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	**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.

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.

N13904A

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 $s_A^2 = 0.495$ mm^2. A sample of size 25 is taken from area *B* and the unbiased estimate for the population variance is $s_B^2 = 1.04$ 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

Paper Reference(s)

6686

Edexcel GCE

Statistics S4

Advanced/Advanced Subsidiary

Wednesday 16 June 2004 – 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.

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 8 and 12 degrees of freedom.

Find $\mathrm{P}\left(\frac{1}{5.67} < X < 2.85\right)$.

(4)

---

2. A mechanic is required to change car tyres. An inspector timed a random sample of 20 tyre changes and calculated the unbiased estimate of the population variance to be 6.25 minutes^2. Test, at the 5% significance level, whether or not the standard deviation of the population of times taken by the mechanic is greater than 2 minutes. State your hypotheses clearly.

(6)

---

3. It is suggested that a Poisson distribution with parameter λ can model the number of currants in a currant bun. A random bun is selected in order to test the hypotheses H_0: $\lambda = 8$ against H_1: $\lambda \neq 8$, using a 10% level of significance.

(*a*) Find the critical region for this test, such that the probability in each tail is as close as possible to 5%.

(5)

(*b*) Given that $\lambda = 10$, find

(i) the probability of a type II error,

(ii) the power of the test.

(4)

---

4. A doctor believes that the span of a person's dominant hand is greater than that of the weaker hand. To test this theory, the doctor measures the spans of the dominant and weaker hands of a random sample of 8 people. He subtracts the span of the weaker hand from that of the dominant hand. The spans, in mm, are summarised in the table below.

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Dominant hand	202	251	215	235	210	195	191	230
Weaker hand	195	249	218	234	211	197	181	225

Test, at the 5% significance level, the doctor's belief.

(9)

---

5. (*a*) Explain briefly what you understand by

(i) an unbiased estimator,

(ii) a consistent estimator.

of an unknown population parameter θ.

(3)

From a binomial population, in which the proportion of successes is p, 3 samples of size n are taken. The number of successes X_1, X_2, and X_3 are recorded and used to estimate p.

(*b*) Determine the bias, if any, of each of the following estimators of p.

$$\hat{p}_1 = \frac{X_1 + X_2 + X_3}{3n},$$

$$\hat{p}_2 = \frac{X_1 + 3X_2 + X_3}{6n},$$

$$\hat{p}_3 = \frac{2X_1 + 3X_2 + X_3}{6n}.$$

(4)

(*c*) Find the variance of each of these estimators.

(4)

(*d*) State, giving a reason, which of the three estimators for p is

(i) the best estimator,

(ii) the worst estimator.

(4)

6. A supervisor wishes to check the typing speed of a new typist. On 10 randomly selected occasions, the supervisor records the time taken for the new typist to type 100 words. The results, in seconds, are given below.

110, 125, 130, 126, 128, 127, 118, 120, 122, 125

The supervisor assumes that the time taken to type 100 words is normally distributed.

(*a*) Calculate a 95% confidence interval for

(i) the mean,

(ii) the variance

of the population of times taken by this typist to type 100 words.

(13)

The supervisor requires the average time needed to type 100 words to be no more than 130 seconds and the standard deviation to be no more than 4 seconds.

(*b*) Comment on whether or not the supervisor should be concerned about the speed of the new typist.

(3)

7. A grocer receives deliveries of cauliflowers from two different growers, A and B. The grocer takes random samples of cauliflowers from those supplied by each grower. He measures the weight x, in grams, of each cauliflower. The results are summarised in the table below.

	Sample size	Σx	Σx^2
A	11	6600	3960540
B	13	9815	7410579

(*a*) Show, at the 10% significance level, that the variances of the populations from which the samples are drawn can be assumed to be equal by testing the hypothesis $H_0: \sigma_A^2 = \sigma_B^2$ against hypothesis $H_1: \sigma_A^2 \neq \sigma_B^2$.

(You may assume that the two samples come from normal populations.)

(6)

The grocer believes that the mean weight of cauliflowers provided by B is at least 150 g more than the mean weight of cauliflowers provided by A.

(*b*) Use a 5% significance level to test the grocer's belief.

(8)

(*c*) Justify your choice of test.

(2)

END

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

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	**Items included with question papers**
Mathematical Formulae (Green)	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.

N22342A

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:

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

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% levcel of significnace, 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

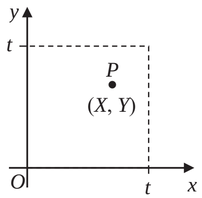


Figure 1 shows a square of side t and area t^2 which lies in the first quadrant with one vertex at the origin. A point P with coordinates (X, Y) is selected at random inside the square and the coordinates are used to estimate t^2. It is assumed that X and Y are independent random variables each having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

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

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

(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 $\mathrm{Var}\, S = \dfrac{7t^4}{9}$.

(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 $q = \dfrac{3}{2}$.

(3)

(*e*) Find 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

Paper Reference(s)

6686/01

Edexcel GCE

Statistics S4

Advanced Level

Friday 22 June 2007 – Morning

Time: 1 hour 30 minutes

Materials required for examination	**Items included with question papers**
Mathematical Formulae (Green)	Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulas stored in them.

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 7 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. A medical student is investigating two methods of taking a person's blood pressure. He takes a random sample of 10 people and measures their blood pressure using an arm cuff and a finger monitor. The table below shows the blood pressure for each person, measured by each method.

Person	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Arm cuff	140	110	138	127	142	112	122	128	132	160
Finger monitor	154	112	156	152	142	104	126	132	144	180

(*a*) Use a paired *t*-test to determine, at the 10% level of significance, whether or not there is a difference in the mean blood pressure measured using the two methods. State your hypotheses clearly.

(8)

(*b*) State an assumption about the underlying distribution of measured blood pressure required for this test.

(1)

2. The value of orders, in £, made to a firm over the internet has distribution $N(\mu, \sigma^2)$. A random sample of *n* orders is taken and $\overline{X}$ denotes the sample mean.

(*a*) Write down the mean and variance of $\overline{X}$ in terms of μ and σ^2.

(2)

A second sample of *m* orders is taken and $\overline{Y}$ denotes the mean of this sample.

An estimator of the population mean is given by

$$U = \frac{n\overline{X} + m\overline{Y}}{n+m}.$$

(*b*) Show that *U* is an unbiased estimator for μ.

(3)

(*c*) Show that the variance of *U* is $\dfrac{\sigma^2}{n+m}$.

(4)

(*d*) State which of $\overline{X}$ or *U* is a better estimator for μ. Give a reason for your answer.

(2)

3. The lengths, *x* mm, of the forewings of a random sample of male and female adult butterflies are measured. The following statistics are obtained from the data.

	No. of butterflies	Sample mean $\bar{x}$	$\sum x^2$
Females	7	50.6	17 956.5
Males	10	53.2	28 335.1

(*a*) Assuming the lengths of the forewings are normally distributed 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, at the 5% level of significance, whether the mean length of the forewings of the female butterflies is less than the mean length of the forewings of the male butterflies.

(6)

4. The length *X* mm of a spring made by a machine is normally distributed $N(\mu, \sigma^2)$. A random sample of 20 springs is selected and their lengths measured in mm. Using this sample the unbiased estimates of μ and σ^2 are

$$\bar{x} = 100.6, \qquad s^2 = 1.5.$$

Stating your hypotheses clearly test, at the 10% level of significance,

(*a*) whether or not the variance of the lengths of springs is different from 0.9,

(6)

(*b*) whether or not the mean length of the springs is greater than 100 mm.

(6)

5. The number of tornadoes per year to hit a particular town follows a Poisson distribution with mean λ. A weatherman claims that due to climate changes the mean number of tornadoes per year has decreased. He records the number of tornadoes *x* to hit the town last year.

To test the hypotheses H_0: $\lambda = 7$ and H_1: $\lambda < 7$, a critical region of $x \leq 3$ is used.

(*a*) Find, in terms λ the power function of this test.

(3)

(*b*) Find the size of this test.

(2)

(*c*) Find the probability of a Type II error when $\lambda = 4$.

(2)

6. A butter packing machine cuts butter into blocks. The weight of a block of butter is normally distributed with a mean weight of 250 g and a standard deviation of 4 g. A random sample of 15 blocks is taken to monitor any change in the mean weight of the blocks of butter.

