Write your name here

Surname	Other names

Pearson Edexcel GCE

Centre Number	Candidate Number

Statistics S4

Advanced/Advanced Subsidiary

Wednesday 28 June 2017 – Morning
Time: 1 hour 30 minutes

Paper Reference
6686/01

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 ▶

P49069A

©2017 Pearson Education Ltd.

1/1/1/

P 4 9 0 6 9 A 0 1 2 4

Pearson

Leave blank

1. The times taken by children to run 150m are normally distributed. The times taken, x seconds, by a random sample of 9 boys and an independent random sample of 6 girls are recorded. The following statistics are obtained.

	Number of children	**Sample mean $\bar{x}$**	$\sum x^2$
Boys	9	22.8	4693.60
Girls	6	29.5	5236.12

(a) Test, at the 10% level of significance, whether or not the variances of the two distributions are equal. State your hypotheses clearly.

(7)

The Headteacher claims that the mean time taken for the girls is more than 5 seconds greater than the mean time taken for the boys.

(b) Stating your hypotheses clearly, test the Headteacher's claim. Use a 1% level of significance and show your working clearly.

(7)

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

(Total 14 marks)

Q1

Leave blank

2. The number of accidents per year in *Daftstown* follows a Poisson distribution with mean λ. The value of λ has previously been 6 but Jonty claims that since the Council increased the speed limit, the value of λ has increased.

Jonty records the number of accidents in *Daftstown* in the first year after the speed limit was increased. He plans to test, at the 5% significance level, whether or not there is evidence of an increase in the mean number of accidents in *Daftstown* per year.

(a) Stating your hypotheses clearly, calculate the probability of a Type I error for this test.

(4)

Given that there were 9 accidents in the first year after the speed limit was increased,

(b) state, giving a reason, whether or not there is evidence to support Jonty's claim.

(2)

(c) Given that the value of λ has actually increased to 8, calculate the probability of drawing the conclusion, using this test, that the number of accidents per year in *Daftstown* has not increased.

(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 2 continued

Q2

(Total 8 marks)

Leave blank

3. The lengths, X mm, of the wings of adult blackbirds follow a normal distribution. A random sample of 5 adult blackbirds is taken and the lengths of the wings are measured. The results are summarised below

$$\sum x = 655 \text{ and } \sum x^2 = 85\,845$$

(a) Test, at the 10% level of significance, whether or not the mean length of an adult blackbird's wing is less than 135 mm. State your hypotheses clearly.

(7)

(b) Find the 90% confidence interval for the variance of the lengths of adult blackbirds' wings. Show your working clearly.

(4)

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. A coach believes that the average score in the final round of a golf tournament is more than one point below the average score in the first round. To test this belief, the scores of 8 randomly selected players are recorded. The results are given in the table below.

Player	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
First round	76	80	72	78	83	88	81	72
Final round	70	78	75	75	79	84	83	69

(a) (i) State why a paired *t*-test is suitable for use with these data.

(ii) State an assumption that needs to be made in order to carry out a paired *t*-test in this case.

(2)

(b) Test, at the 5% level of significance, whether or not there is evidence to support the coach's belief. Show your working clearly.

(8)

(c) Explain, in the context of the coach's belief, what a Type II error would be in this case.

(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

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

Q4

(Total 12 marks)

Leave blank

5. *Jamland* and *Goodjam* are two suppliers of jars of jam. The weights of the jars of jam produced by each supplier can be assumed to be normally distributed with unknown, but equal, variances. A random sample of 20 jars of jam is taken from those supplied by *Jamland*.

Based on this sample, the 95% confidence interval for the mean weight of a jar of *Jamland* jam, in grams, is

[492, 507]

A random sample of 10 jars of jam is selected from those supplied by *Goodjam*. The weight of each jar of *Goodjam* jam, y grams, is recorded. The results are summarised as follows

$$\bar{y} = 480 \qquad s_y^2 = 280$$

Find a 90% confidence interval for the value by which the mean weight of a jar of jam supplied by *Jamland* exceeds the mean weight of a jar of jam supplied by *Goodjam*.

(11)

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 11 marks)

Leave blank

6. The independent random variables X_1 and X_2 are each distributed B(n, p), where $n > 1$
An unbiased estimator for p is given by

$$\hat{p} = \frac{aX_1 + bX_2}{n}$$

where a and b are constants.

[You may assume that if X_1 and X_2 are independent then $E(X_1X_2) = E(X_1)E(X_2)$]

(a) Show that $a + b = 1$

(2)

(b) Show that $\text{Var}(\hat{p}) = \dfrac{(2a^2 - 2a + 1)p(1 - p)}{n}$

(4)

(c) Hence, justifying your answer, determine the value of a and the value of b for which $\hat{p}$ has minimum variance.

(5)

(d) (i) Show that $\hat{p}^2$ is a biased estimator for p^2

(ii) Show that the bias $\rightarrow 0$ as $n \rightarrow \infty$

(5)

(e) By considering $E[X_1(X_1 - 1)]$ find an unbiased estimator for p^2

(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 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 19 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