

Paper Reference(s)

6686

Edexcel GCE

Statistics S4

Advanced/Advanced Subsidiary

Tuesday 10 June 2003 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination

Answer Book (AB16)
Graph Paper (ASG2)
Mathematical Formulae (Lilac)

Items included with question papers

Nil

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.

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.

1. A beach is divided into two areas A and B . A random sample of pebbles is taken from each of the two areas and the length of each pebble is measured. A sample of size 26 is taken from area A and the unbiased estimate for the population variance is $s_A^2 = 0.495 \text{ mm}^2$. A sample of size 25 is taken from area B and the unbiased estimate for the population variance is $s_B^2 = 1.04 \text{ mm}^2$.
- (a) Stating your hypotheses clearly test, at the 10% significance level, whether or not there is a difference in variability of pebble length between area A and area B . (5)
- (b) State the assumption you have made about the populations of pebble lengths in order to carry out the test. (1)
-

2. A random sample of 10 mustard plants had the following heights, in mm, after 4 days growth.

5.0, 4.5, 4.8, 5.2, 4.3, 5.1, 5.2, 4.9, 5.1, 5.0

Those grown previously had a mean height of 5.1 mm after 4 days. Using a 2.5% significance level, test whether or not the mean height of these plants is less than that of those grown previously.

(You may assume that the height of mustard plants after 4 days follows a normal distribution.)

(9)

3. A train company claims that the probability p of one of its trains arriving late is 10%. A regular traveller on the company's trains believes that the probability is greater than 10% and decides to test this by randomly selecting 12 trains and recording the number X of trains that were late. The traveller sets up the hypotheses $H_0: p = 0.1$ and $H_1: p > 0.1$ and accepts the null hypothesis if $x \leq 2$.

- (a) Find the size of the test. (1)

- (b) Show that the power function of the test is

$$1 - (1 - p)^{10} (1 + 10p + 55p^2).$$

(4)

- (c) Calculate the power of the test when

(i) $p = 0.2$,

(ii) $p = 0.6$.

(3)

- (d) Comment on your results from part (c).

(1)

4. A random sample of 15 tomatoes is taken and the weight x grams of each tomato is found. The results are summarised by $\sum x = 208$ and $\sum x^2 = 2962$.

(a) Assuming that the weights of the tomatoes are normally distributed, calculate the 90% confidence interval for the variance σ^2 of the weights of the tomatoes. (7)

(b) State with a reason whether or not the confidence interval supports the assertion $\sigma^2 = 3$. (2)

5. (a) Define

(i) a Type I error,

(ii) a Type II error. (2)

A small aviary, that leaves the eggs with the parent birds, rears chicks at an average rate of 5 per year. In order to increase the number of chicks reared per year it is decided to remove the eggs from the aviary as soon as they are laid and put them in an incubator. At the end of the first year of using an incubator 7 chicks had been successfully reared.

(b) Assuming that the number of chicks reared per year follows a Poisson distribution test, at the 5% significance level, whether or not there is evidence of an increase in the number of chicks reared per year. State your hypotheses clearly. (4)

(c) Calculate the probability of the Type I error for this test. (3)

(d) Given that the true average number of chicks reared per year when the eggs are hatched in an incubator is 8, calculate the probability of a Type II error. (2)

6. A random sample of three independent variables X_1, X_2 and X_3 is taken from a distribution with mean μ and variance σ^2 .

(a) Show that $\frac{2}{3}X_1 - \frac{1}{2}X_2 + \frac{5}{6}X_3$ is an unbiased estimator for μ . (3)

An unbiased estimator for μ is given by $\hat{\mu} = aX_1 + bX_2$ where a and b are constants.

(b) Show that $\text{Var}(\hat{\mu}) = (2a^2 - 2a + 1)\sigma^2$. (6)

(c) Hence determine the value of a and the value of b for which $\hat{\mu}$ has minimum variance. (5)

7. Two methods of extracting juice from an orange are to be compared. Eight oranges are halved. One half of each orange is chosen at random and allocated to Method A and the other half is allocated to Method B . The amounts of juice extracted, in ml, are given in the table.

	Orange							
	1	2	3	4	5	6	7	8
Method A	29	30	26	25	26	22	23	28
Method B	27	25	28	24	23	26	22	25

One statistician suggests performing a two-sample t -test to investigate whether or not there is a difference between the mean amounts of juice extracted by the two methods.

(a) Stating your hypotheses clearly and using a 5% significance level, carry out this test.
(You may assume $\bar{x}_A = 26.125$, $s_A^2 = 7.84$, $\bar{x}_B = 25$, $s_B^2 = 4$ and $\sigma_A^2 = \sigma_B^2$) (7)

Another statistician suggests analysing these data using a paired t -test.

(b) Using a 5% significance level, carry out this test. (9)

(c) State which of these two tests you consider to be more appropriate. Give a reason for your choice. (1)

END