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.

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

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

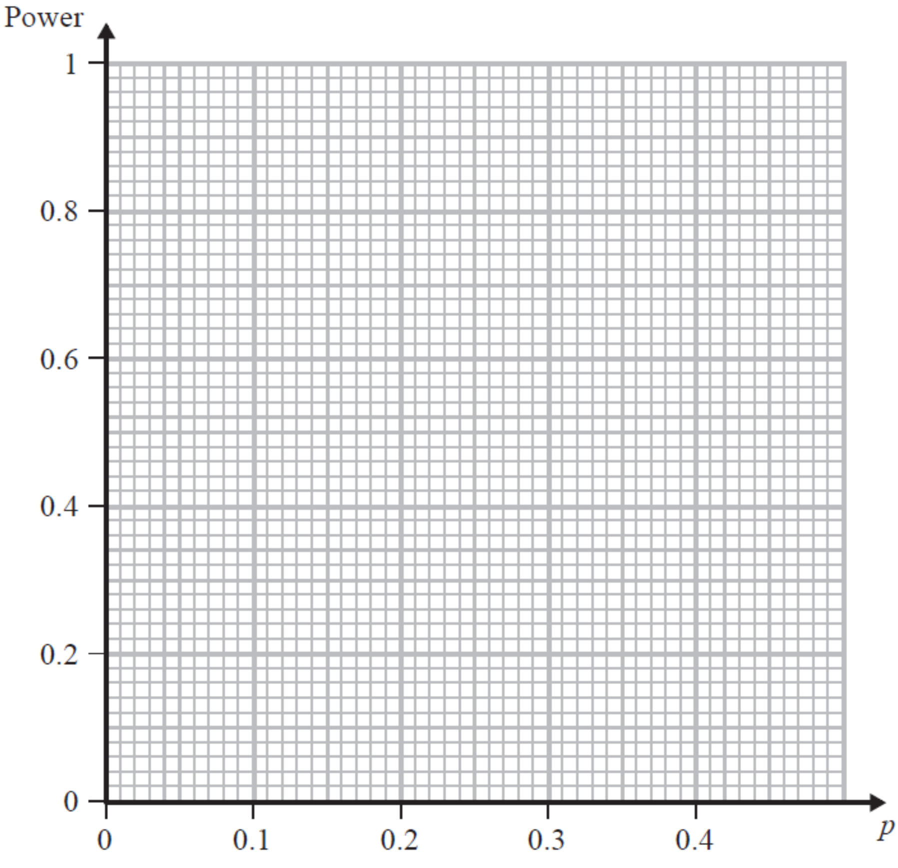
Table 1

(*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

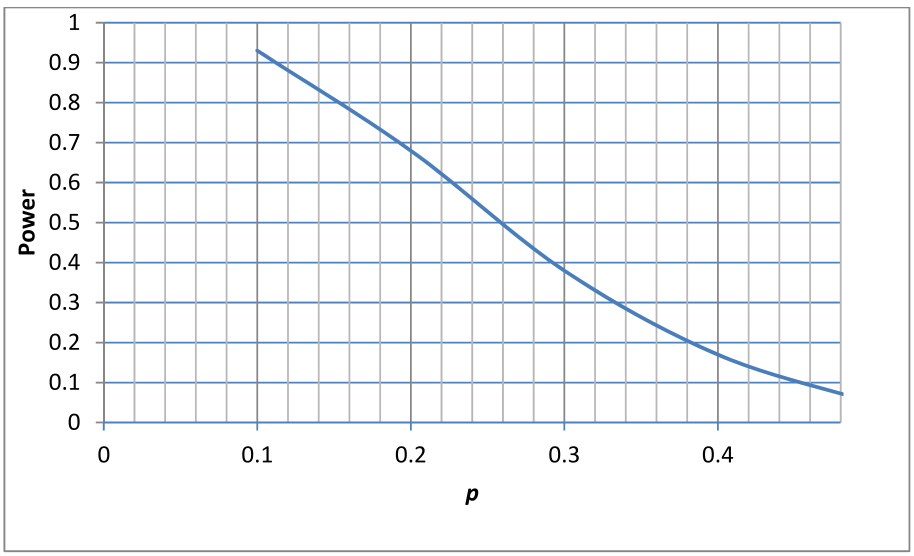
END

Question Number	Scheme	Marks
1.	$H_0 : \mu = 100 \quad H_1 : \mu < 100$	B1
	$t = \frac{\lvert\bar{x}-\mu\rvert}{s/\sqrt{n}} = \frac{\lvert 92.875-100\rvert}{8.3055/\sqrt{8}} = 2.4264\ldots$ or $\frac{c-100}{8.3055/\sqrt{8}} = -1.895 \therefore$ CR $c < 94.435$	M1A1
	$t_7(5\%) = \pm\ 1.895$	B1
	There is evidence to reject H_0. Malcolm's belief is supported or there is evidence that the amount of oil placed in bottles is less than 100mm	A1ft
		(5)
	Notes	
	B1 both hypotheses	
	M1 either $\frac{\lvert 92.875-100\rvert}{8.3055/\sqrt{8}}$ or $p = 0.0228$ or $\frac{c-100}{8.3055/\sqrt{8}} = -(\text{a } t \text{ value})$	
	A1 awrt 2.43 or awrt 94.4 or awrt 0.0228	
	B1 ± 1.895 or $0.0228 < 0.05$ (must have correct comparison for hypotheses) A1ft Do Not allow contradictions	

