Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel International Advanced Level

Centre Number | Candidate Number

Wednesday 12 June 2019

Morning (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.
- 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

Pearson

June 2019
S1

Leave blank

Answer ALL questions. Write your answers in the spaces provided.

1. The heights, x metres, of 40 children were recorded by a teacher. The results are summarised as follows

$$\sum x = 58 \quad \sum x^2 = 84.829$$

(a) Find the mean and the variance of the heights of these 40 children.
(3)

The teacher decided that these statistics would be more useful in centimetres.

(b) Find

(i) the mean of these heights in centimetres,

(ii) the standard deviation of these heights in centimetres.
(2)

Two more children join the group. Their heights are 130 cm and 160 cm.

(c) (i) State, giving a reason, the mean height of the 42 children.

(ii) Without recalculating the standard deviation, state, giving a reason, whether the standard deviation of the heights of the 42 children will be greater than, less than or the same as the standard deviation of the heights of the group of 40 children.
(4)

(a) Mean:

$\frac{\Sigma x}{n} = \frac{58}{40} = 1.45$

Variance

$\frac{\Sigma x^2}{n} - (1.45)^2$

$= 0.018225$

$= \underline{0.0182}$

(b) 1m → 100cm.

$1.45 \times 100 = \underline{145}$

SD = $\sqrt{var}$

$= \sqrt{0.018225} \times 100$

$= \underline{13.5}$

c i) $\frac{130+160}{2} = \underline{\underline{145}}$

→ The mean remains the same as the mean of the 2 new children is the same as the mean of the 40 old children.

ii) SD will increase
→ Because the extra children are more than 1 s.d from the mean ∴ there is an increase in spread.

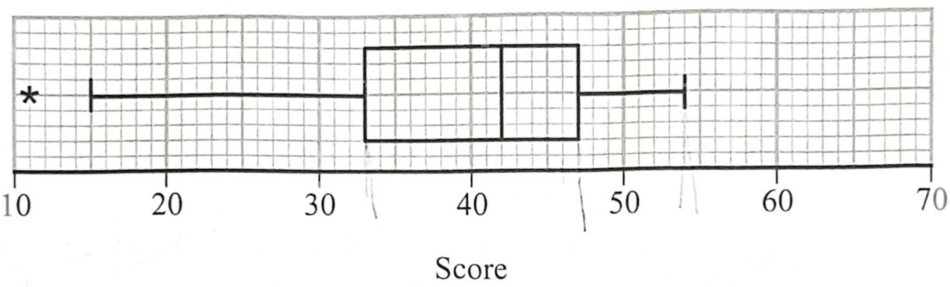
DO NOT WRITE IN THIS AREA

2. Chi wanted to summarise the scores of the 39 competitors in a village quiz. He started to produce the following stem and leaf diagram

Key: 2|5 is a score of 25

Score	
1	1 5 8 9
2	0 2 5 8 9
3	3 5 5 7 8 9 …

He did not complete the stem and leaf diagram but instead produced the following box plot.



Chi defined an outlier as a value that is

greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$

or less than $Q_1 - 1.5 \times (Q_3 - Q_1)$

(a) Find

(i) the interquartile range

(ii) the range.

(2)

(b) Describe, giving a reason, the skewness of the distribution of scores.

(2)

Albert and Beth asked for their scores to be checked.

Albert's score was changed from 25 to 37
Beth's score was changed from 54 to 60

(c) On the grid on page 5, draw an updated box plot.
Show clearly any calculations that you used.

(7)

Some of the competitors complained that the questions were biased towards the younger generation. The product moment correlation coefficient between the age of the competitors and their score in the quiz is –0.187

(d) State, giving a reason, whether or not the complaint is supported by this statistic.

(2)

Leave blank

Question 2 continued

(a) $IQR = Q_3 - Q_1$

$= 47 - 33 = \underline{\underline{14}}$

ii) Range $= 54 - 11$

$= \underline{43}$

b) negatively skewed as

$Q_3 - Q_2 < Q_2 - Q_1$

(c) $25 \rightarrow 37$

$\therefore$ new $Q_1 = 35$.

