

Paper Reference(s)

6686/01

Edexcel GCE

Statistics S4

Advanced Level

Wednesday 21 June 2006 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination

Mathematical Formulae (Green)

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 6 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.

1. Historical records from a large colony of squirrels show that the weight of squirrels is normally distributed with a mean of 1012 g. Following a change in the diet of squirrels, a biologist is interested in whether or not the mean weight has changed.

A random sample of 14 squirrels is weighed and their weights x , in grams, recorded. The results are summarised as follows:

$$\Sigma x = 13\,700, \quad \Sigma x^2 = 13\,448\,750.$$

Stating your hypotheses clearly test, at the 5% level of significance, whether or not there has been a change in the mean weight of the squirrels.

(7)

2. The weights, in grams, of apples are assumed to follow a normal distribution.

The weights of apples sold by a supermarket have variance σ_s^2 . A random sample of 4 apples from the supermarket had weights

$$114, 100, 119, 123.$$

- (a) Find a 95% confidence interval for σ_s^2 .

(7)

The weights of apples sold on a market stall have variance σ_M^2 . A second random sample of 7 apples was taken from the market stall. The sample variance s_M^2 of the apples was 318.8.

- (b) Stating your hypotheses clearly test, at the 1% level of significance, whether or not there is evidence that $\sigma_M^2 > \sigma_s^2$.

(5)

3. As part of an investigation into the effectiveness of solar heating, a pair of houses was identified where the mean weekly fuel consumption was the same. One of the houses was then fitted with solar heating and the other was not. Following the fitting of the solar heating, a random sample of 9 weeks was taken and the table below shows the weekly fuel consumption for each house.

Week	1	2	3	4	5	6	7	8	9
Without solar heating	19	19	18	14	6	7	5	31	43
With solar heating	13	22	11	16	14	1	0	20	38

Units of fuel used per week

- (a) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not there is evidence that the solar heating reduces the mean weekly fuel consumption. **(8)**

- (b) State an assumption about weekly fuel consumption that is required to carry out this test. **(1)**

4. Two machines A and B produce the same type of component in a factory. The factory manager wishes to know whether the lengths, x cm, of the components produced by the two machines have the same mean. The manager took a random sample of components from each machine and the results are summarised in the table below.

	Sample size	Mean \bar{x}	Standard deviation s
Machine A	9	4.83	0.721
Machine B	10	4.85	0.572

The lengths of components produced by the machines can be assumed to follow normal distributions.

- (a) Use a two tail test to show, at the 10% significance level, that the variances of the lengths of components produced by each machine can be assumed to be equal. **(4)**

- (b) Showing your working clearly, find a 95% confidence interval for $\mu_B - \mu_A$, where μ_A and μ_B are the mean lengths of the populations of components produced by machine A and machine B respectively. **(7)**

There are serious consequences for the production at the factory if the difference in mean lengths of the components produced by the two machines is more than 0.7 cm.

- (c) State, giving your reason, whether or not the factory manager should be concerned. **(2)**

5. Rolls of cloth delivered to a factory contain defects at an average rate of λ per metre. A quality assurance manager selects a random sample of 15 metres of cloth from each delivery to test whether or not there is evidence that $\lambda > 0.3$. The criterion that the manager uses for rejecting the hypothesis that $\lambda = 0.3$ is that there are 9 or more defects in the sample.

(a) Find the size of the test.

(2)

Table 1 gives some values, to 2 decimal places, of the power function of this test.

λ	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Power	0.15	0.34	r	0.72	0.85	0.92	0.96

Table 1

(b) Find the value of r .

(2)

The manager would like to design a test, of whether or not $\lambda > 0.3$, that uses a smaller length of cloth. He chooses a length of 10 m and requires the probability of a type I error to be less than 10%.

(c) Find the criterion to reject the hypothesis that $\lambda = 0.3$ which makes the test as powerful as possible.

(2)

(d) Hence state the size of this second test.

(1)

Table 2 gives some values, to 2 decimal places, of the power function for the test in part (c).

λ	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Power	0.21	0.38	0.55	0.70	s	0.88	0.93

Table 2

(e) Find the value of s .

(2)

(f) Using the same axes, on graph paper draw the graphs of the power functions of these two tests.

(4)

(g) (i) State the value of λ where the graphs cross.

(ii) Explain the significance of λ being greater than this value.

(2)

The cost of wrongly rejecting a delivery of cloth with $\lambda = 0.3$ is low. Deliveries of cloth with $\lambda > 0.7$ are unusual.

(h) Suggest, giving your reasons, which the test manager should adopt.

(2)

6.

Figure 1

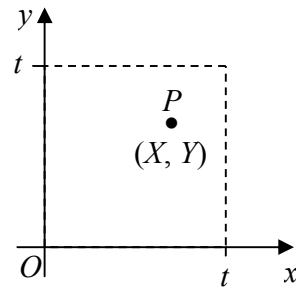


Figure 1 shows a square of side t and area t^2 which lies in the first quadrant with one vertex at the origin. A point P with coordinates (X, Y) is selected at random inside the square and the coordinates are used to estimate t^2 . It is assumed that X and Y are independent random variables each having a continuous uniform distribution over the interval $[0, t]$.

[You may assume that $E(X^n Y^n) = E(X^n)E(Y^n)$, where n is a positive integer.]

(a) Use integration to show that $E(X^n) = \frac{t^n}{n+1}$. (3)

The random variable $S = kXY$, where k is a constant, is an unbiased estimator for t^2 .

(b) Find the value of k . (3)

(c) Show that $\text{Var } S = \frac{7t^4}{9}$. (3)

The random variable $U = q(X^2 + Y^2)$, where q is a constant, is also an unbiased estimator for t^2 .

(d) Show that the value of $q = \frac{3}{2}$. (3)

(e) Find $\text{Var } U$. (3)

(f) State, giving a reason, which of S and U is the better estimator of t^2 . (1)

The point $(2, 3)$ is selected from inside the square.

(g) Use the estimator chosen in part (f) to find an estimate for the area of the square. (1)

TOTAL FOR PAPER: 75 MARKS

END