Paper Reference(s)

6686

Edexcel GCE

Statistics S4

Advanced/Advanced Subsidiary
Thursday 30 May 2002 – Morning
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. Pages 6, 7 and 8 are blank.

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. The random variable X has an F distribution with 10 and 12 degrees of freedom. Find a and b such that $P(a < X < b) = 0.90$.

(3)

2. A chemist has developed a fuel additive and claims that it reduces the fuel consumption of cars. To test this claim, 8 randomly selected cars were each filled with 20 litres of fuel and driven around a race circuit. Each car was tested twice, once with the additive and once without. The distances, in miles, that each car travelled before running out of fuel are given in the table below.

Car	1	2	3	4	5	6	7	8
Distance without additive	163	172	195	170	183	185	161	176
Distance with additive	168	185	187	172	180	189	172	175

Assuming that the distances travelled follow a normal distribution and stating your hypotheses clearly test, at the 10% level of significance, whether or not there is evidence to support the chemist's claim.

(8)

3. A technician is trying to estimate the area μ^2 of a metal square. The independent random variables X_1 and X_2 are each distributed $N(\mu, \sigma^2)$ and represent two measurements of the sides of the square. Two estimators of the area, A_1 and A_2, are proposed where

$$A_1 = X_1 X_2 \quad \text{and} \quad A_2 = \left(\frac{X_1 + X_2}{2} \right)^2.$$

[You may assume that if X_1 and X_2 are independent random variables then
$$E(X_1 X_2) = E(X_1)E(X_2)]$$

(*a*) Find $E(A_1)$ and show that $E(A_2) = \mu^2 + \dfrac{\sigma^2}{2}$.

(4)

(*b*) Find the bias of each of these estimators.

(2)

The technician is told that $\text{Var}(A_1) = \sigma^4 + 2\mu^2\sigma^2$ and $\text{Var}(A_2) = \frac{1}{2}\sigma^4 + 2\mu^2\sigma^2$. The technician decided to use A_1 as the estimator for μ^2.

(*c*) Suggest a possible reason for this decision.

(1)

A statistician suggests taking a random sample of n measurements of sides of the square and finding the mean $\overline{X}$. He knows that $E(\overline{X}^2) = \mu^2 + \dfrac{\sigma^2}{n}$ and

$\text{Var}(\overline{X}^2) = \dfrac{2\sigma^4}{n^2} + \dfrac{4\sigma^2\mu^2}{n}$.

(*d*) Explain whether or not $\overline{X}^2$ is a consistent estimator of μ^2.

(3)

4. A recent census in the U.K. revealed that the heights of females in the U.K. have a mean of 160.9 cm. A doctor is studying the heights of female Indians in a remote region of South America. The doctor measured the height, x cm, of each of a random sample of 30 female Indians and obtained the following statistics.

$$\Sigma x = 4\,400.7, \quad \Sigma x^2 = 646\,904.41.$$

The heights of female Indians may be assumed to follow a normal distribution.

The doctor presented the results of the study in a medical journal and wrote 'the female Indians in this region are more than 10 cm shorter than females in the U.K.'

(*a*) Stating your hypotheses clearly and using a 5% level of significance, test the doctor's statement.

(6)

The census also revealed that the standard deviation of the heights of U.K. females was 6.0 cm.

(*b*) Stating your hypotheses clearly test, at the 5% level of significance, whether or not there is evidence that the variance of the heights of female Indians is different from that of females in the U.K.

(6)

---

5. The times, x seconds, taken by the competitors in the 100 m freestyle events at a school swimming gala are recorded. The following statistics are obtained from the data.

	No. of competitors	Sample Mean $\bar{x}$	$\sum x^2$
Girls	8	83.10	55 746
Boys	7	88.90	56 130

Following the gala a proud parent claims that girls are faster swimmers than boys. Assuming that the times taken by the competitors are two independent random samples from normal distributions,

(a) test, at the 10% level of significance, whether or not the variances of the two distributions are the same. State your hypotheses clearly.

(7)

(*b*) Stating your hypotheses clearly, test the parent's claim. Use a 5% level of significance.

(6)

6. A nutritionist studied the levels of cholesterol, X mg/cm^3, of male students at a large college. She assumed that X was distributed $N(\mu, \sigma^2)$ and examined a random sample of 25 male students. Using this sample she obtained unbiased estimates of μ and σ^2 as

$$\hat{\mu} = 1.68, \quad \hat{\sigma}^2 = 1.79.$$

(*a*) Find a 95% confidence interval for μ.

(4)

(*b*) Obtain a 95% confidence interval for σ^2.

(5)

A cholesterol reading of more than 2.5 mg/cm^3 is regarded as high.

(*c*) Use appropriate confidence limits from parts (*a*) and (*b*) to find the lowest estimate of the proportion of male students in the college with high cholesterol.

(4)

TURN OVER FOR QUESTION 7

7. A proportion p of the items produced by a factory is defective. A quality assurance manager selects a random sample of 5 items from each batch produced to check whether or not there is evidence that p is greater than 0.10. The criterion that the manager uses for rejecting the hypothesis that p is 0.10 is that there are more than 2 defective items 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.

Table 1

p	0.15	0.20	0.25	0.30	0.35	0.40
Power	0.03	r	0.10	0.16	0.24	0.32

(*b*) Find the value of r.

(3)

One day the manager is away and an assistant checks the production by random sample of 10 items from each batch produced. The hypothesis that $p = 0.10$ is rejected if more than 4 defectives are found in the sample.

(*c*) Find P(Type I error) using the assistant's test.

(2)

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

Table 2

p	0.15	0.20	0.25	0.30	0.35	0.40
Power	0.01	0.03	0.08	0.15	0.25	s

(*d*) Find the value of s.

(1)

(*e*) Using the same axes, draw the graphs of the power functions of these two tests.

(4)

(*f*) (i) State the value of p where these graphs cross.

(ii) Explain the significance if p is greater than this value.

(2)

The manager studies the graphs in part (*e*) but decides to carry on using the test based on a sample of size 5.

(*g*) Suggest 2 reasons why the manager might have made this decision.

(2)

END