > 4.3)} = \frac{1}{3}$

$P(D > g) = \frac{1}{3} \times 0.289257... = 0.096419...$

$P(D < g) = 0.903580... \Rightarrow z = 1.302228...$

$\frac{g - 3.8}{0.9} = 1.302228...$

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

$g = 3.8 + 0.9 \times 1.302228... = 4.97200...$

$g = 4.97$

(d) P(medal winner is gold) $= \frac{1}{3}$, P(not gold) $= \frac{2}{3}$

P(no gold medal winners) $= \left(\frac{2}{3}\right)^3 = \frac{8}{27}$

P(at least one gold) $= 1 - \frac{8}{27} = \frac{19}{27}$

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 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

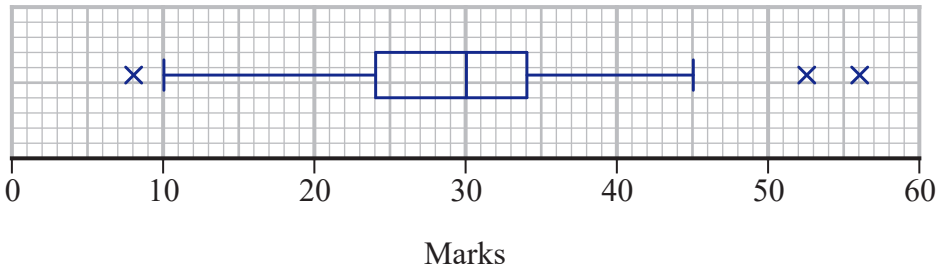
more than 1.5 × IQR below the lower quartile or
more than 1.5 × IQR above the upper quartile

(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.

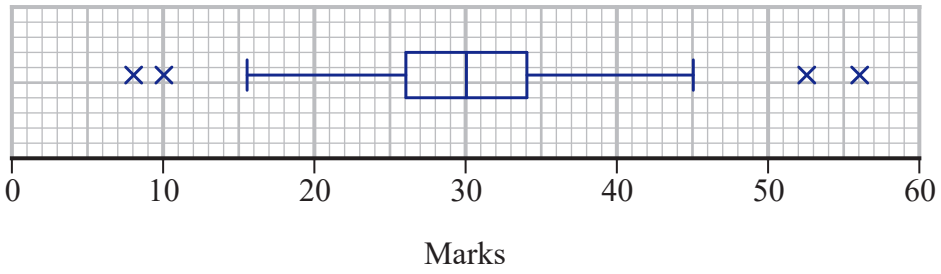
(2)

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)

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) $Q_3 = Q_1 + IQR = 24 + 10 = 34$

Lower limit $= 24 - 1.5 \times 10 = 24 - 15 = 9$

Upper limit $= 34 + 1.5 \times 10 = 34 + 15 = 49$

$8 < 9$, but $10 > 9$; $52.5 > 49$ and $56 > 49$, but $45 < 49$

Outliers are 8, 52.5 and 56

(b) Box plot drawn on the grid.

(c) $Q_2 - Q_1 = 30 - 24 = 6$, $Q_3 - Q_2 = 34 - 30 = 4$

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

(d) New $IQR = 34 - 26 = 8$

Lower limit $= 26 - 1.5 \times 8 = 14$, upper limit $= 34 + 1.5 \times 8 = 46$

Outliers: 8 and 10 (< 14) and 52.5 and 56 (> 46)

Lower whisker to 15.5, upper whisker to 45

(e) Q_1 has increased and the median is unchanged, so one mark is between 26 and 30 (inclusive).

Q_3 is unchanged, so the other mark is between 34 and 45 (inclusive) (both marks are less than 45).

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

Leave blank

Question 4 continued

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

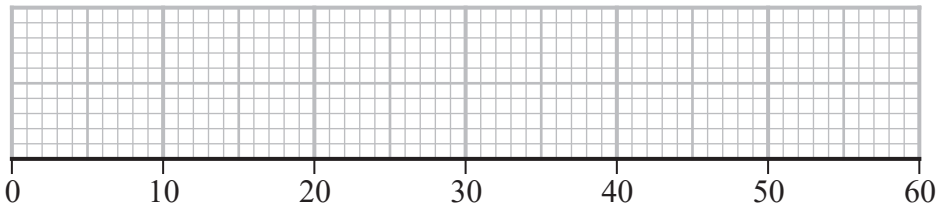
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Leave blank

