Write your name here

Surname	Other names

Pearson Edexcel GCE

Centre Number	Candidate Number

Statistics S1

Advanced/Advanced Subsidiary

Wednesday 7 June 2017 – Morning
Time: 1 hour 30 minutes

Paper Reference
6683/01

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

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 ▶

P48947A

©2017 Pearson Education Ltd.

1/1/1/1/

P 4 8 9 4 7 A 0 1 2 4

Pearson

Leave blank

1. A clothes shop manager records the weekly sales figures, £ s, and the average weekly temperature, t °C, for 6 weeks during the summer. The sales figures were coded so that

$$w = \frac{s}{1000}$$

The data are summarised as follows

$$S_{ww} = 50 \quad \sum wt = 784 \quad \sum t^2 = 2435 \quad \sum t = 119 \quad \sum w = 42$$

(a) Find S_{wt} and S_{tt}

(3)

(b) Write down the value of S_{ss} and the value of S_{st}

(2)

(c) Find the product moment correlation coefficient between s and t.

(2)

The manager of the clothes shop believes that a linear regression model may be appropriate to describe these data.

(d) State, giving a reason, whether or not your value of the correlation coefficient supports the manager's belief.

(1)

(e) Find the equation of the regression line of w on t, giving your answer in the form $w = a + bt$

(3)

(f) Hence find the equation of the regression line of s on t, giving your answer in the form $s = c + dt$, where c and d are correct to 3 significant figures.

(2)

(g) Using your equation in part (f), interpret the effect of a 1°C increase in average weekly temperature on weekly sales during the summer.

(1)

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

Leave blank

Question 1 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 1 continued

Q1

(Total 14 marks)

Leave blank

2. An estate agent is studying the cost of office space in London. He takes a random sample of 90 offices and calculates the cost, £x per square foot. His results are given in the table below.

Cost (£x)	Frequency (f)	Midpoint (£y)
$20 \leqslant x < 40$	12	30
$40 \leqslant x < 45$	13	42.5
$45 \leqslant x < 50$	25	47.5
$50 \leqslant x < 60$	32	55
$60 \leqslant x < 80$	8	70

(You may use $\sum f y^2 = 226\,687.5$)

A histogram is drawn for these data and the bar representing $50 \leqslant x < 60$ is 2 cm wide and 8 cm high.

(a) Calculate the width and height of the bar representing $20 \leqslant x < 40$

(3)

(b) Use linear interpolation to estimate the median cost.

(2)

(c) Estimate the mean cost of office space for these data.

(2)

(d) Estimate the standard deviation for these data.

(2)

(e) Describe, giving a reason, the skewness.

(1)

Rika suggests that the cost of office space in London can be modelled by a normal distribution with mean £50 and standard deviation £10

(f) With reference to your answer to part (e), comment on Rika's suggestion.

(1)

(g) Use Rika's model to estimate the 80th percentile of the cost of office space in London.

(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

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

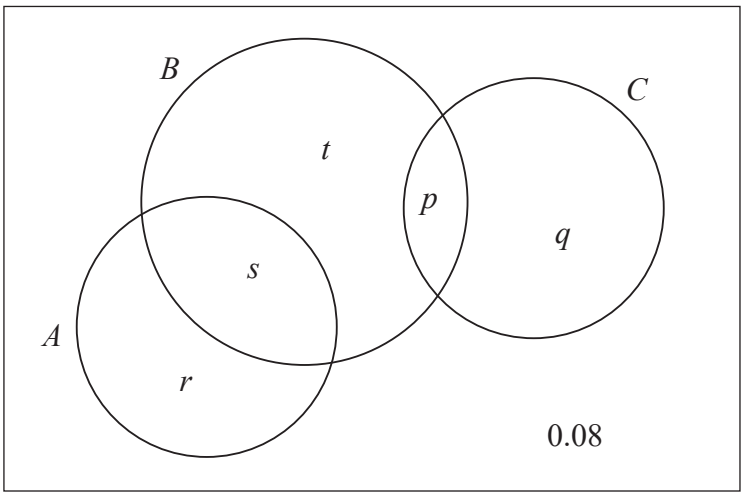
Question 2 continued

(Total 14 marks)

Q2

Leave blank

3. The Venn diagram shows three events A, B and C, where p, q, r, s and t are probabilities.



$P(A) = 0.5$, $P(B) = 0.6$ and $P(C) = 0.25$ and the events B and C are independent.

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

(2)

(b) Find the value of r.

(2)

(c) Hence write down the value of s and the value of t.

(2)

(d) State, giving a reason, whether or not the events A and B are independent.

(2)

(e) Find $P(B \mid A \cup C)$.

(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 3 continued

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 11 marks)

Q3

Leave blank

4. The discrete random variable X has probability distribution

x	–1	0	1	2
$P(X = x)$	a	b	b	c

The cumulative distribution function of X is given by

x	–1	0	1	2
$F(x)$	$\frac{1}{3}$	d	$\frac{5}{6}$	e

(a) Find the values of a, b, c, d and e.

(5)

(b) Write down the value of $P(X^2 = 1)$.

(1)

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

Q4

(Total 6 marks)

Leave blank

5. Yuto works in the quality control department of a large company. The time, T minutes, it takes Yuto to analyse a sample is normally distributed with mean 18 minutes and standard deviation 5 minutes.

(a) Find the probability that Yuto takes longer than 20 minutes to analyse the next sample.

(3)

The company has a large store of samples analysed by Yuto with the time taken for each analysis recorded. Serena is investigating the samples that took Yuto longer than 15 minutes to analyse.

She selects, at random, one of the samples that took Yuto longer than 15 minutes to analyse.

(b) Find the probability that this sample took Yuto more than 20 minutes to analyse.

(4)

Serena can identify, in advance, the samples that Yuto can analyse in under 15 minutes and in future she will assign these to someone else.

(c) Estimate the median time taken by Yuto to analyse samples in future.

(5)

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

Q5

(Total 12 marks)

Leave blank

6. The score, X, for a biased spinner is given by the probability distribution

x	0	3	6
$P(X = x)$	$\frac{1}{12}$	$\frac{2}{3}$	$\frac{1}{4}$

Find

(a) $E(X)$

(2)

(b) $Var(X)$

(3)

A biased coin has one face labelled 2 and the other face labelled 5
The score, Y, when the coin is spun has

$$P(Y = 5) = p \quad \text{and} \quad E(Y) = 3$$

(c) Form a linear equation in p and show that $p = \frac{1}{3}$

(3)

(d) Write down the probability distribution of Y.

(1)

Sam plays a game with the spinner and the coin.
Each is spun once and Sam calculates his score, S, as follows

if $X = 0$ then $S = Y^2$
if $X \neq 0$ then $S = XY$

(e) Show that $P(S = 30) = \frac{1}{12}$

(2)

(f) Find the probability distribution of S.

(3)

(g) Find $E(S)$.

(2)

Charlotte also plays the game with the spinner and the coin.
Each is spun once and Charlotte ignores the score on the coin and just uses X^2 as her score.
Sam and Charlotte each play the game a large number of times.

(h) State, giving a reason, which of Sam and Charlotte should achieve the higher total score.

(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

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 18 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