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,

(*b*) 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

Question Number	Scheme	Marks
1. (a)	$H_0: \mu_d = 0$, $H_1: \mu_d > 0$ (or $H_1: \mu_d < 0$)	B1
	where μ_d is the (population) mean difference :- BP sitting down – BP standing. (BP standing – BP sitting down)	
	Assume the differences are normally distributed	B1
		(2)
(b)	d: 4, -1, 6, 6, 3, -2, 9, -1, 4, 7, -11, 7	M1
	($\Sigma d = 31$, $\Sigma d^2 = 419$) $\bar{d} = \pm 2.5833$; sd = 5.55073. (or Var = 30.8106)	A1; A1
	$t = \dfrac{\pm 2.5833\sqrt{12}}{5.55073} = \pm 1.612\ldots$ Formula and substitution, 1.61	M1, A1
	Critical value $t_{11}(1\%) = 2.718$(1 tail)	B1
	Not significant. Insufficient evidence to support that the blood pressure of a person sitting down is more than the blood pressure of a person after standing up.	A1 ft
		(7) **9 marks**

Question Number	Scheme	Marks
2. (a)	$S_F^2 = \frac{1}{5}\{2308.01 - 6\times 19.6^2\} = 0.61$	B1
	$S_M^2 = \frac{1}{11}\{2262.57 - 12\times 13.7^2\} = 0.93545..$	B1
	H_0: $\mu_F = \mu_M + 5$; H_1: $\mu_F \neq \mu_M + 5$ both	B1
	CR: $t_{16}(0.025) > 2.120$ 2.12	B1
	$S_p^2 = \frac{5\times 0.61 + 11\times 0.93545...}{16} = 0.83375$	M1 A1
	$t = \frac{19.6 - 13.7 - 5}{\sqrt{0.83375\left(\frac{1}{6}+\frac{1}{12}\right)}} = 1.971$	M1 A1ftA1
	Since 1.971 is not in the critical region we accept H_0 and conclude that the mean shell length of female turtles does exceed the shell length of male turtles by 5cm.(or Biologists claim is correct)	A1 ft
		(10)
(b)(i)	$-1.96 < \frac{\overline{X}_F - \overline{X}_M - 5}{\sqrt{\left(\frac{0.9}{6}+\frac{0.9}{12}\right)}} < 1.96$	B1 M1
	$4.07 < \overline{X}_F - \overline{X}_M < 5.93$	A1cso
(ii)	P(Type II error) = P($4.07 < \overline{X}_F - \overline{X}_M < 5.93$ | N(6, 0.225))	M1
	$= P(\frac{4.07-6}{\sqrt{0.225}} < z < \frac{5.93-6}{\sqrt{0.225}})$	M1
	= 0.44 awrt 0.44	A1
		(6) **16 marks**
3.	H_0: $\sigma_A^2 = \sigma_B^2$; H_1: $\sigma_A^2 \neq \sigma_B^2$	B1
	$S_A^2 / S_B^2 = \frac{225}{36} = 6.25 \quad \left(\frac{36}{225} = 0.16\right)$	M1A1
	CR: $F_{10,8} > 3.35 \quad \left(\frac{1}{F_{10,8}} = 0.299\right)$	B1
	Since 6.25 is in the critical region we can assume that the lengths of paving slabs sold by the builders merchant differ in variability.	A1ft
		(5) **5 marks**

Question Number	Scheme	Marks
4. (a)	$\bar{x} = 4.9$	B1
	$s = \sqrt{0.191..}$ (0.437...)	B1
	(NB: $\Sigma x = 49$; $\Sigma x^2 = 241.82$)	
(i)	95 % confidence interval is given by	
	$4.9 \pm 2.262 \times \sqrt{\dfrac{0.191..}{10}}$	M1A1ft B1
	i.e: (4.587…, 5.212 …)	A1 A1
(ii)	95% confidence interval is given by	
	$\dfrac{9\times 0.437...^2}{19.023} < \sigma^2 < \dfrac{9\times 0.437...^2}{2.7}$ use of $\dfrac{(n-1)s^2}{\chi^2_{n-1}}$	M1B1B1A1
	i.e. (0.0904, 0.63704)	A1 A1
		(13)
(b)	5 lies inside the confidence interval	B1ft
	0.49(0.7^2) lies inside the confidence interval	B1ft
	Yes it does meet the time requirement	B1 ft
		(3)
		16 marks