(*a*) Find the critical region of a suitable test using a 2% level of significance.
(3)

(*b*) Assuming the mean weight of a block of butter has increased to 254 g, find the probability of a Type II error.
(5)

---

7. A doctor wishes to study the level of blood glucose in males. The level of blood glucose is normally distributed. The doctor measured the blood glucose of 10 randomly selected male students from a school. The results, in mmol/litre, are given below.

4.7 3.6 3.8 4.7 4.1 2.2 3.6 4.0 4.4 5.0

(*a*) Calculate a 95% confidence interval for the mean.
(7)

(*b*) Calculate a 95% confidence interval for the variance.
(4)

A blood glucose reading of more than 7 mmol/litre is counted as high.

(*c*) Use appropriate confidence limits from parts (*a*) and (*b*) to find the highest estimate of the proportion of male students in the school with a high blood glucose level.
(4)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)
6686/01

Edexcel GCE

Statistics S4

Advanced Level

Wednesday 18 June 2008 – Morning

Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Mathematical Formulae (Green)	Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulas stored in them.

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 7 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.

H31327A

1. A random sample $X_1, X_2, ..., X_{10}$ is taken from a population with mean μ and variance σ^2.

(*a*) Determine the bias, if any, of each of the following estimators of μ.

$$\theta_1 = \frac{X_3 + X_4 + X_5}{3},$$

$$\theta_2 = \frac{X_{10} - X_1}{3},$$

$$\theta_3 = \frac{3X_1 + 2X_2 + X_{10}}{6}.$$

(4)

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

(5)

(*c*) State, giving reasons, which of these three estimators for μ is

(i) the best estimator,

(ii) the worst estimator.

(4)

---

2. A large number of students are split into two groups *A* and *B*. The students sit the same test but under different conditions. Group *A* has music playing in the room during the test, and group *B* has no music playing during the test. Small samples are then taken from each group and their marks recorded. The marks are normally distributed.

The marks are as follows:

Sample from Group *A*	42	40	35	37	34	43	42	44	49
Sample from Group *B*	40	44	38	47	38	37	33		

(*a*) Stating your hypotheses clearly, and using a 10% level of significance, test whether or not there is evidence of a difference between the variances of the marks of the two groups.

(8)

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

(1)

(*c*) Use a two tailed test, with a 5% level of significance, to determine if the playing of music during the test has made any difference in the mean marks of the two groups. State your hypotheses clearly.

(7)

(*d*) Write down what you can conclude about the effect of music on a student's performance during the test.

(1)

3. The weights, in grams, of mice are normally distributed. A biologist takes a random sample of 10 mice. She weighs each mouse and records its weight.

The ten mice are then fed on a special diet. They are weighed again after two weeks.

Their weights in grams are as follows:

Mouse	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Weight before diet	50.0	48.3	47.5	54.0	38.9	42.7	50.1	46.8	40.3	41.2
Weight after diet	52.1	47.6	50.1	52.3	42.2	44.3	51.8	48.0	41.9	43.6

Stating your hypotheses clearly, and using a 1% level of significance, test whether or not the diet causes an increase in the mean weight of the mice.

(8)

---

4. A town council is concerned that the mean price of renting two bedroom flats in the town has exceeded £650 per month. A random sample of eight two bedroom flats gave the following results, £*x*, per month.

705, 640, 560, 680, 800, 620, 580, 760

[You may assume $\sum x = 5345$, $\sum x^2 = 3\,621\,025$.]

(*a*) Find a 90% confidence interval for the mean price of renting a two bedroom flat.

(6)

(*b*) State an assumption that is required for the validity of your interval in part (*a*).

(1)

(*c*) Comment on whether or not the town council is justified in being concerned. Give a reason for your answer.

(2)

---

5. A machine is filling bottles of milk. A random sample of 16 bottles was taken and the volume of milk in each bottle was measured and recorded. The volume of milk in a bottle is normally distributed and the unbiased estimate of the variance, s^2, of the volume of milk in a bottle is 0.003.

(*a*) Find a 95% confidence interval for the variance of the population of volumes of milk from which the sample was taken.

(5)

The machine should fill bottles so that the standard deviation of the volumes is equal to 0.07.

(*b*) Comment on this with reference to your 95% confidence interval.

(3)

6. A drug is claimed to produce a cure to a certain disease in 35% of people who have the disease. To test this claim a sample of 20 people having this disease is chosen at random and given the drug. If the number of people cured is between 4 and 10 inclusive the claim will be accepted. Otherwise the claim will not be accepted.

(*a*) Write down suitable hypotheses to carry out this test.
(2)

(*b*) Find the probability of making a Type I error.
(3)

The table below gives the value of the probability of the Type II error, to 4 decimal places, for different values of p where p is the probability of the drug curing a person with the disease.

P(cure)	0.2	0.3	0.4	0.5
P(Type II error)	0.5880	r	0.8565	s

(*c*) Calculate the value of r and the value of s.
(3)

(*d*) Calculate the power of the test for $p = 0.2$ and $p = 0.4$
(2)

(*e*) Comment, giving your reasons, on the suitability of this test procedure.
(2)

---

7. An engineering firm buys steel rods. The steel rods from its present supplier are known to have a mean tensile strength of 230 N/mm^2.

A new supplier of steel rods offers to supply rods at a cheaper price than the present supplier. A random sample of ten rods from this new supplier gave tensile strengths, x N/mm^2, which are summarised below.

Sample size	Σx	Σx^2
10	2283	524 079

(*a*) Stating your hypotheses clearly, and using a 5% level of significance, test whether or not the rods from the new supplier have a tensile strength lower than the present supplier. (You may assume that the tensile strength is normally distributed).
(7)

(*b*) In the light of your conclusion to part (*a*) write down what you would recommend the engineering firm to do.
(1)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

6686/01
Edexcel GCE
Statistics S4
Advanced Level
Friday 19 June 2009 – Afternoon
Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Mathematical Formulae (Orange or Green)	Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulas stored in them.

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 not gain full credit.

M34281A

1. A company manufactures bolts with a mean diameter of 5 mm. The company wishes to check that the diameter of the bolts has not decreased. A random sample of 10 bolts is taken and the diameters, x mm, of the bolts are measured. The results are summarised below.

$$\Sigma x = 49.1 \qquad \Sigma x^2 = 241.2$$

Using a 1% level of significance, test whether or not the mean diameter of the bolts is less than 5 mm.

(You may assume that the diameter of the bolts follows a normal distribution.)

(8)

---

2. An emission-control device is tested to see if it reduces CO_2 emissions from cars. The emissions from 6 randomly selected cars are measured with and without the device. The results are as follows.

Car	*A*	*B*	*C*	*D*	*E*	*F*
Emissions without device	151.4	164.3	168.5	148.2	139.4	151.2
Emissions with device	148.9	162.7	166.9	150.1	140.0	146.7

(*a*) State an assumption that needs to be made in order to carry out a *t*-test in this case.

(1)

(*b*) State why a paired *t*-test is suitable for use with these data.

(1)

(*c*) Using a 5% level of significance, test whether or not there is evidence that the device reduces CO_2 emissions from cars.

(8)

(*d*) Explain, in context, what a type II error would be in this case.

(2)

---

3. Define, in terms of H_0 and/or H_1,

(*a*) the size of a hypothesis test,

(1)

(*b*) the power of a hypothesis test.

(1)

The probability of getting a head when a coin is tossed is denoted by p. This coin is tossed 12 times in order to test the hypotheses H_0: $p = 0.5$ against H_1: $p \neq 0.5$, using a 5% level of significance.

(*c*) Find the largest critical region for this test, such that the probability in each tail is less than 2.5%.

(4)

(*d*) Given that $p = 0.4$

(i) find the probability of a type II error when using this test,

(ii) find the power of this test.

(4)

(*e*) Suggest two ways in which the power of the test can be increased.

(2)

---

4. A farmer set up a trial to assess whether adding water to dry feed increases the milk yield of his cows. He randomly selected 22 cows. Thirteen of the cows were given dry feed and the other 9 cows were given the feed with water added. The milk yields, in litres per day, were recorded with the following results.

	Sample size	Mean	s^2
Dry feed	13	25.54	2.45
Feed with water added	9	27.94	1.02

You may assume that the milk yield from cows given the dry feed and the milk yield from cows given the feed with water added are from independent normal distributions.

