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 $\bar{X} = \frac{X_1 + X_2}{2}$ 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

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

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

(4)

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

(*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)

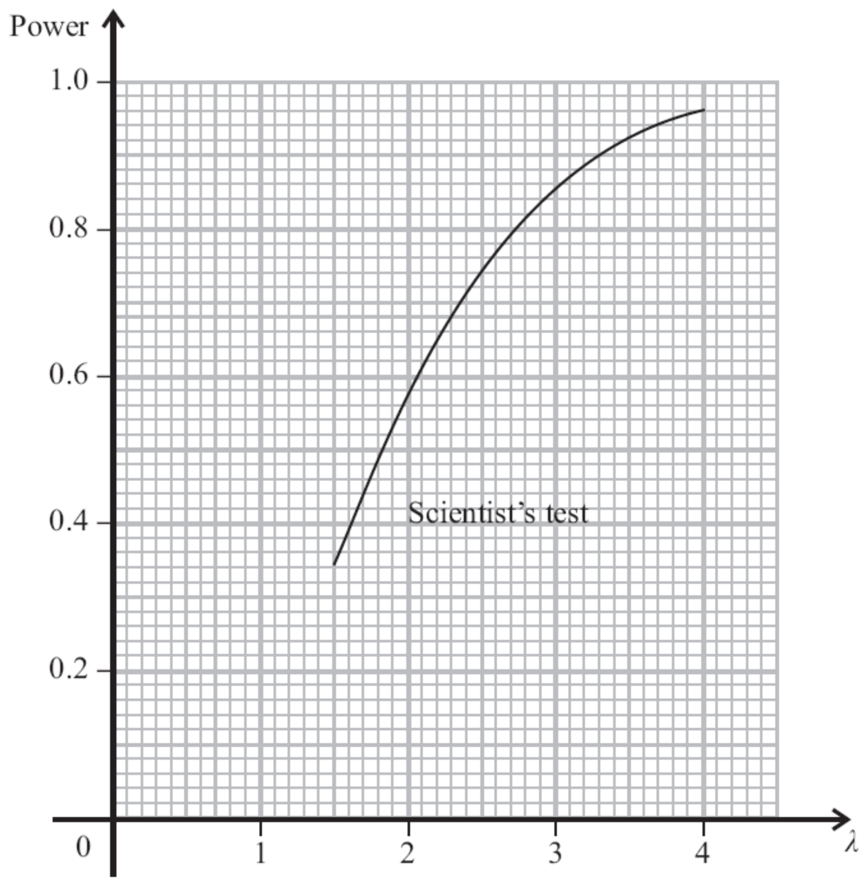


Figure 1

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

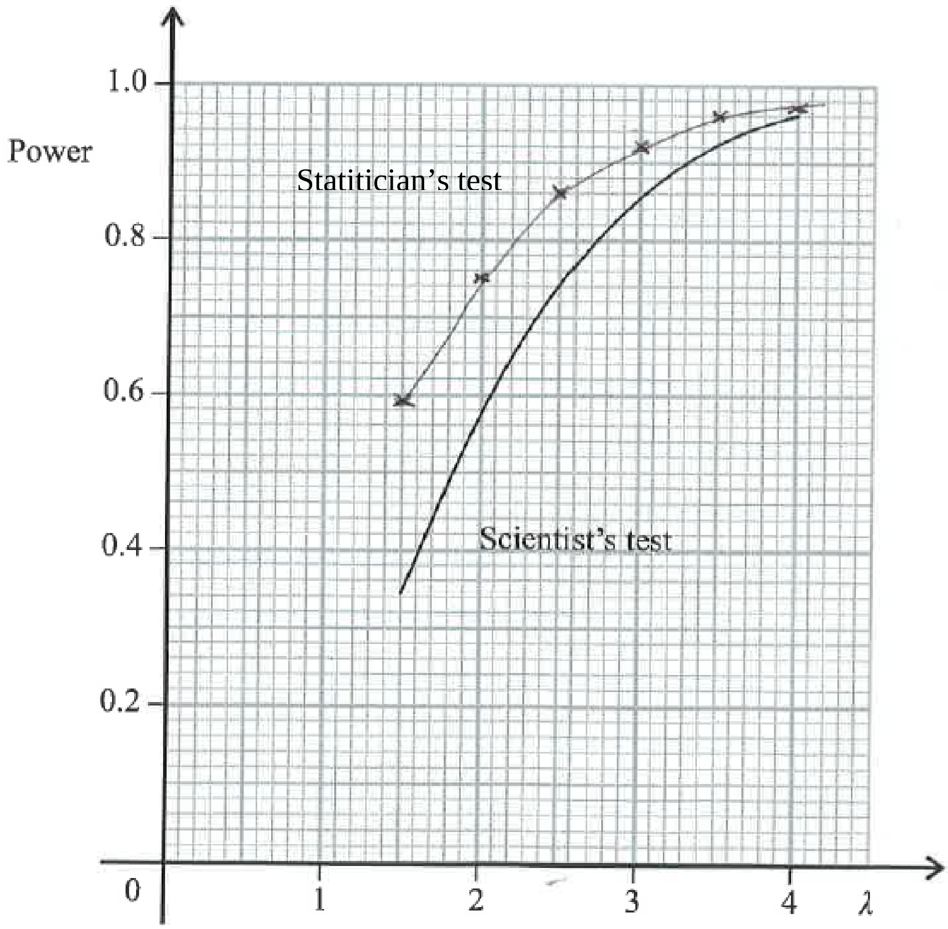
END

Question Number	Scheme	Marks
1.	$H_0: \sigma^2 = 2.4^2 \quad H_1: \sigma^2 \neq 2.4^2$	B1
	$s^2 = \dfrac{1414.08 - 10\times\left(\dfrac{113.4}{10}\right)^2}{9} = 14.236$	M1 A1
	$\chi^2 = \dfrac{9s^2}{\sigma^2} = \dfrac{9\times 14.236}{2.4^2} = 22.24375$	M1 A1
	Critical Value $\chi_9^2(0.025) = 19.023$	B1
	Significant result, there is evidence of a change in standard deviation or the data do not support George's belief	A1cso **(7)** **[7]**
	Notes	
	1st B1 Both hypotheses, must use σ. Allow $H_0: \sigma = 2.4 \quad H_1: \sigma \neq 2.4$ 1st M1 correct method used 1st A1 awrt 14.2 2nd M1 $\chi^2 = \dfrac{9\times\text{"}their\ s^2\text{"}}{2.4^2}$ 2nd A1 awrt 22.2 2nd B1 for critical value, this should be compatible with their alternative hypothesis (16.919 for one tail test) 3rd A1ft fully correct solution only	