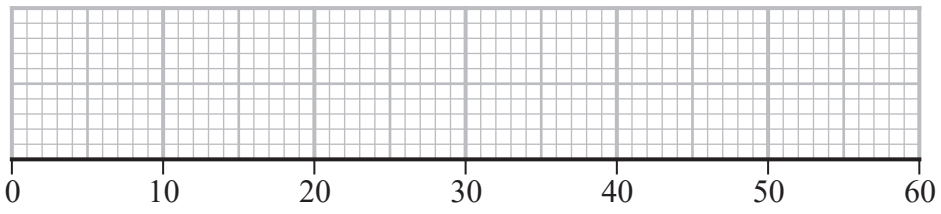
Question 4 continued

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)

(a) $800\ m^2 \Rightarrow x = 80$: $y = 6.066 + 0.136 \times 80 = 16.946$

Estimated annual rent = \$16 946

(b) $S_{yy} = 3434 - \frac{183^2}{10} = 3434 - 3348.9 = 85.1$

$S_{xx} = 84\,818 - \frac{900^2}{10} = 84\,818 - 81\,000 = 3818$

(c) $b = \frac{S_{xy}}{S_{xx}} \Rightarrow S_{xy} = 0.136 \times 3818 = 519.248$

$r = \frac{S_{xy}}{\sqrt{(S_{xx}\,S_{yy})}} = \frac{519.248}{\sqrt{(3818 \times 85.1)}}$

$= 0.910944\ldots = 0.911$

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

(d) Original $\bar{x} = \frac{900}{10} = 90$; new $\bar{x} = \frac{900 + 90}{11} = 90$

11th shop has $x = 90$, so its term $(x - \bar{x})(y - \bar{y}) = 0 \times (y - \bar{y}) = 0$

$\bar{x}$ is unchanged, so the other 10 terms $(x - \bar{x})(y - \bar{y})$ sum to the same value as before: $\Sigma(x - \bar{x})(y - \bar{y}) = \Sigma(x - \bar{x})y - \bar{y}\Sigma(x - \bar{x})$

and $\Sigma(x - \bar{x}) = 0$, so $\bar{y}$ changing does not matter.

$\therefore S_{xy}$ is the same for the 11 shops.

(e) S_{xx} is unchanged (new term $(90 - 90)^2 = 0$) and S_{xy} is unchanged, so $b = 0.136$ is the same.

New $\bar{y} = \frac{183 + 15}{11} = \frac{198}{11} = 18$ (old $\bar{y} = 18.3$)

$\bar{x}$ unchanged and $\bar{y}$ reduced by 0.3, so a is reduced by 0.3:

$a = 6.066 - 0.3 = 5.766$

$y = 5.766 + 0.136x$

(f) 3000 $m^2 \Rightarrow x = 300$, far outside the range of the data ($\bar{x} = 90$).

This is extrapolation, so the regression line is not suitable.

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 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)

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

(a) $E(A) = 1 \times 0.40 + 4 \times 0.20 + 5 \times 0.25 + 7 \times 0.15$

$= 0.4 + 0.8 + 1.25 + 1.05 = 3.5$

(b) $E(A^2) = 1^2 \times 0.40 + 4^2 \times 0.20 + 5^2 \times 0.25 + 7^2 \times 0.15$

$= 0.4 + 3.2 + 6.25 + 7.35 = 17.2$

$\text{Var}(A) = E(A^2) - [E(A)]^2 = 17.2 - 3.5^2 = 4.95$

(c) Discrete uniform distribution

(d) $E(B) = 3.5$ and the distribution is symmetrical about $E(B)$,

so $3.5 - 1 = k - 3.5$: $k = 6$

(e) Sam: $\mu = 4$, $\sigma = 3$: $P(X > 3.5) = P\left(z > \frac{3.5 - 4}{3}\right) = P\left(z > -\frac{1}{6}\right)$

Tim: $\sigma = 4$. Tim wins if $\frac{3.5 - A}{4} < -\frac{1}{6} \Rightarrow A > 3.5 + \frac{4}{6} = 4.166\ldots$

So Tim needs $A = 5$ or $A = 7$ ($A = 4$ gives $z = -0.125 > -0.1667$)

$P(\text{Tim wins}) = P(A = 5) + P(A = 7) = 0.25 + 0.15 = 0.4$

(f) Largest value needs the largest μ and the smallest σ:

$\mu = 7$ and $\sigma = 1$

$P(X > 3.5) = P\left(z > \frac{3.5 - 7}{1}\right) = P(z > -3.5) = P(z < 3.5)$

$= 0.999767\ldots = 0.9998$

(g) Need $A = 7$ and $B = 1$:

$P(A = 7) \times P(B = 1) = 0.15 \times 0.25 = 0.0375$

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 15 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