(*a*) Test, at the 10% level of significance, whether or not the variances of the populations from which the samples are drawn are the same. State your hypotheses clearly.

(5)

(*b*) Calculate a 95% confidence interval for the difference between the two mean milk yields.

(7)

(*c*) Explain the importance of the test in part (*a*) to the calculation in part (*b*).

(2)

---

5. A machine fills jars with jam. The weight of jam in each jar is normally distributed. To check the machine is working properly the contents of a random sample of 15 jars are weighed in grams. Unbiased estimates of the mean and variance are obtained as

$$\hat{\mu} = 560 \qquad s^2 = 25.2$$

Calculate a 95% confidence interval for,

(a) the mean weight of jam,

(4)

(b) the variance of the weight of jam.

(5)

A weight of more than 565 g is regarded as too high and suggests the machine is not working properly.

(c) Use appropriate confidence limits from parts (a) and (b) to find the highest estimate of the proportion of jars that weigh too much.

(5)

---

6. A continuous uniform distribution on the interval $[0, k]$ has mean $\frac{k}{2}$ and variance $\frac{k^2}{12}$.

A random sample of three independent variables X_1, X_2 and X_3 is taken from this distribution.

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

(3)

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

(b) Show that $\text{Var}(\hat{k}) = (a^2 - 2a + 2)\frac{k^2}{6}$

(6)

(c) Hence determine the value of a and the value of b for which $\hat{k}$ has minimum variance, and calculate this minimum variance.

(6)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

6686/01

Edexcel GCE

Statistics S4

Advanced Level

Thursday 24 June 2010 – Morning

Time: 1 hour 30 minutes

<u>Materials required for examination</u>	<u>Items included with question papers</u>
Mathematical Formulae (Pink)	Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulas stored in them.

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 not gain full credit.

M35398A

1. A teacher wishes to test whether playing background music enables students to complete a task more quickly. The same task was completed by 15 students, divided at random into two groups. The first group had background music playing during the task and the second group had no background music playing.

The times taken, in minutes, to complete the task are summarised below.

	Sample size n	Standard deviation s	Mean $\bar{x}$
With background music	8	4.1	15.9
Without background music	7	5.2	17.9

You may assume that the times taken to complete the task by the students are two independent random samples from normal distributions.

(*a*) Stating your hypotheses clearly, test, at the 10% level of significance, whether or not the variances of the times taken to complete the task with and without background music are equal.
(5)

(*b*) Find a 99% confidence interval for the difference in the mean times taken to complete the task with and without background music.
(7)

Experiments like this are often performed using the same people in each group.

(*c*) Explain why this would not be appropriate in this case.
(1)

---

2. As part of an investigation, a random sample of 10 people had their heart rate, in beats per minute, measured whilst standing up and whilst lying down. The results are summarized below.

Person	1	2	3	4	5	6	7	8	9	10
Heart rate lying down	66	70	59	65	72	66	62	69	56	68
Heart rate standing up	75	76	63	67	80	75	65	74	63	75

(*a*) State one assumption that needs to be made in order to carry out a paired *t*-test.
(1)

(*b*) Test, at the 5% level of significance, whether or not there is any evidence that standing up increases people's mean heart rate by more than 5 beats per minute. State your hypotheses clearly.
(8)

---

3. A manager in a sweet factory believes that the machines are working incorrectly and the proportion p of underweight bags of sweets is more than 5%. He decides to test this by randomly selecting a sample of 5 bags and recording the number X that are underweight. The manager sets up the hypotheses H_0: $p = 0.05$ and H_1: $p > 0.05$ and rejects the null hypothesis if $x > 1$.

(*a*) Find the size of the test.
(2)

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

$$1 - (1 - p)^4(1 + 4p)$$
(3)

The manager goes on holiday and his deputy checks the production by randomly selecting a sample of 10 bags of sweets. He rejects the hypothesis that $p = 0.05$ if more than 2 underweight bags are found in the sample.

(*c*) Find the probability of a Type I error using the deputy's test.
(2)

The table below gives some values, to 2 decimal places, of the power function for the deputy's test.

p	0.10	0.15	0.20	0.25
Power	0.07	s	0.32	0.47

(*d*) Find the value of s.
(1)

Question 3 continues on page 4

The graph of the power function for the manager's test is shown in Figure 1.

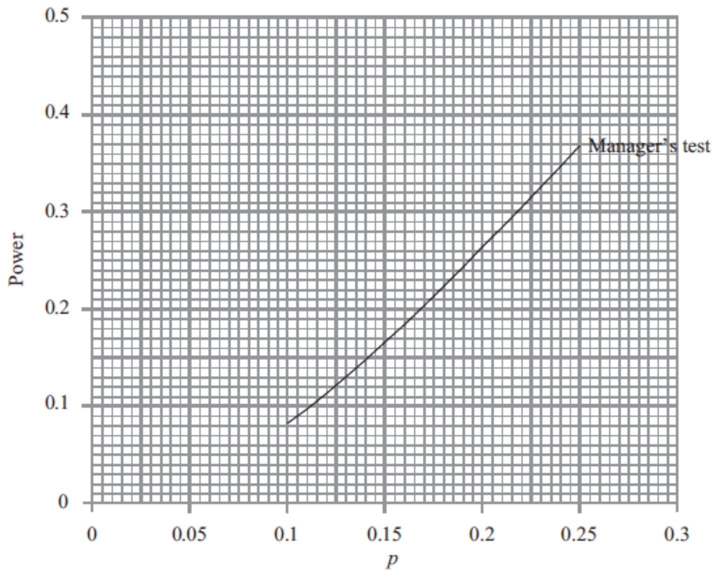


Figure 1

(e) On the same axes, draw the graph of the power function for the deputy's test. **(1)**

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

(ii) Compare the effectiveness of the two tests if p is greater than this value. **(2)**

The deputy suggests that they should use his sampling method rather than the manager's.

(g) Give a reason why the manager might not agree to this change. **(1)**

4. A random sample of 15 strawberries is taken from a large field and the weight x grams of each strawberry is recorded. The results are summarised below.

$$\sum x = 291 \qquad \sum x^2 = 5968$$

Assume that the weights of strawberries are normally distributed.

Calculate a 95% confidence interval for

(a) (i) the mean of the weights of the strawberries in the field,

(ii) the variance of the weights of the strawberries in the field. **(12)**

Strawberries weighing more than 23 g are considered to be less tasty.

(b) Use appropriate confidence limits from part (a) to find the highest estimate of the proportion of strawberries that are considered to be less tasty. **(4)**

5. A car manufacturer claims that, on a motorway, the mean number of miles per gallon for the Panther car is more than 70. To test this claim a car magazine measures the number of miles per gallon, x, of each of a random sample of 20 Panther cars and obtained the following statistics.

$$\bar{x} = 71.2 \qquad s = 3.4$$

The number of miles per gallon may be assumed to be normally distributed.

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

The standard deviation of the number of miles per gallon for the Tiger car is 4.

(b) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not there is evidence that the variance of the number of miles per gallon for the Panther car is different from that of the Tiger car. **(6)**

6. Faults occur in a roll of material at a rate of λ per m^2. To estimate λ, three pieces of material of sizes 3 m^2, 7 m^2 and 10 m^2 are selected and the number of faults X_1, X_2 and X_3 respectively are recorded.

The estimator $\hat{\lambda}$, where

$$\hat{\lambda} = k(X_1 + X_2 + X_3)$$

is an unbiased estimator of λ.

(*a*) Write down the distributions of X_1, X_2 and X_3 and find the value of k.
(4)

(*b*) Find Var ($\hat{\lambda}$).
(3)

A random sample of n pieces of this material, each of size 4 m^2, was taken. The number of faults on each piece, Y, was recorded.

(*c*) Show that $\frac{1}{4}\overline{Y}$ is an unbiased estimator of λ.
(2)

(*d*) Find Var ($\frac{1}{4}\overline{Y}$).
(3)

(*e*) Find the minimum value of n for which $\frac{1}{4}\overline{Y}$ becomes a better estimator of λ than $\hat{\lambda}$.
(2)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)
6686/01

Edexcel GCE

Statistics S4

Advanced Level

Thursday 23 June 2011 – Morning

Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Mathematical Formulae (Pink)	Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulas stored in them.

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 7 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 not gain full credit.

