Pearson Edexcel
International
Advanced Level

Statistics S1

Advanced/Advanced Subsidiary

Wednesday 18 October 2017 – Morning
Time: 1 hour 30 minutes

Paper Reference
WST01/01

You must have:
Mathematical Formulae and Statistical Tables (Blue)

Total Marks

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. 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). Coloured pencils and highlighter pens must not be used.
- **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. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information

- 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.
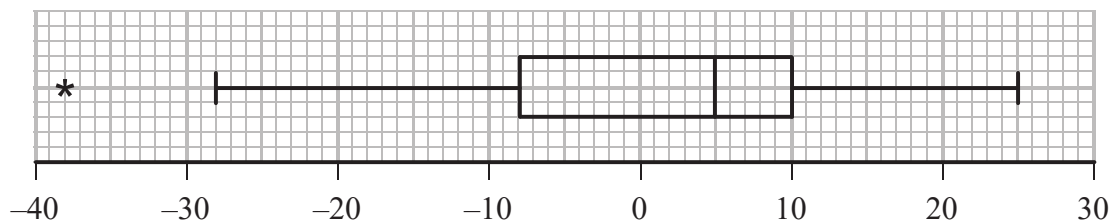- Check your answers if you have time at the end.

Turn over ▶

1. At the start of a course, an instructor asked a group of 80 apprentices to estimate the length of a piece of pipe. The error (true length – estimated length) was recorded in centimetres. The results are summarised in the box plot below.

(a) Find the range for these data.

(1)

(b) Find the interquartile range for these data.

(1)

One month later, the instructor asked the 80 apprentices to estimate the length of a different piece of pipe and recorded their errors. The results are summarised in the table below.

Error (e cm)	Number of apprentices
$-40 < e \leqslant -16$	2
$-16 < e \leqslant -8$	18
$-8 < e \leqslant 0$	33
$0 < e \leqslant 8$	14
$8 < e \leqslant 16$	10
$16 < e \leqslant 40$	3

(c) Use linear interpolation to estimate the median error for these data.

(2)

(d) Show that the upper quartile for these data, to the nearest centimetre, is 4.

(1)

For these data, the lower quartile is –8 and the five worst errors were –25, –21, 18, 23, 28

An outlier is a value that falls either
    more than $1.5 \times$ (interquartile range) above the upper quartile or
    more than $1.5 \times$ (interquartile range) below the lower quartile.

(e) (i) Show that there are only 2 outliers for these data.

(ii) Draw a box plot for these data on the grid on page 3.

(6)

(f) State, giving reasons, whether or not the apprentices' ability to estimate the length of a piece of pipe has improved over the first month of the course.

(3)

Question 1 continued

−40	−30	−20	−10	0	10	20	30

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

P 5 1 5 1 0 A 0 3 2 4

Question 1 continued

Question 1 continued

Only use this diagram if you need to redraw your plot.

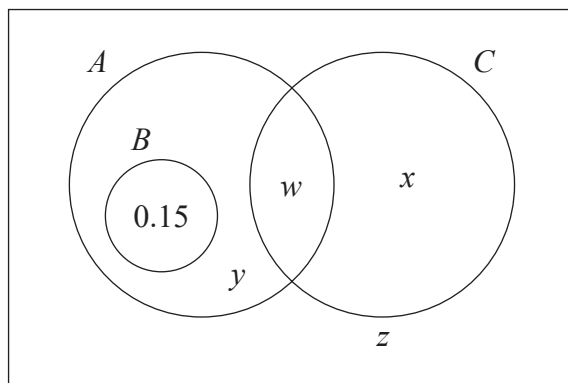|						
−40	−30	−20	−10	0	10	20	30
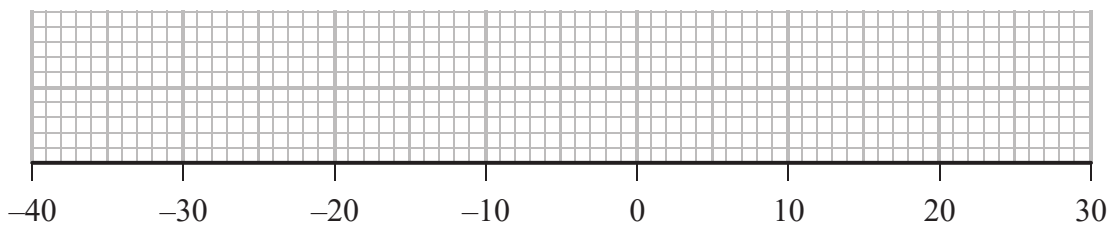
(Total 14 marks)

2. The Venn diagram, where w, x, y and z are probabilities, shows the probabilities of a group of students buying each of 3 magazines.

A represents the event that a student buys magazine A and $P(A) = 0.60$

B represents the event that a student buys magazine B and $P(B) = 0.15$

C represents the event that a student buys magazine C and $P(C) = 0.35$

(a) State which two of the three events A, B and C are mutually exclusive.

(1)

The events A and C are independent.

(b) Show that $w = 0.21$

(1)

(c) Find the value of x, the value of y and the value of z.

(4)

(d) Find the probability that a student selected at random buys only one of these magazines.

(1)

(e) Find the probability that a student selected at random buys magazine B or magazine C.

(1)

(f) Find $P(A \mid [B \cup C])$

(3)

P 5 1 5 1 0 A 0 6 2 4

P51510A0724

Leave blank

Question 2 continued

P51510A0824

Question 2 continued

(Total 11 marks)

Q2

P51510A0924

3. Hei and Tang are designing some pieces of art. They collected a large number of sticks. The random variable L represents the length of a stick in centimetres and has a normal distribution with mean μ and standard deviation σ.