Question Number	Scheme	Marks
5. (a)	$H_0: \sigma^2 = 36$; $H_1: \sigma^2 > 36$	B1
	$\upsilon = 24$, X^2_{24} (0.05) = 36.415	B1
	$\dfrac{(n-1)S^2}{\sigma^2} = \dfrac{24 \times 55}{36} = 36.67$	M1 A1
	Since 36.67 > 36.415 there is sufficient evidence to reject H_0.	A1 ft
	There is evidence to suggest that the variance is greater than 36.	A1 ft
		(6)
(b)		
	$H_0 : \mu = 450$ $H_1 : \mu > 450$	B1
	$t_{24} = 1.711$	B1
	$t = \pm \dfrac{455 - 450}{\sqrt{\dfrac{55}{25}}} = \pm 3.37\ldots$	M1 A1
	Significant; The **mean weight** of chocolates is **greater than 450,** Or **μ is more than 450**	A1ft; A1ft **(6)**
(c)	The **weights** are normally distributed	B1 **(1)** **13 marks**

Question Number	Scheme	Marks
6. (a)(i)	$\mathrm{E}(\hat{p}_1) = \mathrm{E}\left(\frac{X}{n}\right)$	
	$= \frac{1}{n}\mathrm{E}(X)$	
	$= \frac{1}{n} \times np$	M1
	$= p$ unbiased	A1cso
(ii)	$\mathrm{Var}(\hat{p}_1) = \mathrm{Var}\left(\frac{X}{n}\right)$	
	$= \frac{1}{n^2}\mathrm{Var}(X)$	M1
	$= \frac{1}{n^2} \times np(1-p)$	
	$= \frac{p(1-p)}{n}$	A1
		(4)
b(i)	$\mathrm{E}(\hat{p}_3) = 3a\,\mathrm{E}(\hat{p}_1) + 2a\,\mathrm{E}(\hat{p}_2)$	M1
	$= 3ap + 2ap$	
	$= 5ap$	
	$5ap = p$	M1
	$a = \frac{1}{5}$	A1
(ii)	$\mathrm{Var}(\hat{p}_3) = \frac{9}{25}\mathrm{Var}(\hat{p}_1) + \frac{4}{25}\mathrm{Var}(\hat{p}_2)$	M1
	$= \frac{9p(1-p)}{25n} + \frac{4p(1-p)}{25m}$	M1d
	$= \frac{p(1-p)}{25}\left(\frac{9}{n} + \frac{4}{m}\right)$	A1
		(6)

Question Number	Scheme	Marks
6. (c)	$\frac{p(1-p)}{25}\left(\frac{9}{n}+\frac{4}{m}\right) < \frac{p(1-p)}{n}$	M1
	$9m + 4n < 25m$	
	$4n < 16m$	
	$\frac{n}{m} < 4$	
	$\frac{p(1-p)}{25}\left(\frac{9}{n}+\frac{4}{m}\right) < \frac{p(1-p)}{m}$	M1
	$9m + 4n < 25n$	
	$9m < 21n$	
	$\frac{9}{21} < \frac{n}{m}$ or $\frac{3}{7} < \frac{n}{m}$	
	$\frac{3}{7} < \frac{n}{m} < 4$	A1
		(3)
(d)	$\text{Var}(\hat{p}_1) = 0.05\,p(1-p)$ $\text{Var}(\hat{p}_2) = 0.0167\,p(1-p)$ $\text{Var}(\hat{p}_3) = 0.0207\,p(1-p)$	M1
	Or since $\frac{1}{3}$ is not in the range $\frac{9}{21} < \frac{n}{m} < 4$ $\text{Var}(\hat{p}_3)$ is not the smallest variance. $\text{Var}(\hat{p}_1) = 0.05\,p(1-p)$ $\text{Var}(\hat{p}_2) = 0.0167\,p(1-p)$	
	Therefore $\hat{p}_2$; is the best estimator as it has the smallest variance	A1ft; A1ft
		(3)
		16 marks