P35418A

1. Find the value of the constant a such that

$$P(a < F_{8,\,10} < 3.07) = 0.94.$$

(2)

---

2. Two independent random samples $X_1,\ X_2,\ ...,\ X_7$ and $Y_1,\ Y_2,\ Y_3,\ Y_4$ were taken from different normal populations with a common standard deviation σ.

The following sample statistics were calculated.

$$s_x = 14.67 \qquad s_y = 12.07$$

Find the 99% confidence interval for σ^2 based on these two samples.

(5)

---

3. Manuel is planning to buy a new machine to squeeze oranges in his cafe and he has two models, at the same price, on trial. The manufacturers of machine B claim that their machine produces more juice from an orange than machine A. To test this claim Manuel takes a random sample of 8 oranges, cuts them in half and puts one half in machine A and the other half in machine B. The amount of juice, in ml, produced by each machine is given in the table below.

Orange	1	2	3	4	5	6	7	8
Machine A	60	58	55	53	52	51	54	56
Machine B	61	60	58	52	55	50	52	58

Stating your hypotheses clearly, test, at the 10% level of significance, whether or not the mean amount of juice produced by machine B is more than the mean amount produced by machine A.

(8)

Question 4 follows on page 4

4. A proportion p of letters sent by a company are incorrectly addressed and if p is thought to be greater than 0.05 then action is taken.

Using H_0: $p = 0.05$ and H_1: $p > 0.05$, a manager from the company takes a random sample of 40 letters and rejects H_0 if the number of incorrectly addressed letters is more than 3.

(*a*) Find the size of this test. **(2)**

(*b*) Find the probability of a Type II error in the case where p is in fact 0.10. **(2)**

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

p	0.075	0.100	0.125	0.150	0.175	0.200	0.225
Power	0.35	s	0.75	0.87	0.94	0.97	0.99

Table 1

(*c*) Write down the value of s. **(1)**

A visiting consultant uses an alternative system to test the same hypotheses. A sample of 15 letters is taken. If these are all correctly addressed then H_0 is accepted. If 2 or more are found to have been incorrectly addressed then H_0 is rejected. If only one is found to be incorrectly addressed then a further random sample of 15 is taken and H_0 is rejected if 2 or more are found to have been incorrectly addressed in this second sample, otherwise H_0 is accepted.

(*d*) Find the size of the test used by the consultant. **(3)**

Figure 1 shows the graph of the power function of the test used by the consultant.

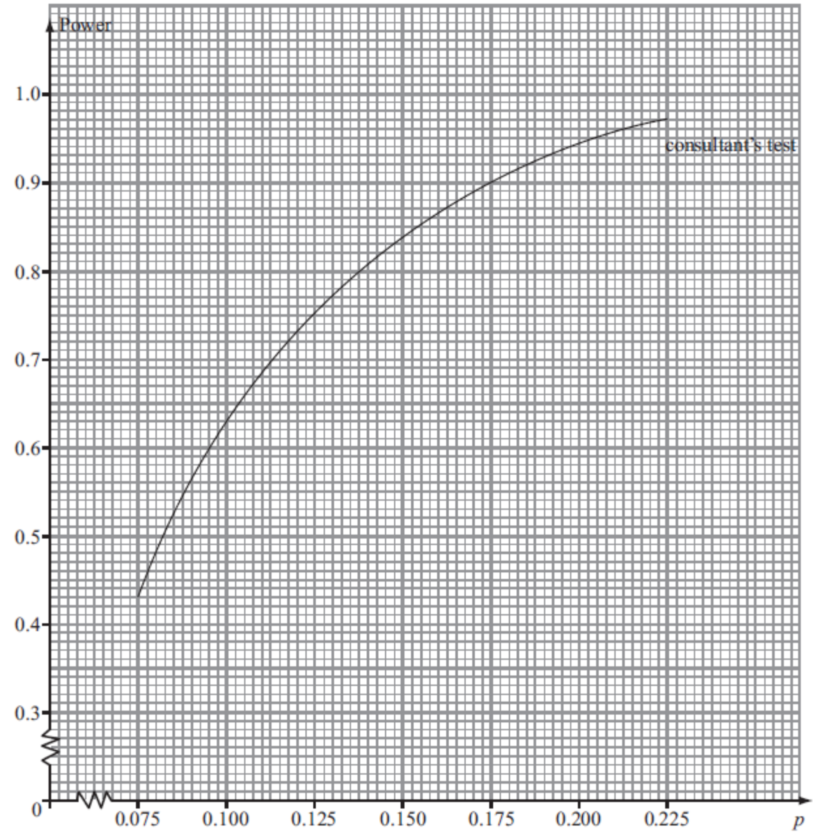


Figure 1

(*e*) On Figure 1 draw the graph of the power function of the manager's test. **(2)**

(*f*) State, giving your reasons, which test you would recommend. **(2)**

5. The weights of the contents of breakfast cereal boxes are normally distributed.

A manufacturer changes the style of the boxes but claims that the weight of the contents remains the same. A random sample of 6 old style boxes had contents with the following weights (in grams).

512 503 514 506 509 515

The weights, y grams, of the contents of an independent random sample of 5 new style boxes gave

$$\bar{y} = 504.8 \text{ and } s_y = 3.420$$

(*a*) Use a two-tail test to show, at the 10% level of significance, that the variances of the weights of the contents of the old and new style boxes can be assumed to be equal. State your hypotheses clearly.
(5)

(*b*) Showing your working clearly, find a 90% confidence interval for $\mu_x - \mu_y$, where μ_x and μ_y are the mean weights of the contents of old and new style boxes respectively.
(7)

(*c*) With reference to your confidence interval comment on the manufacturer's claim.
(2)

6. A random sample $X_1, X_2, ..., X_n$ is taken from a population where each of the X_i have a continuous uniform distribution over the interval $[0, \beta]$.

The random variable $Y = \max\{X_1, X_2, ..., X_n\}$.

The probability density function of Y is given by

$$f(y) = \begin{cases} \dfrac{n}{\beta^n} y^{n-1} & 0 \le y \le \beta, \\ 0 & \text{otherwise.} \end{cases}$$

(*a*) Show that $E(Y^m) = \dfrac{n}{n+m}\beta^m$.
(3)

(*b*) Write down $E(Y)$.
(1)

(*c*) Using your answers to parts (*a*) and (*b*), or otherwise, show that

$$\text{Var}(Y) = \frac{n}{(n+1)^2(n+2)}\beta^2.$$
(3)

(*d*) State, giving your reasons, whether or not Y is a consistent estimator of β.
(3)

The random variables $M = 2\bar{X}$, where $\bar{X} = \dfrac{1}{n}(X_1 + X_2 + ... + X_n)$, and $S = kY$, where k is a constant, are both unbiased estimators of β.

(*e*) Find the value of k in terms of n.
(1)

(*f*) State, giving your reasons, which of M and S is the better estimator of β in this case.
(3)

Five observations of X are: 8.5 6.3 5.4 9.1 7.6

(*g*) Calculate the better estimate of β.
(2)

7. A machine produces components whose lengths are normally distributed with mean 102.3 mm and standard deviation 2.8 mm. After the machine had been serviced, a random sample of 20 components were tested to see if the mean and standard deviation had changed. The lengths, x mm, of each of these 20 components are summarised as

$$\sum x = 2072 \qquad \sum x^2 = 214\,856$$

(*a*) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not there is evidence of a change in standard deviation.
(7)

(*b*) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not the mean length of the components has changed from the original value of 102.3 mm using

(i) a normal distribution,

(ii) a *t* distribution.
(9)

(*c*) Comment on the mean length of components produced after the service in the light of the tests from part (*a*) and part (*b*). Give a reason for your answer.
(2)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

6686/01

Edexcel GCE

Statistics S4

Advanced Level

Friday 22 June 2012 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Mathematical Formulae (Pink)	Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulas stored in them.

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 not gain full credit.

P40473A

1. A medical student is investigating whether there is a difference in a person's blood pressure when sitting down and after standing up. She takes a random sample of 12 people and measures their blood pressure, in mmHg, when sitting down and after standing up.

The results are shown below.

Person	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*
Sitting down	135	146	138	146	141	158	136	135	146	161	119	151
Standing up	131	147	132	140	138	160	127	136	142	154	130	144

The student decides to carry out a paired *t*-test to investigate whether, on average, the blood pressure of a person when sitting down is more than their blood pressure after standing up.