They sorted the sticks into lengths and painted them.

They found that 60% of the sticks were longer than 45 cm and these were painted red, whilst 15% of the sticks were shorter than 35 cm and these were painted blue. The remaining sticks were painted yellow.

(a) Show that μ and σ satisfy

$$45 + 0.2533\sigma = \mu$$

(2)

(b) Find a second equation in μ and σ.

(2)

(c) Hence find the value of μ and the value of σ.

(3)

(d) Find   (i) $P(L > 35 \,|\, L < 45)$

           (ii) $P(L < 45 \,|\, L > 35)$

(3)

Hei created her piece of art using a random selection of blue and yellow sticks.

Tang created his piece of art using a random selection of red and yellow sticks.

Hei and Tang each used the same number of sticks to create their piece of art.

George is viewing Hei's and Tang's pieces of art. He finds a yellow stick on the floor that has fallen from one of these pieces.

(e) With reference to your answers to part (d), state, giving a reason, whether the stick is more likely to have fallen from Hei's or Tang's piece of art.

(2)

P 5 1 5 1 0 A 0 1 0 2 4

Question 3 continued

Question 3 continued

Q3

(Total 12 marks)

P 5 1 5 1 0 A 0 1 3 2 4

Turn over

4. The following incomplete tree diagram shows the relationships between the event A and the event B.

$\frac{2}{5}$ B

.... A

B'

....

$\frac{7}{10}$ B

A'

....

B'

....
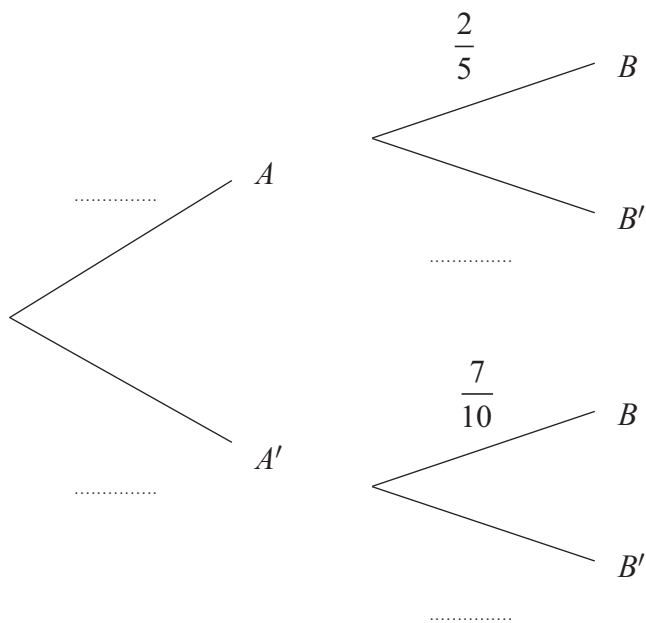
Given that $P(B) = \dfrac{9}{20}$

(a) find $P(A)$ and complete the tree diagram,

(6)

(b) find $P(A' \,|\, B')$.

(2)

Question 4 continued

(Total 8 marks)

Q4

5. A company wants to pay its employees according to their performance at work. Last year's performance score x and annual salary y, in thousands of dollars, were recorded for a random sample of 10 employees of the company.

The performance scores were

$$15 \quad 24 \quad 32 \quad 39 \quad 41 \quad 18 \quad 16 \quad 22 \quad 34 \quad 42$$

(You may use $\sum x^2 = 9011$)

(a) Find the mean and the variance of these performance scores.

(3)

The corresponding y values for these 10 employees are summarised by

$$\sum y = 306.1 \quad \text{and} \quad S_{yy} = 546.3$$

(b) Find the mean and the variance of these y values.

(2)

The regression line of y on x based on this sample is

$$y = 12.0 + 0.659x$$

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

(5)

(d) State, giving a reason, whether or not the value of the product moment correlation coefficient supports the use of a regression line to model the relationship between performance score and annual salary.

(1)

The company decides to use this regression model to determine future salaries.

(e) Find the proposed annual salary, in dollars, for an employee who has a performance score of 35

(2)

Question 5 continued

P51510A01724

Question 5 continued

Question 5 continued

(Total 13 marks)

Q5

6. The discrete random variable D with the following probability distribution represents the score when a 4-sided die is rolled.

d	1	2	3	4
$P(D = d)$	$\dfrac{1}{4}$	$\dfrac{1}{4}$	$\dfrac{1}{4}$	$\dfrac{1}{4}$

(a) Write down the name of this distribution.

(1)

The die is used to play a game and the random variable X represents the number of points scored. The die is rolled once and if $D = 2$, 3 or 4 then $X = D$. If $D = 1$ the die is rolled a second time and $X = 0$ if $D = 1$ again, otherwise X is the sum of the two scores on the die.

(b) Show that the probability of scoring 3 points in this game is $\dfrac{5}{16}$

(2)

(c) Find the probability of scoring 0 in this game.

(1)

The table below shows the probability distribution for the remaining values of X.

x	0	2	3	4	5
$P(X = x)$		$\dfrac{1}{4}$		$\dfrac{5}{16}$	$\dfrac{1}{16}$

(d) Find $E(X)$

(2)

(e) Find $Var(X)$

(3)

The discrete random variable R represents the number of times the die is rolled in the game.

(f) Write down the probability distribution of R.

(2)

The random variable $Y = 2R + 0.5$

(g) Show that $E(Y) = E(X)$

(3)

The game is played once.

(h) Find $P(X > Y)$

(3)

Question 6 continued

P 5 1 5 1 0 A 0 2 1 2 4

Question 6 continued

P51510A02224

Question 6 continued

Q6

(Total 17 marks)

TOTAL FOR PAPER: 75 MARKS

END

P51510A02424