$54 \rightarrow 60$ $\therefore$ no change in Q_3.

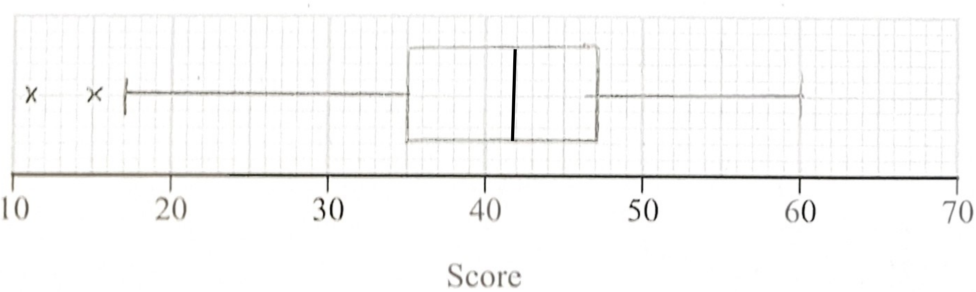
new $IQR = 47 - 35$

$= 12$

$35 - 1.5(12) = 17$

$47 + 1.5(12) = 65$

(d) Weak correlation $\therefore$ complaint not supported.



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

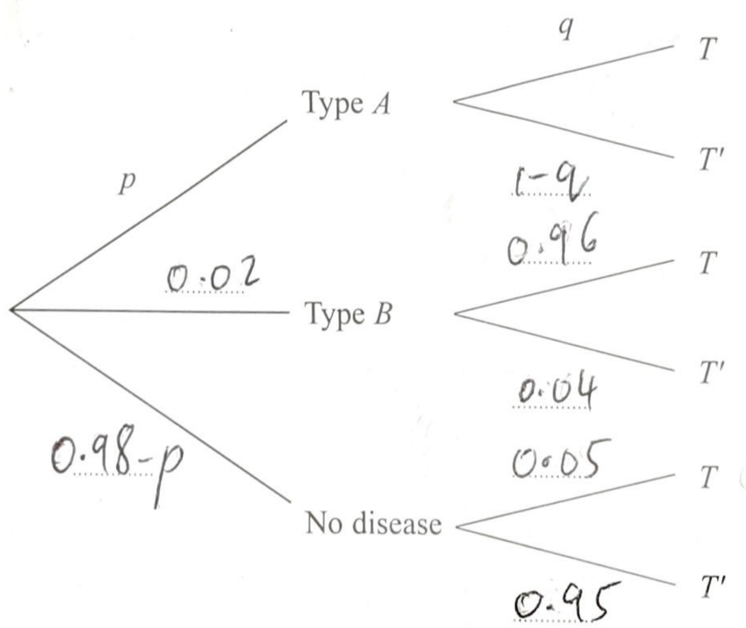
3. A certain disease occurs in a population in 2 mutually exclusive types.

It is difficult to diagnose people with type A of the disease and there is an unknown proportion p of the population with type A.
It is easier to diagnose people with type B of the disease and it is known that 2% of the population have type B.

A test has been developed to help diagnose whether or not a person has the disease. The event T represents a positive result on the test. After a large-scale trial of the test, the following information was obtained.

For a person with type B of the disease the probability of a positive test result is 0.96
For a person who does not have the disease the probability of a positive test result is 0.05
For a person with type A of the disease the probability of a positive test result is q

(a) Complete the tree diagram.



(2)

The probability of a randomly selected person having a positive test result is 0.169
For a person with a positive test result, the probability that they do not have the disease is $\frac{41}{169}$

(b) Find the value of p and the value of q.

(7)

A doctor is about to see a person who she knows does not have type B of the disease but does have a positive test result.

(c) (i) Find the probability that this person has type A of the disease.

(3)

(ii) State, giving a reason, whether or not the doctor will find the test useful.

(1)

Leave blank

Question 3 continued

(b) ① Positive test result.