(*a*) State clearly the hypotheses that should be used and any necessary assumption that needs to be made.

(2)

(*b*) Carry out the test at the 1% level of significance.

(7)

2. A biologist investigating the shell size of turtles takes random samples of adult female and adult make turtles and records the length, x cm, of the shell. The results are summarised below.

	Number in sample	Sample mean $\bar{x}$	$\sum x^2$
Female	6	19.6	2308.01
Male	12	13.7	2262.57

You may assume that samples come from independent normal distributions with the same variance.

The biologist claims that the mean shell length of adult female turtles is 5 cm longer than the mean shell length of adult male turtles.

(*a*) Test the biologist's claim at the 5% level of significance.

(10)

(*b*) Given that the true values for the variance of the population of adult male turtles and adult female turtles are both 0.9 cm^2,

(i) show that when samples of 6 and 12 are used with a 5% level of significance, the biologist's claim will be accepted if $4.07 < \bar{X}_F - \bar{X}_M < 5.93$, where $\bar{X}_F$ and $\bar{X}_M$ are the mean shell lengths of females and males respectively.

(ii) Hence find the probability of a type II error for this test if in fact the true mean shell length of adult female turtles is 6 cm more than the mean shell length of adult male turtles.

(6)

3. The sample variance of the lengths of a random sample of 9 paving slabs sold by a builders' merchant is 36 mm^2. The sample variance of the lengths of a random sample of 11 paving slabs sold by a second builder's merchant is 225 mm^2. Test at the 10% significance level whether or not there is evidence that the lengths of paving slabs sold by these builders' merchants differ in variability. State your hypotheses clearly.

(You may assume the lengths of paving slabs are normally distributed.)

(5)

4. A newspaper runs a daily Sudoku. A random sample of 10 people took the following times, in minutes, to complete the Sudoku.

5.0 4.5 4.7 5.3 5.2 4.1 5.3 4.8 5.5 4.6

Given that the time to complete the Sudoku follow a normal distribution,

(*a*) calculate a 95% confidence interval for

(i) the mean,

(ii) the variance,

of the times taken by people to complete the Sudoku.

(13)

The newspaper requires the average time needed to complete the Sudoku to be 5 minutes with a standard deviation of 0.7 minutes.

(*b*) Comment on whether or not the Sudoku meets this requirement. Give a reason for your answer.

(3)

---

5. Boxes of chocolates manufactured by Philippe have a mean weight of μ grams and a standard deviation of σ grams. A random sample of 25 of these boxes are weighed. Using this sample, the unbiased estimate of μ is 455 and the unbiased estimate of σ^2 is 55.

(*a*) Test, at the 5% level of significance, whether or not σ is greater than 6. State your hypotheses clearly.

(6)

(*b*) Test, at the 5% level of significance, whether or not μ is more than 450.

(6)

(*c*) State an assumption you have made in order to carry out the above tests.

(1)

---

6. When a tree is planted the probability of it germinating is p.

A random sample of size n is taken and the number of tree seeds, X, which germinate is recorded.

(*a*) (i) Show that $\hat{p}_1 = \dfrac{X}{n}$ is an unbiased estimator of p.

(ii) Find the variance of $\hat{p}_1$.

(4)

A second sample of size m is taken and the number of tree seeds, Y, which germinate is recorded.

Given that $\hat{p}_2 = \dfrac{Y}{m}$ and that $\hat{p}_3 = a(3\hat{p}_1 + 2\hat{p}_2)$ is an unbiased estimator of p,

(*b*) show that

(i) $a = \dfrac{1}{5}$,

(ii) $\text{Var}(\hat{p}_3) = \dfrac{p(1-p)}{25}\left(\dfrac{9}{n} + \dfrac{4}{m}\right)$.

(6)

(*c*) Find the range of values of $\dfrac{n}{m}$ for which

$$\text{Var}(\hat{p}_3) < \text{Var}(\hat{p}_1) \text{ and } \text{Var}(\hat{p}_3) < \text{Var}(\hat{p}_2).$$

(3)

(*d*) Given that $n = 20$ and $m = 60$, explain which of $\hat{p}_1$, $\hat{p}_2$ or $\hat{p}_3$ is the best estimator.

(3)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

6686/01R

Edexcel GCE

Statistics S4 (R)

Advanced/Advanced Subsidiary

Friday 21 June 2013 – Morning

Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Mathematical Formulae (Pink)	Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer for each question in the space following the question.
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.
The marks for the parts of questions are shown in round brackets, e.g. **(2)**.
There are 8 questions in this question paper. The total mark for this paper is 75.
There are 24 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 not gain full credit.

P42960A

1. (*a*) Find the value of the constant *a* such that

$$P(1.690 < \chi^2_7 < a) = 0.95$$

(2)

The random variable *Y* follows an *F*-distribution with 6 and 4 degrees of freedom.

(b) (i) Find the upper 1% critical value for *Y*.

(ii) Find the lower 1% critical value for *Y*.

(2)

---

2. The time, *t* hours, that a typist can sit before incurring back pain is modelled by N(μ, σ^2). A random sample of 30 typists gave unbiased estimates for μ and σ^2 as shown below.

$$\hat{\mu} = 2.5 \qquad s^2 = 0.36$$

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

(5)

(*b*) State with a reason whether or not the confidence interval supports the assertion that $\sigma^2 = 0.495$.

(2)

---

3. The number of houses sold per week by a firm of estate agents follows a Poisson distribution with mean 2. The firm believes that the appointment of a new salesman will increase the number of houses sold. The firm tests its belief by recording the number of houses sold, x, in the week following the appointment. The firm sets up the hypotheses $H_0: \lambda = 2$ and $H_1: \lambda > 2$, where λ is the mean number of houses sold per week, and rejects the null hypothesis if $x \geq 3$.

(*a*) Find the size of the test.
(2)

(*b*) Show that the power function for this test is

$$1 - \frac{1}{2}e^{-\lambda}(2 + 2\lambda + \lambda^2)$$

(3)

The table below gives the values of the power function to 2 decimal places.

λ	2.5	3.0	3.5	4.0	5.0	7.0
Power	0.46	r	0.68	s	0.88	0.97

Table 1

(*c*) Calculate the values of r and s.
(2)

(*d*) Draw a graph of the power function on the graph paper provided on page 4.
(2)

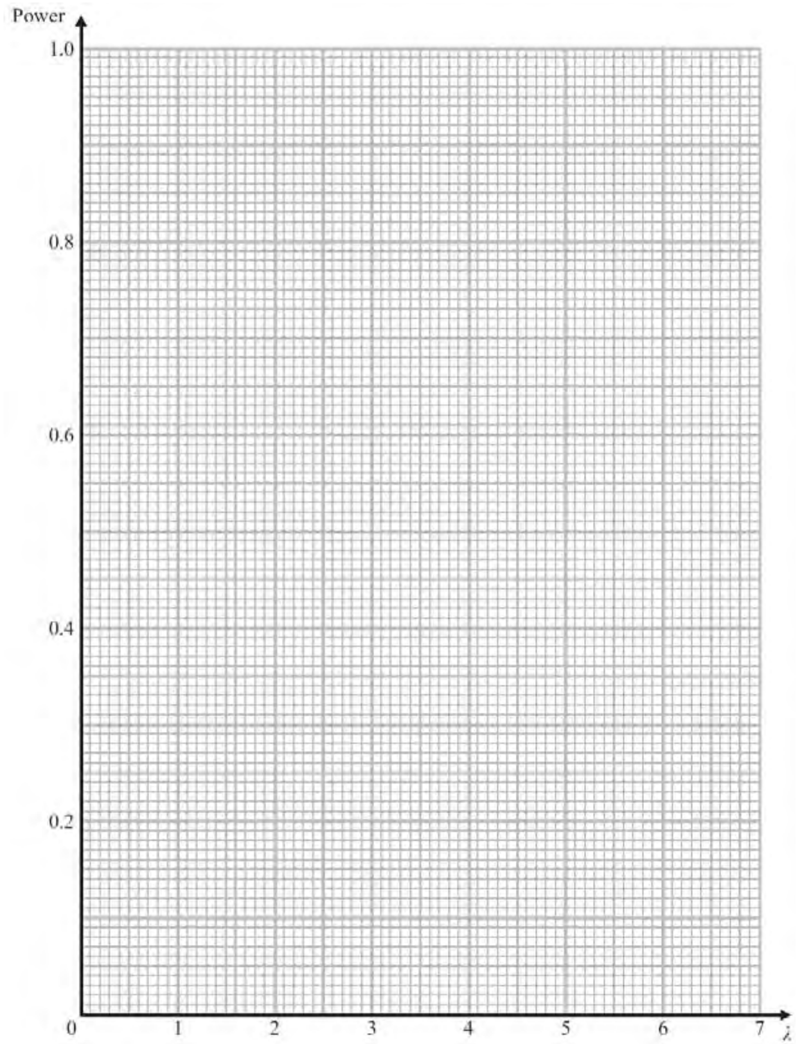
(*e*) Find the range of values of λ for which the power of this test is greater than 0.6.
(1)

Question 3 continued

For your convenience Table 1 is repeated here.