Question Number	Scheme	Marks
2(a)(i)	Type I - H_0 rejected when it is true	B1
(ii)	Type II - H_0 is accepted when it is false	B1
		(2)
(b)	$P(X< c \mid \lambda = 6) \approx 0.05$	M1
	$P(X \leq 2) = 0.0620$	
	$P(X \leq 1) = 0.0174$	
	Critical region = $X \leq 2$	A1
	P(Type 1 error) = $P(X \leq 2 \mid \lambda = 6) = 0.062$	A1cao
		(3)
(c)	P(Type 2 error) = $P(X \geq 3 \mid \lambda = 4)$	M1
	$= 1 - 0.2381$	
	$= 0.7619$	A1
		(2)
	Notes	
(b)	M1 use of Po(6) A1 correct CR. May be implied by correct probability. Alow if written as part of a 2 tailed CR A1 awrt 0.062	
(c)	M1 using Po(4) and 1 - $P(X \leq 2)$,ft their CR in (b) if one tail A1awrt 0.762	

Question Number	Scheme	Marks
3(a) (i)	$\bar{x} = \frac{181}{9} = 20.111\ldots$	B1
	$s_x^{\ 2} = \left(\frac{3913 - 9\times\bar{x}^2}{8} = \right) 34.1111 \quad (s_x = 5.84)$	B1
	$t_8(0.025)$ cv = 2.306	B1
	95% CI for μ is $= 20.111 \pm 2.306 \times \frac{5.84}{\sqrt{9}}$	M1
	= (15.6, 24.6) awrt **<u>(15.6, 24.6)</u>**	A1, A1
(ii)	$\chi_8^{\ 2}(0.025) = 2.18(0)$, $\chi_8^{\ 2}(0.975) = 17.535$	B1B1
	95% CI for σ^2 is given by $2.180 < \frac{8s_x^{\ 2}}{\sigma^2} < 17.535$	M1
	So 95% CI for σ^2 is = **<u>awrt (15.6, 125)</u>**	A1
		(10)
(b)	Require P(X < 16) = $\text{P}\left(Z < \frac{16-\mu}{\sigma}\right)$ to be as small as possible OR $\frac{16-\mu}{\sigma}$ to be as large as possible but negative; **imply lowest σ and largest μ.**	M1
	$\text{P}(Z < \frac{16-24.6}{\sqrt{15.6}})$; = 1 – 0.9854 = **<u>0.0146 or 0.0147</u>**	M1A1ft;A1
		(4)
	Notes	
(a)(i)	1st M1 ' their $\bar{x}$ ' $\pm\ t\ value \times \frac{\text{'their } s\text{'}}{\sqrt{9}}$ 1st A1 awrt 15.6 2nd A1 awrt 24.6	
(ii)	2nd M1 $\chi^2 < \frac{8s^2}{\sigma^2} < \chi^2$ A1 awrt 15.6 and 125	
(b)	M1 Identify must use **lowest σ and largest μ** M1 standardising and finding correct area use either limit for μ and σ A1 ft their **lowest σ and largest μ** A1 awrt 0.0146 or 0.0147	

Question Number	Scheme	Marks
4(a)	The differences in the mean number of hours sleep are normally distributed	B1
		(1)
(b)	Differences are 0.8, 0.7, -0.3, 1.2, 0.7, 2.9, 1.3, 0.8	M1
	$\bar{d} = = \frac{8.1}{8} = 1.0125$	M1
	$s_d = \sqrt{\frac{13.89 - 8\times 1.0125^2}{7}} = 0.901\ldots$ both $\bar{d}$ and s	M1
	$H_0: \mu_D = 1/6 \quad H_1: \mu_D > 1/6$	B1
	$t = \frac{1.0125 - 1/6}{0.901/\sqrt{8}} =$ awrt 2.65 or $\frac{c - 1/6}{0.901/\sqrt{8}} = 2.988 \therefore$ CR $c >$ awrt 1.12	M1A1
	$t_7(1\%) = 2.998$ (or prob. = awrt 0.0164)	B1
	There is insufficient evidence to suggest the drug increases the mean number of hours slept by more than 10 minutes.	A1ft
		(8)
	Notes	
(a)	B1 for a comment that mentions "differences" and "normal" distribution	
(b)	1st M1 for attempting the *d*s 2nd M1 for attempting $\bar{d}$ 1st M1 for s_d or s_d^2 1st B1 for both hypotheses correct in terms of μ or μ_d .(allow a defined symbol) Do not allow 10 instead of 1/6(awrt 0.167) unless working in minutes throughout 3rd M1 for attempting the correct test statistic $\frac{\bar{d} - 1/6}{s_d/\sqrt{8}}$ or p = awrt 0.016 or $\frac{c - 1/6}{0.901/\sqrt{8}} = t$ value 2nd A1 awrt 2.65 /2.655 or awrt 1.12 or awrt 0.016 2nd B1 2.988 or 0.0164 3rd A1ft for a correct comment in context based on their test statistic and their cv. Do not allow contradictions.	

