Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel GCE

Centre Number | Candidate Number

Tuesday 25 June 2019

Morning (Time: 1 hour 30 minutes) | Paper Reference **6686/01**

Statistics S4

Advanced/Advanced Subsidiary

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

Total Marks

Candidates may use any calculator allowed 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). 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 ▶

P55865A

1/1/1/1/e2/

P 5 5 8 6 5 A 0 1 2 4

Pearson

1. A factory worker has two alternative routes to work.

The times taken, in minutes, on 9 randomly chosen occasions that the factory worker used route 1 are summarised by

$$\sum x = 182 \qquad \sum x^2 = 4202$$

The times taken, in minutes, on 12 randomly chosen occasions that the factory worker used route 2 are summarised by

$$\sum y = 310 \qquad \sum y^2 = 8396$$

You may assume that the times are independent and normally distributed.

(a) Stating your hypotheses clearly, test, at the 10% level of significance, whether or not there is evidence that the variances of the times taken on route 1 and on route 2 are the same.

(6)

(b) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not there is evidence that there is a difference in the means of the times taken on the two routes. You must show your working clearly and state the critical value used.

(7)

(c) Explain why your result in part (a) enables you to carry out the test in part (b).

(1)

Question 1 continued

Question 1 continued

Question 1 continued

Q1

(Total 14 marks)

2. An athlete is investigating the times taken, to the nearest minute, for runners to complete a 10 km road race and a 10 km cross-country race. She takes a random sample of 10 runners and records their times taken to complete each race.

The results are shown below.

Runner	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Time taken (road race)	43	33	52	61	36	32	31	45	42	52
Time taken (cross-country race)	48	34	55	61	38	36	31	49	45	54

The athlete claims that the race times for the road race are quicker than the race times for the cross-country race.

(a) Test, at the 1% level of significance, whether or not the results support the athlete's claim. State your hypotheses and show your working clearly.

(8)

(b) State a necessary assumption for the test in part (a).

(1)

Question 2 continued

Q2

(Total 9 marks)

3. When a biased six-sided die is rolled, the probability that a six will be obtained is an unknown constant p. The die is rolled n times and the number, X, of sixes obtained is recorded. The die is then rolled a further $2n$ times and the number, Y, of sixes obtained is also recorded.

(a) Show that

$$T_1 = \frac{3Y - 2X}{4n}$$

and

$$T_2 = \frac{5X + 3Y}{11n}$$

are both unbiased estimators for p.

(4)

(b) Find the variance for each of these estimators.

(4)

(c) State, with a reason, which of the two estimators for p is the better.

(1)

Question 3 continued

Question 3 continued

Question 3 continued

Q3

(Total 9 marks)

4. The weights of apples are normally distributed. A farmer records the weights, in grams, of a random sample of 10 apples from his orchard.

Based on his sample, the 95% confidence interval for the variance of the weights of apples from his orchard is

[12.31, 86.70]

(a) Show that the sample standard deviation, s, is 5.1 grams to 1 decimal place.

(4)

Given that the mean weight of the sample of the 10 apples is 85 grams,

(b) calculate a 99% confidence interval for the mean weight of the apples in the farmer's orchard. Show your working clearly.

(4)

Question 4 continued

Question 4 continued

Question 4 continued

Q4

(Total 8 marks)

5. The weights of Great Dane dogs are normally distributed.

A random sample of 12 Great Dane dogs is taken and their weights, to the nearest kilogram, are recorded. The results are given below.

73 68 67 69 71 72 70 77 74 72 70 69

Stating your hypotheses and showing your working clearly, test at the 5% level of significance,

(a) whether or not the mean weight of Great Dane dogs is less than 72 kg,

(8)

(b) whether or not the standard deviation of the weights of Great Dane dogs is 6 kg.

(5)

Question 5 continued

Q5

(Total 13 marks)

6. A factory uses a machine to fill bags with sand. The weight of sand in a bag, measured in kg, is distributed normally with a standard deviation of 2.5 and an intended mean of 38

The manager of the factory is concerned that the machine is not set correctly and it does not put enough sand into the bags. He therefore takes a random sample of 20 bags and weighs them in order to verify that the machine has been set correctly.

(a) Write down the appropriate null and alternative hypotheses to test the manager's concern.

(1)

(b) Using a 5% level of significance, find the critical region for the mean weight of sand in a sample of 20 bags for this hypothesis test.

(4)

Given that the machine is filling the bags with a mean weight of sand of 37.5 kg,

(c) find the probability of a Type II error for this test.

(3)

The manager decides to take a larger sample of size n. The manager requires the probability of a Type II error to be less than 0.2 when the machine is filling the bags with a mean weight of sand of 37.5 kg.

(d) Calculate the minimum value of n.

(7)

Question 6 continued

Question 6 continued

Question 6 continued

Q6

(Total 15 marks)

7. A random variable, X, has a probability density function given by

$$f(x) = \begin{cases} kx & \lambda < x < \lambda + 10 \\ 0 & \text{otherwise} \end{cases}$$

where λ and k are positive constants.

Hypotheses are set up such that

$$H_0: \lambda = 0 \qquad H_1: \lambda > 0$$

The suggested test procedure, based on a single observation on X, is to reject H_0 in favour of H_1 only if the observation is greater than 9.5

(a) Determine the size of this test.

(3)

Given that $\lambda = 2$

(b) calculate the power of this test.

(4)

Question 7 continued

Question 7 continued

Q7

(Total 7 marks)

TOTAL FOR PAPER: 75 MARKS

END