λ	2.5	3.0	3.5	4.0	5.0	7.0
Power	0.46	r	0.68	s	0.88	0.97

Table 1



4. A company carries out an investigation into the strengths of rods from two different suppliers, Ardo and Bards. Independent random samples of rods were taken from each supplier and the force, x kN, needed to break each rod was recorded. The company wrote the results on a piece of paper but unfortunately spilt ink on it so some of the results can not be seen.
The paper with the results on is shown below.

Ardo: 13.1 13.6 13.2 13.8 12.8 13.5 13.8

Bards: 15.3 15.5 14.1 15.4 14.2 15.4

Ardo $n_A = 7$ $\bar{x}_A = 13.4$

Bards $n_B = 9$ $\bar{x}_B = 14.8$

Pooled estimate of variance = 0.261

(*a*) (i) Use the data from Ardo to calculate an unbiased estimate, s_A^2, of the variance.

(i) Hence find an unbiased estimate, s_B^2, of the variance for the sample of 9 values from Bards.

(4)

(*b*) Stating your hypotheses clearly, test at the 10% level of significance whether or not there is a difference in variability of strength between the rods from Ardo and the rods from Bards.
(You may assume the two samples come from independent normal distributions.)

(5)

(*c*) Use a 5% level of significance to test whether the mean strength of rods from Bards is more than 0.9 kN greater than the mean strength of rods from Ardo.

(6)

5. Students studying for their Mathematics GCSE are assessed by two examination papers. A teacher believes that on average the score on paper I is more than 1 mark higher than the score on paper II. To test this belief the scores of 8 randomly selected students are recorded. The results are given in the table below.

Student	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Score on paper I	57	63	68	81	43	65	52	31
Score on paper II	53	62	61	78	44	64	43	29

Assuming that the scores are normally distributed and stating your hypotheses clearly, test at the 5% level of significance whether or not there is evidence to support the teacher's belief.

(8)

6. A machine fills bottles with water. The amount of water in each bottle is normally distributed. To check the machine is working properly, a random sample of 12 bottles is selected and the amount of water, in ml, in each bottle is recorded. Unbiased estimates for the mean and variance are

$$\hat{\mu} = 502 \qquad s^2 = 5.6$$

Stating your hypotheses clearly, test at the 1% level of significance

(*a*) whether or not the mean amount of water in a bottle is more than 500 ml,

(5)

(*b*) whether or not the standard deviation of the amount of water in a bottle is less than 3 ml.

(5)

7. A machine produces bricks. The lengths, x mm, of the bricks are distributed $N(\mu, 2^2)$. At the start of each week a random sample of n bricks is taken to check the machine is working correctly.
A test is then carried out at the 1% level of significance with

$$H_0: \mu = 202 \quad \text{and} \quad H_1: \mu < 202$$

(*a*) Find, in terms of n, the critical region of the test.
(3)

The probability of a type II error, when $\mu = 200$, is less than 0.05.

(*b*) Find the minimum value of n.
(6)

---

8. A random sample $W_1, W_2 \ldots, W_n$ is taken from a distribution with mean μ and variance σ^2.

(*a*) Write down $\mathrm{E}\left(\sum_{i=1}^{n} W_i\right)$ and show that $\mathrm{E}\left(\sum_{i=1}^{n} W_i^2\right) = n(\sigma^2 + \mu^2)$.
(4)

An estimator for μ is

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} W_i$$

(*b*) Show that $\bar{X}$ is a consistent estimator for μ.
(3)

An estimator of σ^2 is

$$U = \frac{1}{n}\sum_{i=1}^{n} W_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} W_i\right)^2$$

(*c*) Find the bias of U.
(4)

(*d*) Write down an unbiased estimator of σ^2 in the form kU, where k is in terms of n.
(1)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

6686/01

Edexcel GCE

Statistics S4

Advanced/Advanced Subsidiary

Friday 21 June 2013 – Morning

Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Mathematical Formulae (Pink)	Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer for each question in the space following the question.
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.
The marks for the parts of questions are shown in round brackets, e.g. **(2)**.
There are 6 questions in this question paper. The total mark for this paper is 75.
There are 20 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 not gain full credit.

P41806A

1. George owns a garage and he records the mileage of cars, x thousands of miles, between services. The results from a random sample of 10 cars are summarised below.

$$\sum x = 113.4 \qquad \sum x^2 = 1414.08$$

The mileage of cars between services is normally distributed and George believes that the standard deviation is 2.4 thousand miles.

Stating your hypotheses clearly, test, at the 5% level of significance, whether or not these data support George's belief.

(7)

---

2. Every 6 months some engineers are tested to see if their times, in minutes, to assemble a particular component have changed. The times taken to assemble the component are normally distributed. A random sample of 8 engineers was chosen and their times to assemble the component were recorded in January and in July. The data are given in the table below.

Engineer	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
January	17	19	22	26	15	28	18	21
July	19	18	25	24	17	25	16	19

(*a*) Calculate a 95% confidence interval for the mean difference in times.

(7)

(*b*) Use your confidence interval to state, giving a reason, whether or not there is evidence of a change in the mean time to assemble a component. State your hypotheses clearly.

(3)

---

3. An archaeologist is studying the compression strength of bricks at some ancient European sites. He took random samples from two sites A and B and recorded the compression strength of these bricks in appropriate units. The results are summarised below.

Site	Sample size (n)	Sample mean $(\bar{x})$	Standard deviation (s)
A	7	8.43	4.24
B	13	14.31	4.37

It can be assumed that the compression strength of bricks is normally distributed.

(a) Test, at the 2% level of significance, whether or not there is evidence of a difference in the variances of compression strength of the bricks between these two sites. State your hypotheses clearly.

(5)

Site A is older than site B and the archaeologist claims that the mean compression strength of the bricks was greater at the younger site.

(b) Stating your hypotheses clearly and using a 1% level of significance, test the archaeologist's claim.

(6)

(c) Explain briefly the importance of the test in part (a) to the test in part (b).

(1)

4. A random sample of size 2, X_1 and X_2, is taken from the random variable X which has a continuous uniform distribution over the interval $[-a, 2a]$, $a > 0$.

(a) Show that $\bar{X} = \dfrac{X_1 + X_2}{2}$ is a biased estimator of a and find the bias.

(3)

The random variable $Y = k\bar{X}$ is an unbiased estimator of a.

(b) Write down the value of the constant k.

(1)

(c) Find Var(Y).

(4)

The random variable M is the maximum of X_1 and X_2.

The probability density function, $m(x)$, of M is given by

$$m(x) = \begin{cases} \dfrac{2(x+a)}{9a^2} & -a \le x \le 2a \\ 0 & \text{otherwise} \end{cases}$$

(d) Show that M is an unbiased estimator of a.

(4)

Given that $\mathrm{E}(M^2) = \dfrac{3}{2}a^2$.

(e) find Var(M).

(1)

(f) State, giving a reason, whether you would use Y or M as an estimator of a.

(2)

A random sample of two values of X are 5 and –1.

(g) Use your answer to part (f) to estimate a.

(1)

5. Water is tested at various stages during a purification process by an environmental scientist. A certain organism occurs randomly in the water at a rate of λ every 10 ml. The scientist selects a random sample of 20 ml of water to check whether there is evidence that λ is greater than 1. The criterion the scientist uses for rejecting the hypothesis that $\lambda = 1$ is that there are 4 or more organisms in the sample of 20 ml.

