Paper Reference(s)

6686/01

Edexcel GCE

Statistics S4

Advanced Level

Thursday 24 June 2010 – 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, 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.

M35398A

1. A teacher wishes to test whether playing background music enables students to complete a task more quickly. The same task was completed by 15 students, divided at random into two groups. The first group had background music playing during the task and the second group had no background music playing.

The times taken, in minutes, to complete the task are summarised below.

	Sample size n	Standard deviation s	Mean $\bar{x}$
With background music	8	4.1	15.9
Without background music	7	5.2	17.9

You may assume that the times taken to complete the task by the students are two independent random samples from normal distributions.

(*a*) Stating your hypotheses clearly, test, at the 10% level of significance, whether or not the variances of the times taken to complete the task with and without background music are equal.

(5)

(*b*) Find a 99% confidence interval for the difference in the mean times taken to complete the task with and without background music.

(7)

Experiments like this are often performed using the same people in each group.

(*c*) Explain why this would not be appropriate in this case.

(1)

---

2. As part of an investigation, a random sample of 10 people had their heart rate, in beats per minute, measured whilst standing up and whilst lying down. The results are summarized below.

Person	1	2	3	4	5	6	7	8	9	10
Heart rate lying down	66	70	59	65	72	66	62	69	56	68
Heart rate standing up	75	76	63	67	80	75	65	74	63	75

(*a*) State one assumption that needs to be made in order to carry out a paired *t*-test.

(1)

(*b*) Test, at the 5% level of significance, whether or not there is any evidence that standing up increases people's mean heart rate by more than 5 beats per minute. State your hypotheses clearly.

(8)

---

3. A manager in a sweet factory believes that the machines are working incorrectly and the proportion p of underweight bags of sweets is more than 5%. He decides to test this by randomly selecting a sample of 5 bags and recording the number X that are underweight. The manager sets up the hypotheses H_0: $p = 0.05$ and H_1: $p > 0.05$ and rejects the null hypothesis if $x > 1$.

(*a*) Find the size of the test.

(2)

(*b*) Show that the power function of the test is

$$1 - (1 - p)^4(1 + 4p)$$

(3)

The manager goes on holiday and his deputy checks the production by randomly selecting a sample of 10 bags of sweets. He rejects the hypothesis that $p = 0.05$ if more than 2 underweight bags are found in the sample.

(*c*) Find the probability of a Type I error using the deputy's test.

(2)

The table below gives some values, to 2 decimal places, of the power function for the deputy's test.

p	0.10	0.15	0.20	0.25
Power	0.07	s	0.32	0.47

(*d*) Find the value of s.

(1)

Question 3 continues on page 4

The graph of the power function for the manager's test is shown in Figure 1.

