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.

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.
©2007 Edexcel Limited

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 $\mathrm{N}(\mu, \sigma^2)$. 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

June 2007
6686 Statistics S4
Mark Scheme

Question Number	Scheme	Marks
1. a	d: 14 2 18 25 0 -8 4 4 12 20	M1
	$\bar{d} = \pm 9.1$ sd = $\sqrt{106.7} = 10.332..$ $(\sum d = 91,$ $\sum x^2 = 1789)$	A1 A1
	$H_0 : \mu_d = 0$ $H_1 : \mu_d \neq 0$	B1
	$t = \pm \frac{9.1\sqrt{10}}{10.332} = \pm\ 2.785$ awrt ± 2.78 or 2.79	M1 A1
	Critical value $t_9 = \pm 1.833$	B1
	Significant. There is a difference between blood pressure measured by arm cuff and finger monitor.	A1 **(8)**
b.	The difference in measurements of blood pressure is normally distributed	B1 **(1)**
	Notes. (a) One tail test Loses the first B1 . CV is 1.383 in this case. Can get 7/8 (b) looking for the difference in measurements. Not just it is normally distributed.	

Question Number	Scheme	Marks
2. a)	$\mathrm{E}(\overline{X}) = \mu$	B1
	$\mathrm{Var}\,(\overline{X}) = \mathrm{Var}\left(\dfrac{X_1 + X_2 + X_3 + \ldots + X_n}{n}\right)$	
	$= \dfrac{\sigma^2}{n}$	B1
		(2)
b)	$\mathrm{E}(U) = \dfrac{1}{n+m}\left(n\mathrm{E}(\overline{X}) + m\mathrm{E}(\overline{Y})\right)$	M1
	$= \dfrac{1}{n+m}(n\mu + m\mu)$	A1
	$= \mu \quad \Rightarrow U$ is unbiased state unbiased	A1
		(3)
c)	$\mathrm{Var}\,(\overline{Y}) = \dfrac{\sigma^2}{m}$	B1
	$\mathrm{Var}\,(U) = \dfrac{n^2\mathrm{Var}(\overline{X}) + m^2\mathrm{Var}(\overline{Y})}{(n+m)^2}$	M1
	$= \dfrac{n^2\dfrac{\sigma^2}{n} + m^2\dfrac{\sigma^2}{m}}{(n+m)^2}$	A1
	$= \dfrac{n\sigma^2 + m\sigma^2}{(n+m)^2}$	
	$= \dfrac{\sigma^2}{n+m}$ * cso	A1
		(4)
d)	$\dfrac{n\overline{X} + m\overline{Y}}{n+m}$ is a better estimate since variance is smaller.	B1 B1
		(2)

Question Number	Scheme	Marks
3. a	$H_0 : \sigma^2_F = \sigma^2_M \quad H_1 : \sigma^2_F \neq \sigma^2_M$	B1
	$s^2_F = \frac{1}{6}(17956.5 - 7 \times 50.6^2) = \frac{33.98}{6} = 5.66333....$	B1
	$s^2_M = \frac{1}{9}(28335.1 - 10 \times 53.2^2) = \frac{32.7}{9} = 3.63333....$	B1
	$\frac{s^2_F}{s^2_M} = 1.5587...$ (Reciprocal 0.6415)	M1 A1
	$F_{6,9} = 3.37$ (or 0.24)	B1
	Not in critical region. Variances of the two distributions are the same	A1
		(7)
b.	$H_0 : \mu_F = \mu_M \quad H_1 : \mu_F < \mu_M$	B1
	Pooled estimate $s^2 = \frac{6 \times 5.66333.... + 9 \times 3.63333}{15}$	M1
	$= 4.44533$	
	$s = 2.11$	
	$t = \frac{50.6 - 53.2}{2.11\sqrt{\frac{1}{7} + \frac{1}{10}}} = \pm 2.50$	M1 A1
	C.V. $t_{15}(5\%) = \pm 1.753$	B1
	Significant. The mean length of the females forewing is less than the length of the males forewing	A1
		(6)
	Notes (a) need to have variance and the same o.e (b) need female and forewing(wing)	