(*a*) Find the size of the test.
(2)

(*b*) When $\lambda = 2.5$ find P(Type II error).
(2)

A statistician suggests using an alternative test. The statistician's test involves taking a random sample of 10 ml and rejecting the hypothesis that $\lambda = 1$ if 2 or more organisms are present but accepting the hypothesis if no organisms are in the sample. If only 1 organism is found then a second random sample of 10 ml is taken and the hypothesis is rejected if 2 or more organisms are present, otherwise the hypothesis is accepted.

(*c*) Show that the power of the statistician's test is given by

$$1 - e^{-\lambda} - \lambda(1 + \lambda)e^{-2\lambda}$$
(4)

Table 1 below gives some values, to 2 decimal places, of the power function of the statistician's test.

λ	1.5	2	2.5	3	3.5	4
Power	0.59	0.75	0.86	r	0.96	0.97

Table 1

(*d*) Find the value of r.
(1)

Question 5 continues on page 6

Question 5 continued

For your convenience Table 1 is repeated here.

λ	1.5	2	2.5	3	3.5	4
Power	0.59	0.75	0.86	r	0.96	0.97

Table 1

Figure 1 shows a graph of the power function for the scientist's test.

(*e*) On the same axes draw the graph of the power function for the statistician's test.
(2)

Given that it takes 20 minutes to collect and test a 20 ml sample and 15 minutes to collect and test a 10 ml sample

(*f*) show that the expected time of the statistician's test is slower than the scientist's test for $\lambda e^{-\lambda} > \frac{1}{3}$.
(4)

(*g*) By considering the times when $\lambda = 1$ and $\lambda = 2$ together with the power curves in part (*e*) suggest, giving a reason, which test you would use.
(2)

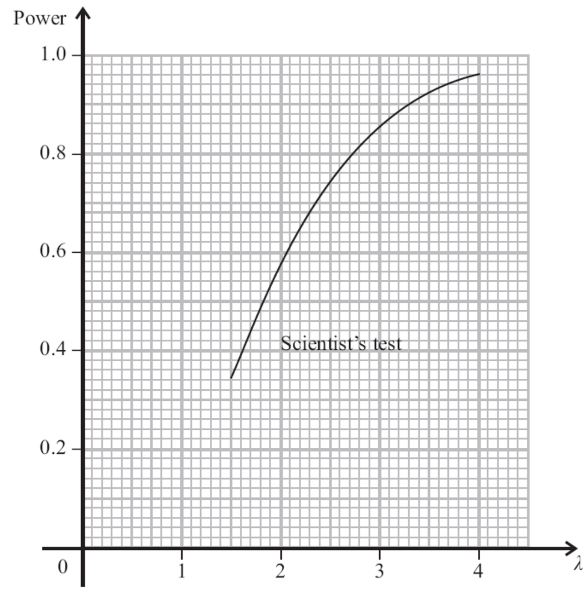


Figure 1

6. The carbon content, measured in suitable units, of steel is normally distributed. Two independent random samples of steel were taken from a refining plant at different times and their carbon content recorded. The results are given below.

Sample *A*:	1.5	0.9	1.3	1.2		
Sample *B*:	0.4	0.6	0.8	0.3	0.5	0.4

(*a*) Stating your hypotheses clearly, carry out a suitable test, at the 10% level of significance, to show that both samples can be assumed to have come from populations with a common variance σ^2.

(7)

(*b*) Showing your working clearly, find the 99% confidence interval for σ^2 based on both samples.

(6)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)
6686/01R

Edexcel GCE

Statistics S4 (R)

Advanced/Advanced Subsidiary

Thursday 12 June 2014 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Mathematical Formulae (Pink)	Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer for each question in the space following the question.
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.
The marks for the parts of questions are shown in round brackets, e.g. **(2)**.
There are 6 questions in this question paper. The total mark for this paper is 75.
There are 24 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 not gain full credit.

P43154A

1. In a trial for a new cough medicine, a random sample of 8 healthy patients were given steadily increasing doses of a pepper extract until they started coughing. The level of pepper that triggered the coughing was recorded. Each patient completed the trial after taking a standard cough medicine and, at a later time, after taking the new medicine. The results are given in the table below.

	Level of pepper extract that triggers coughing							
Patient	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Standard medicine	46	12	18	31	23	16	27	9
New medicine	53	16	13	49	11	34	38	22

(*a*) Using a suitable test, at the 5% level of significance, state whether or not, on the basis of this trial, you would recommend using the new medicine. State your hypotheses clearly.
(8)

(b) State an assumption needed to carry out this test.
(1)

---

2. The cloth produced by a certain manufacturer has defects that occur randomly at a constant rate of λ per square metre. If λ is thought to be greater than 1.5 then action has to be taken.

Using H_0: $\lambda = 1.5$ and H_1: $\lambda > 1.5$ a quality control officer takes a 4 m^2 sample of cloth and rejects H_0 if there are 11 or more defects. If there are 8 or fewer defects she accepts H_0. If there are 9 or 10 defects a second sample of 4 m^2 is taken and H0 is rejected if there are 11 or more defects in this second sample, otherwise it is accepted.

(*a*) Find the size of this test.
(4)

(*b*) Find the power of this test when $\lambda = 2$.
(3)

---

3. A farmer is investigating the milk yields of two breeds of cow. He takes a random sample of 9 cows of breed *A* and an independent random sample of 12 cows of breed *B*. For a 5 day period he measures the amount of milk, *x* gallons, produced by each cow. The results are summarised in the table below.

Breed	**Sample size**	**Mean ($\bar{x}$)**	**Standard deviation (s_x)**
A	9	6.23	2.98
B	12	7.13	2.33

The amount of milk produced by each cow can be assumed to follow a normal distribution.

(*a*) Use a two-tail test to show, at the 10% level of significance, that the variances of the yields of the two breeds can be assumed to be equal. State your hypotheses clearly.
(4)

(*b*) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not there is a difference in the mean yields of the two breeds of cow.
(7)

(*c*) Explain briefly the importance of the test in part (*a*) for the test in part (*b*).
(1)

---

4. At the start of each academic year, a large college carries out a diagnostic test on a random sample of new students. Past experience has shown that the standard deviation of the scores on this test is 19.71.

The admissions tutor claimed that the new students in 2013 would have more varied scores than usual. The scores for the students taking the test can be assumed to come from a normal distribution. A random sample of 10 new students was taken and the score *x*, for each student was recorded. The data are summarised as $\sum x = 619$ $\sum x^2 = 42\,397$.

(*a*) Stating your hypotheses clearly, and using a 5% level of significance, test the admission tutor's claim.
(6)

The admissions tutor decides that in future he will use the same hypotheses but take a larger sample of size 30 and use a significance level of 1%.

(*b*) Use the tables to show that, to 3 decimal places, the critical region for S^2 is $S^2 > 664.281$.
(3)

(*c*) Find the probability of a type II error using this test when the true value of the standard deviation is in fact 22.20.
(3)

---

5. A large company has designed an aptitude test for new recruits. The score, S, for an individual taking the test, has a normal distribution with mean μ and standard deviation σ.

In order to estimate μ and σ, a random sample of 15 new recruits were given the test and their scores, x, are summarised as

$$\sum x = 880 \qquad \sum x^2 = 54\,892$$

(*a*) Calculate a 95% confidence interval for

(i) μ,

(ii) σ.

(11)

The company wants to ensure that no more than 80% of new recruits pass the test.

(*b*) Using values from your confidence intervals in part (*a*), estimate the lowest pass mark they should set.

(5)

6. Emily is monitoring the level of pollution in a river. Over a period of time she has found that the amount of pollution, X, in a 100 ml sample of river water has a continuous distribution with probability density function f(x) given by

$$f(x) = \begin{cases} \dfrac{2x}{a^2} & 0 \le x \le a \\ 0 & \text{otherwise} \end{cases}$$

where a is a constant.

Emily takes a random sample $X_1, X_2, X_3, \ldots, X_n$ to try to estimate the value of a.

(*a*) Show that $\mathrm{E}(\bar{X}) = \dfrac{2a}{3}$ and $\mathrm{Var}(\bar{X}) = \dfrac{a^2}{18n}$.

(4)

The random variable $S = p\bar{X}$, where p is a constant, is an unbiased estimator of a.

(*b*) Write down the value of p and find Var(S).

(2)

Felix suggests using the statistic $M = \max\{X_1, X_2, X_3, \ldots, X_n\}$ as an estimator of a.