Question Number	Scheme	Marks
2. (a)	d = Jan - June: -2, 1, -3, 2, -2, 3, 2, 2	M1
	$\bar{d} = 0.375, \quad \sum d^2 = 39 \Rightarrow s^2 = 5.4107...$ or s = 2.326…	M1, M1
	$t_7(0.025) = 2.365$	B1
	Confidence Interval: $0.375 \pm 2.365 \times \frac{2.326...}{\sqrt{8}}$	M1
	= **<u>(-1.57, 2.32)</u>** (o.e.)	A1,A1 **(7)**
(b)	$H_0: \mu_D = 0 \quad H_1: \mu_D \neq 0$	B1
	Comment that 0 is in the interval	M1
	Not sig, no evidence of a change in **mean time** to **assemble component**	A1ft **(3)** **[10]**
	Notes	
(a)	1st M1 for attempting differences 2nd M1 for attempting $\bar{d}$ 3rd M1 for attempting $s_d{}^2$, correct expression with their $\sum d^2$ and $\bar{d}$ or correct calculation (to 2 sf or better) 4th M1 for use of a correct CI formula, using a value for t and ft their values. 1st A1 for lower limit of -1.57 or -2.32 2nd A1 for corresponding upper limit	
S.C.	Allow A1A1 for (0, 2.32)	
(b)	B1 for both hypotheses using μ_D M1 for a comment about 0 being in (or out) of <u>their</u> interval A1 contextual conclusion – must include assemble components	
S.C.	If they have used difference in means test in part (a) to get the confidence interval then award the B1 for $H_0: \mu_x - \mu_y = 0 \quad H_1: \mu_x - \mu_y \neq 0$ or the correct hypotheses.	