$pq + (0.02 \times 0.96) + (0.98 - p)(0.05) = 0.169.$

$pq - 0.05p = 0.1008$

$p(q - 0.05) = 0.1008.$

② $p(\text{No disease} \mid \text{positive test})$

$= \dfrac{p(\text{no disease} \cap \text{+ve test})}{p(\text{+ve test})} = \dfrac{41}{169}.$

$(0.98 - p)(0.05) = \dfrac{41}{169} \times 0.169$

$0.049 - 0.05p = 0.041$

$0.05p = 8 \times 10^{-3}.$

$\underline{\underline{p = 0.16}}$

$q - 0.05 = \dfrac{0.1008}{0.16}$

$\underline{\underline{q = 0.68}}$

(i) $\dfrac{0.68 \times 0.16}{(0.68 \times 0.16) + (0.82 \times 0.05)}$

$\underline{\underline{= 0.726}}$

(ii) → Should be useful as the doctor knows there are high chances of patients having type A.

Turn over

Leave blank

4. The weights of packages delivered to Susie are normally distributed with a mean of 510 grams and a standard deviation of 45 grams.

(a) Find the probability that a randomly selected package delivered to Susie weighs less than 450 grams.
(3)

The heaviest 5% of packages delivered to Susie are delivered by Rav in his van, the others are delivered by Taruni on foot.

(b) Find the weight of the lightest package that Rav would deliver to Susie.
(3)

Susie randomly selects a package from those delivered by Taruni.

(c) Find the probability that this package weighs more than 450 grams.
(4)

On Tuesday there are 5 packages delivered to Susie.

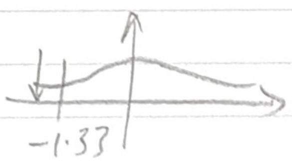
(d) Find the probability that 4 are delivered by Taruni and 1 is delivered by Rav.
(3)

(a) $(510, 45^2)$.

$p(x < 450)$.

$p\left(x < \frac{450-510}{45}\right)$.

$p(x < -1.33)$.



$1 - 0.9082$

$= \underline{0.0918}$

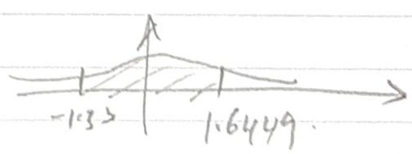
(b) $p(x > w) = 0.05$

$p\left(x > \frac{w-510}{45}\right) = 0.05$

$1.6449 = \frac{w - \cancel{550}\ 510}{45}$

$w = 584.0205$

(c)



$p(-1.33 < x < 1.64)$

$0.95 - 0.0918$

$= \frac{0.8582}{0.95}$ $= \underline{0.903}$

Leave blank

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 4 continued

(d) T T T T R
T T T R T.
T T R T T
T R T T T
R T T T T.

$(0.95)^4 \times 0.05 \times 5$

$= \underline{\underline{0.204}}$

Leave blank

5. The discrete random variable X represents the score when a biased spinner is spun. The probability distribution of X is given by

x	-2	-1	0	2	3
$\mathrm{P}(X = x)$	p	p	q	$\frac{1}{4}$	p

where p and q are probabilities.

(a) Find $\mathrm{E}(X)$.

(2)

Given that $\mathrm{Var}(X) = 2.5$

(b) find the value of p.

(5)

(c) Hence find the value of q.

(1)

Amar is invited to play a game with the spinner.
The spinner is spun once and X_1 is the score on the spinner.

If $X_1 > 0$ Amar wins the game.

If $X_1 = 0$ Amar loses the game.

If $X_1 < 0$ the spinner is spun again and X_2 is the score on this second spin and if $X_1 + X_2 > 0$ Amar wins the game, otherwise Amar loses the game.

(d) Find the probability that Amar wins the game.

(4)

Amar does not want to lose the game.
He says that because $\mathrm{E}(X) > 0$ he will play the game.

(e) State, giving a reason, whether or not you would agree with Amar.