Question Number	Scheme	Marks
4.a)	$H_0 : \sigma^2 = 0.9$ $H_1 : \sigma^2 \neq 0.9$	B1
	$\nu = 19$	
	CR (Lower tail 10.117) Upper tail 30.144	B1 B1
	Test statistic = $\frac{19 \times 1.5}{0.9} = 31.6666$, significant	M1 A1
	There is sufficient evidence that the variance of the length of spring is different to 0.9	A1
		(6)
b)	$H_0 : \mu = 100$ $H_1 : \mu > 100$	B1
	$t_{19} = 1.328$	B1
	$t = \frac{100.6 - 100}{\sqrt{\frac{1.5}{20}}} = 2.19$	M1 A1 A1
	Significant. The mean length of spring is greater than 100	B1 **(6)**
	Notes (a) only need to see 30.144 need variance in conclusion (b) conclusion must be in context. Length of spring needed	

Question Number	Scheme	Marks
5.a)	Power = P (X≤ 3 / λ)	M1
	$= e^{-\lambda} + e^{-\lambda}\lambda + \frac{e^{-\lambda}\lambda^2}{2} + \frac{e^{-\lambda}\lambda^3}{6}$	A1
	$= \frac{e^{-\lambda}}{6}(6 + 6\lambda + 3\lambda^2 + \lambda^3)$	A1 **(3)**
b)	CR is X≤ 3	M1
	Size = P[X≤ 3 / λ =7]	
	= 0.0818	A1 **(2)**
c)	P(Type II error) = 1 – power	
	$= 1 - \frac{e^{-4}}{6}(6 + 6\times 4 + 3\times 4^2 + 4^3)$	M1
	= 0.5665..	A1 **(2)**
6.a)	$\frac{\overline{X} - 250}{\frac{4}{\sqrt{15}}} > 2.3263$ or $\frac{\overline{X} - 250}{\frac{4}{\sqrt{15}}} < -2.3263$ ± 2.3262	B1 M1
	$\overline{X} > 252.40...$ or $\overline{X} < 247.6...$ awrt 252 and 248	A1 **(3)**
b)	$P(\overline{X} < 252.4 / \mu = 254) - P(\overline{X} < 247.6 / \mu = 254)$ using their '252.4' and '247.6	M1
	$= P\left(Z < \frac{252.4 - 254}{\frac{4}{\sqrt{15}}}\right) - P\left(Z < \frac{247.6 - 254}{\frac{4}{\sqrt{15}}}\right)$ stand using 4/√15, 254 their '252.4' or '247.6	M1
	= P(Z < -1.5492) – P(Z < -6.20) -1.5492 and -6.20 o.e.	A1
	= (1 – 0.9394) – (1 – 1)	M1
	= 0.0606	A1 **(5)**
	Notes (a) only needs to try and find one side for M1 (b) only need to see one of the standardisation for second M1 if consider only 252.4 and get 0.0606 they get M0 M1 A0 M1 A1 ie they can get 3/5	

Question Number	Scheme		Marks
7.	$\bar{x} = 4.01$ $s = 0.7992\ldots$		B1 M1 A1
(a)	$4.01 \pm t_9(2.5\%)\dfrac{0.7992..}{\sqrt{10}} = 4.01 \pm 2.262\dfrac{0.7992..}{\sqrt{10}}$	2.262	B1
		their $\bar{x}$ and s and $\sqrt{10}$	M1 A1√
	$= 4.5816\ldots$ and $3.4383\ldots$	awrt 4.58 and 3.44	A1 **(7)**
(b)	$2.700 < \dfrac{9\times 0.7992..^2}{s^2} < 19.023$	2.7,19.023	B1 B1
		$9\times s^2/\sigma^2$	M1
	$\sigma^2 < 2.13, \quad \sigma^2 > 0.302$	both awrt 2.13, 0.302	A1 **(4)**
(c)	$P(X > 7) = P\left(Z > \dfrac{7-\mu}{\sigma}\right)$ needs to be as high as possible		M1
	Therefore μ and σ must be as big as possible		M1
	$= P\left(Z > \dfrac{7-4.581}{\sqrt{2.13}}\right)$		A1√
	$= 1 - 0.9515$		
	$= 0.0485$		
	$= 4.85\%$	4.8 to 4.9	A1 **(4)**
	Notes (a) $s^2 = 0.63877\ldots$ (c) M1 may be implied by them using their highest μ and σ.		