Question Number	Scheme	Marks
3. (a)	$H_0 : \sigma^2{}_A = \sigma^2{}_B \quad H_1 : \sigma^2{}_A \neq \sigma^2{}_B$	B1
	$F = \frac{s_B{}^2}{s_A{}^2} = \frac{4.37^2}{4.24^2} = 1.0622...$	M1A1
	$F_{12,6}$ (0.01) = 7.72	B1
	Not sig, so no evidence of a difference in variances	A1ft **(5)**
(b)	$H_0 : \mu_A = \mu_B \quad H_1 : \mu_A < \mu_B$	B1
	$s_p{}^2 = \frac{6 \times 4.24^2 + 12 \times 4.37^2}{18}$ = 18.7238 or $s_p = 4.327...$	M1
	$t = \pm \frac{14.31 - 8.43}{s_p \sqrt{\frac{1}{7} + \frac{1}{13}}} = \pm\ 2.8985\ldots$ awrt 2.9	M1A1
	$t_{18}(0.01) = 2.552$	B1
	sig, there is evidence to support **archaeologist's** claim <u>or</u> there is evidence that bricks for site *B* have higher **mean compression** strength than those from site *A*.	A1ft **(6)**
(c)	The test in (b) requires $\sigma^2{}_A = \sigma^2{}_B$ and the test in part (a) shows that this is a reasonable assumption. (o.e.)	B1 **(1)**
		[12]
	Notes	
(a)	M1 for use of a correct formula Allow $F = \frac{s_A{}^2}{s_B{}^2} = \frac{4.24^2}{4.37^2} = 0.941...$ with 0.1295..	
(b)	B1 if *A* and *B* not used it must be clear which is *A* and which is *B* 1st M1 for attempt to calculate s_p or $s^2{}_p$ 2nd M1 for attempt correct test statistic 2nd A1 ft need **archaeologist's or compression**	
(c)	Need to refer to 'allows us to assume variances the same' and this is needed in for test. oe	

Question Number	Scheme	Marks
4. (a)	$\text{E}(X)=\mu=\frac{2a-a}{2}=\frac{a}{2}$; $\text{E}(\overline{X})=\mu=\frac{a}{2}$ so biased estimator for a	M1;A1
	Bias = $\frac{a}{2}-a=-\frac{a}{2}$	B1(accept $\pm$) **(3)**
(b)	$k=2$	B1 **(1)**
(c)	$\text{Var}(X)=\sigma^2=\frac{(2a--a)^2}{12}=\frac{9a^2}{12}=\frac{3a^2}{4}$; $\text{Var}(\overline{X})=\frac{\sigma^2}{2}$	B1;B1
	Var(Y) = $k^2\text{Var}(\overline{X})=4,\times\frac{\sigma^2}{2}=4\times\frac{3a^2}{4\times2}=\frac{3}{2}a^2$	M1,A1 **(4)**
(d)	E(M) = $\int\frac{2x(x+a)}{9a^2}\text{ d}x=\left[\frac{2x^3}{27a^2}+\frac{ax^2}{9a^2}\right]_{-a}^{2a},=\left(\frac{16a}{27}+\frac{4a}{9}\right)-\left(-\frac{2a}{27}+\frac{a}{9}\right)$ [$=a$]	M1A1,M1d
	So E(M) = a and therefore M is an unbiased estimator for a	A1cso **(4)**
(e)	Var(M) = $\frac{3}{2}a^2-a^2=\frac{1}{2}a^2$	B1 **(1)**
(f)	Var(M) < Var(Y) , so M is the better estimator of a	M1, A1 **(2)**
(g)	Maximum value = **<u>5</u>**	B1ft **(1)** **[16]**
	Notes	
(a)	M1 for use of formula or integration or symmetry to find E(X)	
(c)	1st B1 for use of formula for variance 2nd B1 for use of $\frac{\sigma^2}{n}$ formula M1 for $k^2\text{Var}(\overline{X})$ and ft their k	
(d)	1st M1 for attempt at correct integration of correct expression 1st A1 for correct integration 2nd M1d dependent on previous M, for attempting to use correct limits 2nd A1 need statement that M is therefore unbiased	
(f)	M1 for comparison of their Var(Y) and their Var(M)	
(g)	B1ft for calculation of their estimate based on their choice in (f). If they choose Y answer is 4 (or twice their k)	