(2)

(a) $E(X) = -2p - p + 0 + \frac{1}{2} + 3p$

$= 1/2$

(b) $E(X^2) - (1/2)^2 = 2.5$

$E(X^2) = \cancel{9/4} \cdot \frac{11}{4}$

$4p + p + 1 + 9p = \frac{11}{4}$

$14p = \frac{7}{4}$

$p = \frac{1}{8}$

Leave blank

Question 5 continued

(c) $\frac{3}{8} + \frac{1}{4} + q = 1$

$q = 3/8$.

(d) $p(X_1 > 0) = \frac{1}{4} + \frac{1}{8}$

when $X_1 < 0$ he spins again

-2, 3 $\quad 1/8 \times 1/8 = 1/64$
-1, 2 $\quad 1/8 \times 1/4 = 1/32$
-1, 3 $\quad 1/8 \times 1/8 = 1/64$

∴ ~~probabi~~ probability that he will win $= 7/16$.

(e) i don't agree as probability of winning (0.4375) is less than 0.5

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

6. *Ranpose* hospital offers services to a large number of clinics that refer patients to a range of hospitals.

The manager at *Ranpose* hospital took a random sample of 16 clinics and recorded

- the distance, x km, of the clinic from *Ranpose* hospital
- the percentage, y%, of the referrals from the clinic who attend *Ranpose* hospital.

The data are summarised as

$$\bar{x} = 8.1 \qquad \bar{y} = 20.5 \qquad \sum y^2 = 8266 \qquad S_{xx} = 368.16 \qquad S_{xy} = -630.9$$

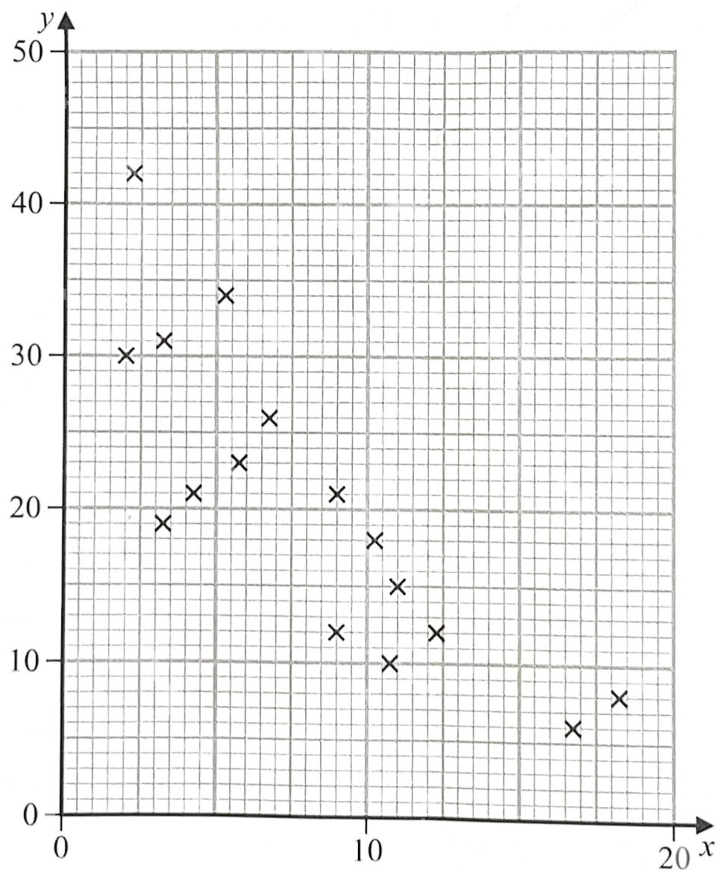
(a) Find the product moment correlation coefficient for these data.

(4)

(b) Give an interpretation of your correlation coefficient.

(1)

The manager at *Ranpose* hospital believes that there may be a linear relationship between the distance of a clinic from the hospital and the percentage of the referrals who attend the hospital. She drew the following scatter diagram for these data.



(c) State, giving a reason, whether or not these data support the manager's belief.