Question Number	Scheme	Marks
5 (a)	$X \sim B(10, 0.5)$ Size = P(reject $H_0 \mid p = 0.5$) $= P(X < 3 \mid p = 0.5)$ $= 0.0547$	 B1 (1)
(b)	Power = $P(X = 2) + P(X = 1) + P(X = 0)$ $= 45p^2(1 - p)^8 + 10p(1 - p)^9 + (1 - p)^{10}$ $= (1 - p)^8(45p^2 + 10p(1 - p) + (1 - p)^2)$ $= (1 - p)^8(36p^2 + 8p + 1)$	M1 A1 A1cso (3)
(c)	$r = 0.68$ $s = 0.17$	B1 B1 (2)
(d)		B1 points B1 curve (2)
(e)	P(Type II error) $\leq$ 0.4 $1 - \text{power} \leq 0.4$ Power ≥ 0.6 $p < 0.23$	M1 A1 A1 (3)
	Notes	
(b) (c) (e)	M1 for a correct expression/selection of probabilities A1 for a fully correct expression SC B1 B0 both correct but not given to 2 dp M1 may be implied by Power ≥ 0.6 or correct value or by correct answer A1 may be implied by correct answer A1 allow number between 0.22 and 0.23 inclusive and either $<$ or $\leq$	

Question Number	Scheme	Marks
6(a)	It is the probability distribution of *T*.	B1 (1)
(b)	An estimator is biased if $\mathrm{E}(T) \neq \theta$	B1 (1)
(c)	$\mathrm{E}(\hat{\mu}_1) = \frac{\mathrm{E}(X_3)+\mathrm{E}(X_5)+\mathrm{E}(X_7)}{3} = \frac{\mu+\mu+\mu}{3} = \mu \quad \therefore \quad \text{Bias} = 0$	M1A1
	$\mathrm{E}(\hat{\mu}_2) = \frac{5E(X_1)+2E(X_2)+E(X_9)}{6} = \frac{5\mu+2\mu+\mu}{6} = \frac{4\mu}{3} \therefore \text{ Bias} = \frac{\mu}{3}$	A1
	$\mathrm{E}(\hat{\mu}_3) = \frac{3\mathrm{E}(X_{10}) - \mathrm{E}(X_1)}{3} = \frac{3\mu-\mu}{3} = \frac{2\mu}{3} \quad \therefore \quad \text{Bias} = -\frac{\mu}{3}$	A1 (4)
(d)	$\mathrm{Var}(\hat{\mu}_1) = \frac{1}{9}(\mathrm{Var}(X_3) + \mathrm{Var}(X_5) + \mathrm{Var}(X_7))$	M1
	$= \frac{1}{9}(\sigma^2 + \sigma^2 + \sigma^2)$	
	$= \frac{\sigma^2}{3}$	A1
	$\mathrm{Var}(\hat{\mu}_2) = \frac{1}{36}(25\mathrm{Var}(X_1) + 4\mathrm{Var}(X_2) + \mathrm{Var}(X_9))$	M1
	$= \frac{1}{36}(25\sigma^2 + 4\sigma^2 + \sigma^2)$	
	$= \frac{5}{6}\sigma^2$	A1
	$\mathrm{Var}(\hat{\mu}_3) = \frac{1}{9}(9\mathrm{Var}(X_{10}) + \mathrm{Var}(X_1))$	M1
	$= \frac{1}{9}(9\sigma^2 + \sigma^2)$	
	$= \frac{10\sigma^2}{9}$	A1 (6)
(e)(i)	$\hat{\mu}_1$ is the best estimator. It has no bias	B1
(ii)	It has <u>same magnitude of bias as $\hat{\mu}_2$</u> but it has the <u>largest variance</u> $\hat{\mu}_3$ is the worst estimator.	B1ft B1dcao (3)
	Notes	
(c)	M1 finding $\mathrm{E}(\hat{\mu})$ A1 bias 0 A1 $\pm\frac{\mu}{3}$ A1 $\pm\frac{\mu}{3}$	
(d)	For method marks allow an incorrect variance, M1 squaring 9, M1 Squaring 5 and 2, M1 adding variances. Do not penalise same mistake twice.	
(e)(ii)	Must have idea that its bias is the same as another ($\hat{\mu}_2$) and state it has largest variance for first B1 . ft their values of Var. Second B1 dependent on first B1cao SC $\hat{\mu}_3$ because <u>largest variance</u> B1 B0	