Question Number	Scheme	Marks
5. (a)	Y = no. of organisms in 20 ml. Y~Po(2λ) Size = P($Y \geq 4$| Y~Po(2)) , = 1 - P($Y \leq 3$) = 1- 0.8571 = **<u>0.1429</u>**	M1, A1 **(2)**
(b)	P(Type II error) = 1 - P($Y \geq 4$| Y~Po(5)) , = P($Y \leq 3$) = **<u>0.2650</u>**	M1, A1 **(2)**
(c)	X = no. of organisms in 10 ml. X ~ Po(λ) Power = P($X \geq 2$) + P(X =1) × P($X \geq 2$) = P($X \geq 2$) [1+ P(X = 1)] = $\left[1-e^{-\lambda}(1+\lambda)\right]\times\left[1+\lambda e^{-\lambda}\right]$ = $1-e^{-\lambda}-\lambda e^{-\lambda}+\lambda e^{-\lambda}-\lambda(1+\lambda)e^{-2\lambda}=1-e^{-\lambda}-\lambda(1+\lambda)e^{-2\lambda}$	 M1 M1A1 A1cso **(4)**
(d)	r = 0.92	B1 **(1)**
(e)	See Graph paper  	B1B1 **(2)**

Question Number	Scheme	Marks
(f)	Expected time for statistician's test: $30\times P(X=1)+15\times\left[1-P(X=1)\right]$	M1

	$= 30\lambda e^{-\lambda} + 15\left(1-\lambda e^{-\lambda}\right) = 15\left(1+\lambda e^{-\lambda}\right)$	A1
	slower if: $15\left(1+\lambda e^{-\lambda}\right) > 20, \quad \Rightarrow \lambda e^{-\lambda} > \frac{1}{3}$	M1,A1cso **(4)**
(g)	$\lambda e^{-\lambda}$ with $\lambda = 1$ is 0.36…, with $\lambda = 2$ is 0.27…so second(statisticians) test is slower if $\lambda = 1$ but faster for $\lambda = 2$. Second test is more powerful for all λ	B1
	Choose second test - more powerful and faster for $\lambda \geq 2$	B1 **(2)** **[17]**
	Notes	
(a)	M1 for correct expression for size using Po(2)	
(b)	M1 for correct expression using Po(5)	
(c)	1st M1 for a correct expression in terms of probabilities Alternate answer 1 – [P(X = 0) + P(X = 1) x P($X \leq 1$)] 2nd M1 for an attempt at a correct equation in λ 1st A1 for a correct expression in λ	
(e)	1st B1 points 2nd B1 curve (or straight lines)	
(f)	1st M1 for an attempt to calculate expected time Alternate method 15 + 15x P(X = 1) 1st A1 for a correct expression in terms of λ 2nd M1 for attempt at correct inequality	
(g)	1st B1 for a comment about power & timings 2nd B1 for selecting second test	

Question Number	Scheme	Marks
6. (a)	$H_0 : \sigma^2{}_A = \sigma^2{}_B \qquad H_1 : \sigma^2{}_A \neq \sigma^2{}_B$	B1
	$s_A{}^2 = (0.25)^2 = 0.0625 \qquad s_B{}^2 = (0.178885...)^2 = 0.032$	B1B1

	$F = \frac{0.0625}{0.032} = 1.953...$	M1A1
	Critical Value: $F_{3,5} = 5.41$	B1
	not sig, samples come from populations with common variance	A1cso **(7)**
(b)	$s_p^{\ 2} = \frac{3\times 0.25^2 + 5\times 0.032}{8} = 0.04343... = (0.0284...)^2$	M1A1
	Use $\frac{8s_p^{\ 2}}{\sigma^2} \sim \chi_8^{\ 2}$	M1
	$1.344 < \frac{8\times 0.0434...}{\sigma^2} < 21.955$	B1,B1
	99% confidence interval is **<u>(0.0158, 0.259)</u>**	A1 **(6)** **[13]**