(1)

Question 6 continues on page 22

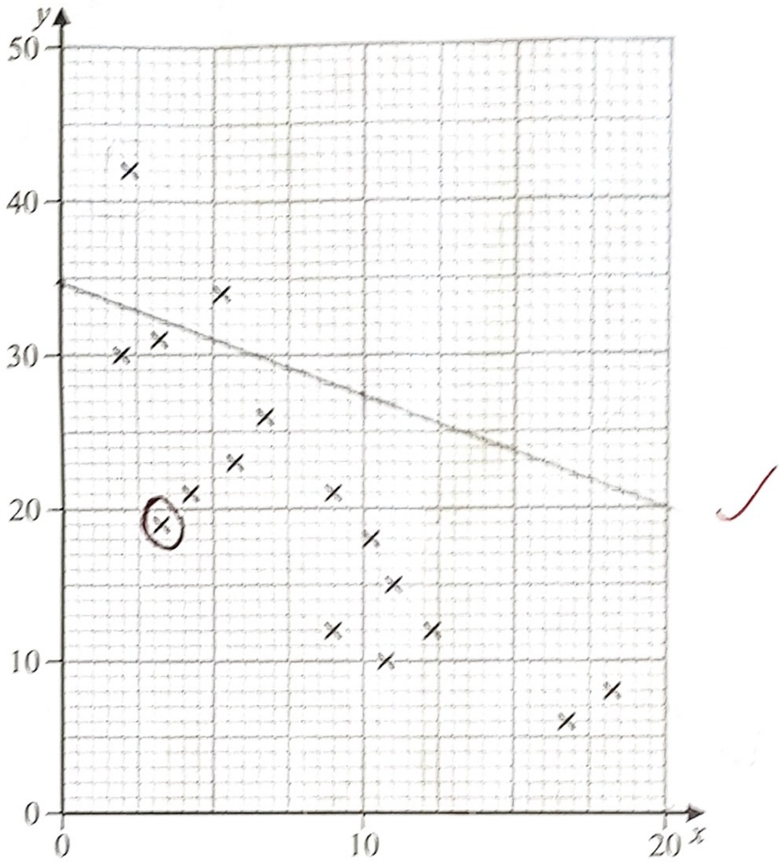
P 5 6 1 6 9 A 0 2 0 3 2

Question 6 continued

[The summary data and the scatter diagram are repeated below.]

The data are summarised as

$\bar{x} = 8.1 \qquad \bar{y} = 20.5 \qquad \sum y^2 = 8266 \qquad S_{xx} = 368.16 \qquad S_{xy} = -630.9$



(d) Find the equation of the regression line of y on x, giving your answer in the form $y = a + bx$

(4)

(e) Give an interpretation of the gradient of your regression line.

(1)

(f) Draw your regression line on the scatter diagram.

(1)

The manager believes that *Ranpose* hospital should be attracting an "above average" percentage of referrals from clinics that are less than 5 km from the hospital. She proposes to target one clinic with some extra publicity about the services *Ranpose* offers.

(g) On the scatter diagram circle the point representing the clinic she should target.

(1)

Leave blank

Question 6 continued

(a) $\frac{Sxy}{\sqrt{Sxx \times Syy}}$.

$Syy = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n}$

$\frac{\Sigma Y}{n} = 20.5$.

$\Sigma Y = 20.5 \times 16$
$= \underline{\underline{328}}$

$8266 - \frac{(328)^2}{16}$

$= \underline{\underline{1542}}$

$\frac{-630.9}{\sqrt{1542 \times 368.16}}$

$= \underline{\underline{-0.837}}$

(b) The further the clinic from Ranpose hospital, the lower the % of referals from that clinic who attend Ranpose hospital

(c) Points close to a straight line $\therefore$ does support the belief

d) $y = a + bx$.

$b = \frac{Sxy}{Sxx}$

$= \frac{-630.9}{368.16} = -1.71$

$a = \bar{y} - b\bar{x}$

$= 20.5 - (1.71)(8.1)$

$a = \underline{34.4}$

$y = 34.4 - 1.71x$.

(e) For every 1km of inc. of the clinic the % of referals who attend Ranpose hospital decrease by 1.71%.

Turn over