He calculates $\mathrm{E}(M) = \dfrac{2n}{2n+1}a$ and $\mathrm{Var}(M) = \dfrac{n}{(n+1)(2n+1)^2}a^2$.

(*c*) State, giving your reasons, whether or not M is a consistent estimator of a.

(3)

The random variable $T = qM$, where q is a constant, is an unbiased estimator of a.

(*d*) Write down, in terms of n, the value of q and find Var(T).

(3)

(*e*) State, giving your reasons, which of S or T you would recommend Emily use as an estimator of a.

(3)

Emily took a sample of 5 values of X and obtained the following:

5.3 4.3 5.7 7.8 6.9

(*f*) Calculate the estimate of a using your recommended estimator from part (*e*).

(2)

(*g*) Find the standard error of your estimate, giving your answer to 2 decimal places.

(2)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

6686/01
Edexcel GCE

Statistics S4

Advanced/Advanced Subsidiary

Thursday 12 June 2014 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Mathematical Formulae (Pink)	Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer for each question in the space following the question.
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.
The marks for 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 24 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 not gain full credit.

P43151A

1. A production line is designed to fill bottles with oil. The amount of oil placed in a bottle is normally distributed and the mean is set to 100 ml.

The amount of oil, x ml, in each of 8 randomly selected bottles is recorded, and the following statistics are obtained.

$$\bar{x} = 92.875 \qquad s = 8.3055$$

Malcolm believes that the mean amount of oil placed in a bottle is less than 100 ml.

Stating your hypotheses clearly, test, at the 5% significance level, whether or not Malcolm's belief is supported.

(5)

---

2. (*a*) Define

(i) a Type I error,

(ii) a Type II error.

(2)

Rolls of material, manufactured by a machine, contain defects at a mean rate of 6 per roll.

The machine is modified. A single roll is selected at random and a test is carried out to see whether or not the mean number of defects per roll has decreased. The significance level is chosen to be as close as possible to 5%.

(*b*) Calculate the probability of a Type I error for this test.

(3)

(*c*) Given that the true mean number of defects per roll of material made by the machine is now 4, calculate the probability of a Type II error.

(2)

---

3. A large number of chicks were fed a special diet for 10 days. A random sample of 9 of these chicks is taken and the weight gained, x grams, by each chick is recorded. The results are summarised below.

$$\sum x = 181 \qquad \sum x^2 = 3913$$

You may assume that the weights gained by the chicks are normally distributed.

Calculate a 95% confidence interval for

(*a*) (i) the mean of the weights gained by the chicks,

(ii) the variance of the weights gained by the chicks.

(10)

A chick which gains less than 16 g has to be given extra feed.

(*b*) Using appropriate confidence limits from part (*a*), find the lowest estimate of the proportion of chicks that need extra feed.

(4)

---

4. A random sample of 8 people were given a new drug designed to help people sleep.

In a two-week period the drug was given for one week and a placebo (a tablet that contained no drug) was given for one week.

In the first week 4 people, selected at random, were given the drug and the other 4 people were given the placebo. Those who were given the drug in the first week were given the placebo in the second week. Those who were given the placebo in the first week were given the drug in the second week.

The mean numbers of hours of sleep per night for each of the people are shown in the table.

Person	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Hours of sleep with drug	10.8	7.2	8.7	6.8	9.4	10.9	11.1	7.6
Hours of sleep with placebo	10.0	6.5	9.0	5.6	8.7	8.0	9.8	6.8

(*a*) State one assumption that needs to be made in order to carry out a paired t-test.

(1)

(*b*) Stating your hypotheses clearly, test, at the 1% level of significance, whether or not the drug increases the mean number of hours of sleep per night by more than 10 minutes. State the critical value for this test.

(8)

---

5. A statistician believes a coin is biased and the probability, p, of getting a head when the coin is tossed is less than 0.5.

The statistician decides to test this by tossing the coin 10 times and recording the number, X, of heads. He sets up the hypotheses $H_0 : p = 0.5$ and $H_1 : p < 0.5$ and rejects the null hypothesis if $x < 3$.

(*a*) Find the size of the test.

(1)

(*b*) Show that the power function of this test is

$$(1 - p)^8 (36p^2 + 8p + 1)$$

(3)

Table 1 gives values, to 2 decimal places, of the power function for the statistician's test.

p	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
Power	0.93	0.82	r	0.53	0.38	0.26	s	0.10

Table 1

(*c*) Calculate the value of r and the value of s.

(2)

Question 5 parts (*d*) and (*e*) continue on page 5

Question 5 continued

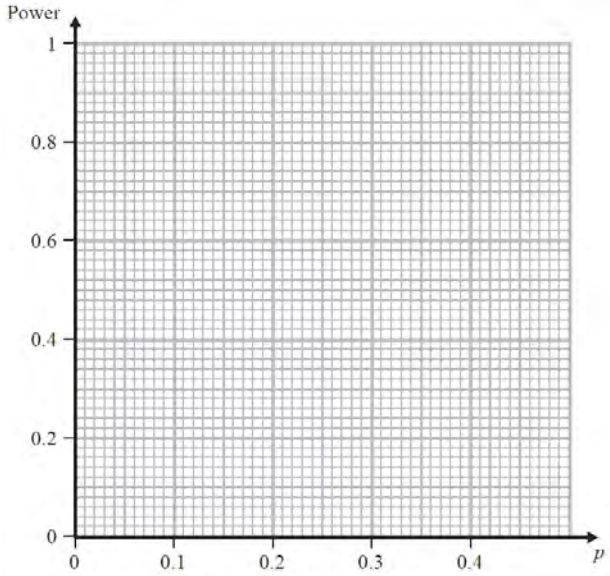
For your convenience Table 1 is repeated here.

p	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
Power	0.93	0.82	r	0.53	0.38	0.26	s	0.10

Table 1

(*d*) On the axes below draw the graph of the power function for the statistician's test.
(2)

(*e*) Find the range of values of p for which the probability of accepting the coin as unbiased, when in fact it is biased, is less than or equal to 0.4.
(3)



6. (*a*) Explain what is meant by the sampling distribution of an estimator T of the population parameter θ.
(1)

(*b*) Explain what you understand by the statement that T is a biased estimator of θ.
(1)

A population has mean μ and variance σ^2.

A random sample $X_1, X_2, \ldots, X_{10}$ is taken from this population.

(*c*) Calculate the bias of each of the following estimators of μ.

$$\hat{\mu}_1 = \frac{X_3 + X_5 + X_7}{3}$$

$$\hat{\mu}_2 = \frac{5X_1 + 2X_2 + X_9}{6}$$

$$\hat{\mu}_3 = \frac{3X_{10} - X_1}{3}$$

(4)

(*d*) Find the variance of each of these three estimators.
(6)

(*e*) State, giving a reason, which of these three estimators for μ is

(i) the best estimator,

(ii) the worst estimator.

(3)

7. Two groups of students take the same examination.

A random sample of students is taken from each of the groups.

The marks of the 9 students from Group 1 are as follows

30 29 35 27 23 33 33 35 28

The marks, x, of the 7 students from Group 2 gave the following statistics

$$\bar{x} = 31.29 \qquad s^2 = 12.9$$

A test is to be carried out to see whether or not there is a difference between the mean marks of the two groups of students.

You may assume that the samples are taken from normally distributed populations and that they are independent.

(*a*) State **one** other assumption that must be made in order to apply this test and show that this assumption is reasonable by testing it at a 10% level of significance. State your hypotheses clearly.
(7)

(*b*) Stating your hypotheses clearly, test, using a significance level of 5%, whether or not there is a difference between the mean marks of the two groups of students.
(7)

TOTAL FOR PAPER: 75 MARKS

END

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.

N10688

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

$$A_1 = X_1X_2 \quad \text{and} \quad A_2 = \left(\frac{X_1 + X_2}{2}\right)^2.$$

[You may assume that if X_1 and X_2 are independent random variables then

$$E(X_1X_2) = E(X_1)E(X_2)]$$

(*a*) Find $E(A_1)$ and show that $E(A_2) = \mu^2 + \frac{\sigma^2}{2}$.
(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 $E(\overline{X}^2) = \mu^2 + \frac{\sigma^2}{n}$ and $\text{Var}(\overline{X}^2) = \frac{2\sigma^4}{n^2} + \frac{4\sigma^2\mu^2}{n}$.

(*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