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 $\dfrac{k}{2}$ and variance $\dfrac{k^2}{12}$.

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

(*a*) Show that $\dfrac{2}{3}X_1 + \dfrac{1}{2}X_2 + \dfrac{5}{6}X_3$ 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 $\text{Var}(\hat{k}) = (a^2 - 2a + 2)\dfrac{k^2}{6}$

(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

Question Number	Scheme	Marks
1	$H_0: \mu = 5$; $H_1: \mu < 5$ both	B1
	CR: $t_9(0.01) > 2.821$	B1
	$\bar{x} = 4.91$	B1
	$s^2 = \frac{1}{9}\left(241.2 - \frac{49.1^2}{10}\right) = 0.0132222$ s= awrt 0.115	M1 A1
	$t = \frac{\lvert 4.91 - 5\rvert}{\frac{\sqrt{0.013222}}{\sqrt{10}}} = \pm 2.475$ 2.47 – 2.48	M1 A1
	Since 2.475 is not in the critical region there is insufficient evidence to reject H_0 and conclude that the mean diameter of the bolts is not less than (not equal to) 5mm.	A1ft **[8]**
2 (a)	The differences are normally distributed	B1 (1)
(b)	The data is collected in pairs or small sample size and variance unknown or samples not independent	B1 (1)
(c)	d: 2.5, 1.6, 1.6, –1.9, –0.6, 4.5 at least 2 correct	M1
	$(\Sigma d = 7.7, \Sigma d^2 = 35.59)$ $\bar{d} = \pm 1.2833$, sd = 2.2675. (Var = 5.141)	A1, A1
	$H_0: \mu_d = 0$, $H_1: \mu_d > 0$ ($H_1: \mu_d < 0$ if d - 2.5, -1.6, -1.6 etc) both depend on their d's	B1
	$t = \frac{\pm 1.2833\sqrt{6}}{2.2675} = \pm 1.386\ldots\ldots$ formula and substitution, 1.38 – 1.39	M1, A1
	Critical value $t_5(5\%) = 2.015$ (1 tail)	B1
	Not significant. Insufficient evidence to support that the device reduces CO_2 emissions.	A1 ft (8)
(d)	The idea that the device reduces $C0_2$ emissions has been rejected when in fact it does reduce emissions. OR Concluding that the device does not reduce emissions when in fact it does (if not in context can get B1 only)	B1 B1 (2) **[12]**
	(b) Allow because the same car has been used (c) awrt ± 1.28 , 2.27	

Question Number	Scheme	Marks
3 (a)	Size is the probability of H_0 being rejected when it is in fact true. or P(reject H_0/ H_0 is true) oe	B1 (1)
(b)	The power of the test is the probability of rejecting H_0 when H_1 is true. or P(rejecting H_0/H_1 is true) / P(rejecting H_0/H_0 is false) oe	B1 (1)
(c)	$X \sim B(12,0.5)$	B1
	$P(X \leq 2) = 0.0193$	M1
	$P(X \geq 10) = 0.0193$	
	$\therefore$ critical region is $\{X \leq 2 \cup X \geq 10\}$	A1A1 (4)
(d)(i)	P(Type II error) = $P(3 \leq X \leq 9 \mid p = 0.4)$	M1
	$= P(X \leq 9) - P(X \leq 2)$	M1dep
	$= 0.9972 - 0.0834$	
	$= 0.9138$	A1
(ii)	Power $= 1 - 0.9138$ $= 0.0862$	B1 ft (4)
(e)	Increase the sample size	B1
	Increase the significance level/larger critical region	B1 (2)
		[12]

Question Number	Scheme	Marks
4 (a)	$H_0 : \sigma_A^2 = \sigma_B^2,\ H_1 : \sigma_A^2 \neq \sigma_B^2$	B1
	critical values $F_{12,8} = 3.28$ and $\frac{1}{F_{8,12}} = 0.35$	B1
	$\frac{s_B^2}{s_A^2} = 2.40 \quad \left(\frac{s_A^2}{s_B^2} = 0.416\right)$	M1A1
	Since 2.40 (0.416) is not in the critical region we accept H_0 and conclude there is no evidence that the two variances are different.	A1ft (5)
(b)	$S_p^2 = \frac{8\times1.02+12\times2.45}{20}$	M1
	$= 1.878$	A1
	$(27.94-25.54) \pm 2.086\times\sqrt{1.878}\times\sqrt{\frac{1}{9}+\frac{1}{13}}$	B1M1 A1ft
	$(1.16,\ 3.64)$	A1 A1 (7)
(c)	To calculate the confidence interval the variances need to be equal. In part (a) the test showed they are equal.	B1 B1 (2)
		[14]

Question Number	Scheme	Marks
5 (a)	95% confidence interval for μ is 2.145	B1
	$560 \pm t_{14}(2.5\%)\sqrt{\frac{25.2}{15}} = 560 \pm 2.145\sqrt{\frac{25.2}{15}} = (557.2, 562.8)$	M1 A1 A1
		(4)
(b)	95% confidence interval for σ^2 is	
	$5.629 < \frac{14\times 25.2}{\sigma^2} < 26.119$	B1, M1, B1
	$\sigma^2 < 62.675$ $\sigma^2 > 13.507$	
	$13.507 < \sigma^2 < 62.675$ awrt 13.5, 62.7	A1, A1
		(5)
(c)	Require $P(X > 565) = P\left(Z > \frac{565-\mu}{\sigma}\right)$ to be as large as possible OR	
	$\frac{565-\mu}{\sigma}$ to be as small as possible; both imply highest σ and μ.	M1
	$\frac{565-562.8}{\sqrt{62.675}} = 0.28$	M1A1
	$P(Z > 0.28) = 1 - 0.6103 = 0.3897$ awrt 0.39 – 0.40	M1 A1
		(5)
		[14]

Question Number	Scheme	Marks
6 (a)	$\mathrm{E}(\frac{2}{3}X_1 + \frac{1}{2}X_2 + \frac{5}{6}X_3) = \frac{2}{3} \times \frac{k}{2} + \frac{1}{2} \times \frac{k}{2} + \frac{5}{6} \times \frac{k}{2} = k$	M1 A1
	$\mathrm{E}(X_1 + X_2 + X_3) = k \Rightarrow$ unbiased	B1
		(3)
(b)	$\mathrm{E}(aX_1 + bX_2) = a\frac{k}{2} + b\frac{k}{2} = k$	M1
	$a + b = 2$	A1
	$\mathrm{Var}(aX_1 + bX_2) = a^2\frac{k^2}{12} + b^2\frac{k^2}{12}$	M1A1
	$= a^2\frac{k^2}{12} + (2-a)^2\frac{k^2}{12}$	M1
	$= (2a^2 - 4a + 4)\frac{k^2}{12}$	
	$= (a^2 - 2a + 2)\frac{k^2}{6}$ (*) since answer given	A1 cso
		(6)
(c)	Min value when $(2a-2)\frac{k^2}{6} = 0$ $\quad \frac{\mathrm{d}}{\mathrm{d}a}(\mathrm{Var}) = 0$, all correct, condone missing $\frac{k^2}{6}$	M1A1
	$\Rightarrow 2a - 2 = 0$ $a = 1, b = 1.$	A1A1
	$\frac{\mathrm{d}^2(\mathrm{Var})}{\mathrm{d}a^2} = \frac{2k^2}{6} > 0$ since $k^2 > 0$ therefore it is a minimum	M1
	min variance $= (1 - 2 + 2)\frac{k^2}{6}$ $= \frac{k^2}{6}$	B1
		(6)
		(15 marks)