Edexcel GCE

Statistics S4

Advanced/Advanced Subsidiary

Friday 27 January 2006 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Answer Book (AB16)	Nil
Graph Paper (ASG2)	
Mathematical Formulae and	
Statistics Tables (Lilac or Green)	

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard 48G.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S4), the paper reference (6686), your surname, other names and signature.
Check that you have the correct question paper.
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 for Candidates

A booklet 'Mathematical Formulae and Statistics Tables' is provided.
Full marks may be obtained for answers to ALL questions.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 7 questions in this question paper. The total mark for this paper is 75.
There are 8 pages in this question paper. Any blank pages are indicated.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You must show sufficient working to make your methods clear to the examiner. Answers without working may gain no credit.

N 2 2 3 4 0 A

Turn over

edexcel

1. A diabetic patient records her blood glucose readings in mmol/l at random times of day over several days. Her readings are given below.

    5.3   5.7   8.4   8.7   6.3   8.0   7.2

    Assuming that the blood glucose readings are normally distributed calculate

    (a) an unbiased estimate for the variance σ^2 of the blood glucose readings,

    (2)

    (b) a 90% confidence interval for the variance σ^2 of blood glucose readings.

    (5)

    (c) State whether or not the confidence interval supports the assertion that $\sigma = 0.9$.

    Give a reason for your answer.

    (1)

    (Total 8 marks)

2. (a) Define

    (i) a Type I error,

    (1)

    (ii) a Type II error.

    (1)

    A manufacturer sells socks in boxes of 50.

    The mean number of faulty socks per box is 7.5. In order to reduce the number of faulty socks a new machine is tried. A box of socks made on the new machine was tested and the number of faulty socks was 2.

    (b) (i) Assuming that the number of faulty socks per box follows a binomial distribution derive a critical region needed to test whether or not there is evidence that the new machine has reduced the mean number of faulty socks per box. Use a 5% significance level.

    (2)

    (ii) Stating your hypotheses clearly, carry out the test in part (i).

    (3)

    (c) Find the probability of the Type I error for this test.

    (2)

    (d) Given that the true mean number of faulty socks per box on the new machine is 5, calculate the probability of a Type II error for this test.

    (3)

    (e) Explain what would have been the effect of changing the significance level for the test in part (b) to 2½%.

    (1)

    (Total 13 marks)

3. A population has mean μ and variance σ^2.
A random sample of size 3 is to be taken from this population and $\overline{X}$ denotes its sample mean.
A second random sample of size 4 is to be taken from this population and $\overline{Y}$ denotes its sample mean.

(a) Show that unbiased estimators for μ are given by

   (i) $\hat{\mu}_1 = \dfrac{1}{3}\overline{X} + \dfrac{2}{3}\overline{Y}$,

   (2)

   (ii) $\hat{\mu}_2 = \dfrac{5\overline{X} + 4\overline{Y}}{9}$.

   (2)

(b) Calculate Var $(\hat{\mu}_1)$

   (2)

(c) Given that Var $(\hat{\mu}_2) = \dfrac{37}{243}\sigma^2$, state, giving a reason, which of these two estimators should be used.

   (1)

   (Total 7 marks)

---

4. The number of accidents that occur at a crossroads has a mean of 3 per month. In order to improve the flow of traffic the priority given to traffic is changed. Colin believes that since the change in priority the number of accidents has increased. He tests his belief by recording the number of accidents x in the month following the change. Colin sets up the hypotheses H_0: $\lambda = 3$ and H_1: $\lambda > 3$, where λ is the mean number of accidents per month, and rejects the null hypothesis if $x > 4$.

(a) Find the size of the test.

   (3)

The table gives the values of the power function of the test to two decimal places.

λ	4	5	6	7
Power	r	0.56	s	0.83

(b) Calculate the value of r and the value of s.

   (2)

(c) Comment on the suitability of the test when $\lambda = 4$.

   (1)

   (Total 6 marks)

5. Seven pipes of equal length are selected at random. Each pipe is cut in half. One piece of each pipe is coated with protective paint and the other is left uncoated. All of the pieces of pipe are buried to the same depth in various soils for 6 months.

The table gives the percentage area of the pieces of pipe in the various soils that are subject to corrosion.

Soil	A	B	C	D	E	F	G
% Corrosion coated pipe	39	40	43	32	42	33	36
% Corrosion uncoated pipe	41	36	61	48	42	48	45

(a) Stating your hypotheses clearly and using a 5% significance level, carry out a paired t-test to assess whether or not there is a difference between the mean percentage of corrosion on the coated pipes and the mean percentage of corrosion on the uncoated pipes.

(9)

(b) (i) State an assumption that has been made in order to carry out this test.

(1)

(ii) Comment on the validity of this assumption.

(1)

(c) State what difference would be made to the conclusion in part (a) if the test had been to determine whether or not the percentage of corrosion on the uncoated pipes was higher than the mean percentage of corrosion on the coated pipes. Justify your answer.

(2)

(Total 13 marks)

6. A tree is cut down and sawn into pieces. Half of the pieces are stored outside and half of the pieces are stored inside. After a year, a random sample of pieces is taken from each location and the hardness is measured. The hardness x units are summarised in the following table.

	Number of pieces sampled	Σx	Σx^2
Stored outside	20	2340	274050
Stored inside	37	4884	645282

(a) Show that unbiased estimates for the variance of the values of hardness for wood stored outside and for the wood stored inside are 14.2 and 16.5, to 1 decimal place, respectively.

(2)

The hardness of wood stored outside and the hardness of wood stored inside can be assumed to be normally distributed with equal variances.

(b) Calculate 95% confidence limits for the difference in mean hardness between the wood that was stored outside and the wood that was stored inside.

(8)

(c) Using your answer to part (b), comment on the means of the hardness of wood stored outside and inside. Give a reason for your answer.

(2)

(Total 12 marks)

7. A psychologist gives a test to students from two different schools, A and B.
   A group of 9 students is randomly selected from school A and given instructions on how to do the test.
   A group of 7 students is randomly selected from school B and given the test without the instructions.

   The table shows the time taken, to the nearest second, to complete the test by the two groups.

A	11	12	12	13	14	15	16	17	17
B	8	10	11	13	13	14	14		

   Stating your hypotheses clearly,

   (a) test at the 10% significance level, whether or not the variance of the times taken to complete the test by students from school A is the same as the variance of the times taken to complete the test by students from school B. (You may assume that times taken for each school are normally distributed.)

   (7)

   (b) test at the 5% significance level, whether or not the mean time taken to complete the test by students from school A is greater than the mean time taken to complete the test by students from school B.

   (7)

   (c) Why does the result to part (a) enable you to carry out the test in part (b)?

   (1)

   (d) Give one factor that has not been taken into account in your analysis.

   (1)

   (Total 16 marks)

   TOTAL FOR PAPER: 75 MARKS

   END