Paper Reference(s)

6686/01

Edexcel GCE

Statistics S4

Advanced Subsidiary

Thursday 16 June 2005 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination	**Items included with question papers**
Answer Book (AB16)	Nil
Graph Paper (ASG2)	
Mathematical Formulae (Lilac)	

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 the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 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 name and signature.
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 Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
This paper has seven questions.
The total mark for this paper is 75.

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.

N21147A

1. The random variable X has a χ^2-distribution with 9 degrees of freedom.

(*a*) Find $P(2.088 < X < 19.023)$.

(3)

The random variable Y follows an F-distribution with 12 and 5 degrees of freedom.

(*b*) Find the upper and lower 5% critical values for Y.

(3)

(Total 6 marks)

---

2. The standard deviation of the length of a random sample of 8 fence posts produced by a timber yard was 8 mm. A second timber yard produced a random sample of 13 fence posts with a standard deviation of 14 mm.

(*a*) Test, at the 10% significance level, whether or not there is evidence that the lengths of fence posts produced by these timber yards differ in variability. State your hypotheses clearly.

(5)

(*b*) State an assumption you have made in order to carry out the test in part (*a*).

(1)

(Total 6 marks)

---

3. A machine is set to fill bags with flour such that the mean weight is 1010 grams.

To check that the machine is working properly, a random sample of 8 bags is selected. The weight of flour, in grams, in each bag is as follows.

1010	1015	1005	1000	998	1008	1012	1007

Carry out a suitable test, at the 5% significance level, to test whether or not the mean weight of flour in the bags is less than 1010 grams. (You may assume that the weight of flour delivered by the machine is normally distributed.)

(Total 8 marks)

---

4. A farmer set up a trial to assess the effect of two different diets on the increase in the weight of his lambs. He randomly selected 20 lambs. Ten of the lambs were given diet *A* and the other 10 lambs were given diet *B*. The gain in weight, in kg, of each lamb over the period of the trial was recorded.

(*a*) State why a paired *t*-test is not suitable for use with these data.

(1)

(*b*) Suggest an alternative method for selecting the sample which would make the use of a paired *t*-test valid.

(1)

(*c*) Suggest two other factors that the farmer might consider when selecting the sample.

(2)

The following paired data were collected.

Diet *A*	5	6	7	4.6	6.1	5.7	6.2	7.4	5	3
Diet *B*	7	7.2	8	6.4	5.1	7.9	8.2	6.2	6.1	5.8

(*d*) Using a paired *t*-test, at the 5% significance level, test whether or not there is evidence of a difference in the weight gained by the lambs using diet *A* compared with those using diet *B*.

(8)

(*e*) State, giving a reason, which diet you would recommend the farmer to use for his lambs.

(1)

(Total 13 marks)

5. Define

(a) a Type I error,

(1)

(b) the size of a test.

(1)

Jane claims that she can read Alan's mind. To test this claim Alan randomly chooses a card with one of 4 symbols on it. He then concentrates on the symbol. Jane then attempts to read Alan's mind by stating what symbol she thinks is on the card. The experiment is carried out 8 times and the number of times X that Jane is correct is recorded.

The probability of Jane stating the correct symbol is denoted by p.

To test the hypothesis H_0: $p = 0.25$ against H_1: $p > 0.25$, a critical region of $X > 6$ is used.

(c) Find the size of this test.

(3)

(d) Show that the power function of this test is $8p^7 - 7p^8$.

(3)

Given that $p = 0.3$, calculate

(e) the power of this test,

(1)

(f) the probability of a Type II error.

(2)

(g) Suggest two ways in which you might reduce the probability of a Type II error.

(2)

(Total 12 marks)

6. Brickland and Goodbrick are two manufacturers of bricks. The lengths of the bricks produced by each manufacturer can be assumed to be normally distributed. A random sample of 20 bricks is taken from Brickland and the length, x mm, of each brick is recorded. The mean of this sample is 207.1 mm and the variance is 3.2 mm^2.

(*a*) Calculate the 98% confidence interval for the mean length of brick from Brickland.

(4)

A random sample of 10 bricks is selected from those manufactured by Goodbrick. The length of each brick, y mm, is recorded. The results are summarised as follows.

$$\sum y = 2046.2 \qquad \sum y^2 = 418\,785.4$$

The variances of the length of brick for each manufacturer are assumed to be the same.

(*b*) Find a 90% confidence interval for the value by which the mean length of brick made by Brickland exceeds the mean length of brick made by Goodbrick.

(8)

(Total 12 marks)

7. A bag contains marbles of which an unknown proportion p is red. A random sample of n marbles is drawn, with replacement, from the bag. The number X of red marbles drawn is noted.

A second random sample of m marbles is drawn, with replacement. The number Y of red marbles drawn is noted.

Given that $p_1 = \dfrac{aX}{n} + \dfrac{bY}{m}$ is an unbiased estimator of p,

(*a*) show that $a + b = 1$.

(4)

Given that $p_2 = \dfrac{(X+Y)}{n+m}$,

(*b*) show that p_2 is an unbiased estimator for p.

(3)

(*c*) Show that the variance of p_1 is $p(1-p)\left(\dfrac{a^2}{n} + \dfrac{b^2}{m}\right)$.

(3)

(*d*) Find the variance of p_2.

(3)

(*e*) Given that $a = 0.4$, $m = 10$ and $n = 20$, explain which estimator p_1 or p_2 you should use.

(4)

(Total 17 marks)

TOTAL FOR PAPER: 75 MARKS

END