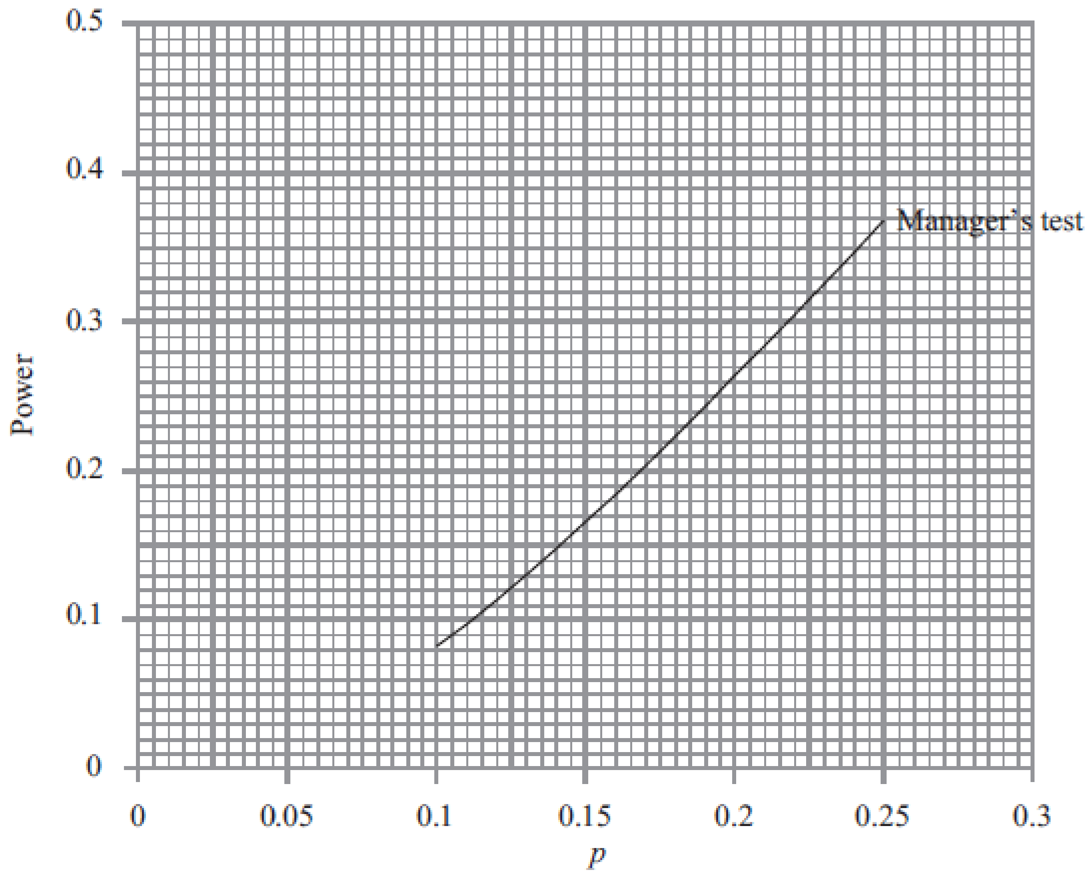


Figure 1

(*e*) On the same axes, draw the graph of the power function for the deputy's test.

(1)

(*f*) (i) State the value of p where these graphs intersect.

(ii) Compare the effectiveness of the two tests if p is greater than this value.

(2)

The deputy suggests that they should use his sampling method rather than the manager's.

(*g*) Give a reason why the manager might not agree to this change.

(1)

4. A random sample of 15 strawberries is taken from a large field and the weight x grams of each strawberry is recorded. The results are summarised below.

$$\sum x = 291 \qquad \sum x^2 = 5968$$

Assume that the weights of strawberries are normally distributed.

Calculate a 95% confidence interval for

(*a*) (i) the mean of the weights of the strawberries in the field,

(ii) the variance of the weights of the strawberries in the field.

(12)

Strawberries weighing more than 23 g are considered to be less tasty.

(*b*) Use appropriate confidence limits from part (*a*) to find the highest estimate of the proportion of strawberries that are considered to be less tasty.

(4)

---

5. A car manufacturer claims that, on a motorway, the mean number of miles per gallon for the Panther car is more than 70. To test this claim a car magazine measures the number of miles per gallon, x, of each of a random sample of 20 Panther cars and obtained the following statistics.

$$\bar{x} = 71.2 \qquad s = 3.4$$

The number of miles per gallon may be assumed to be normally distributed.

(*a*) Stating your hypotheses clearly and using a 5% level of significance, test the manufacturer's claim.

(5)

The standard deviation of the number of miles per gallon for the Tiger car is 4.

(*b*) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not there is evidence that the variance of the number of miles per gallon for the Panther car is different from that of the Tiger car.

(6)

---

6. Faults occur in a roll of material at a rate of λ per m^2. To estimate λ, three pieces of material of sizes 3 m^2, 7 m^2 and 10 m^2 are selected and the number of faults X_1, X_2 and X_3 respectively are recorded.

The estimator $\hat{\lambda}$, where

$$\hat{\lambda} = k(X_1 + X_2 + X_3)$$

is an unbiased estimator of λ.

(*a*) Write down the distributions of X_1, X_2 and X_3 and find the value of k.

(4)

(*b*) Find Var ($\hat{\lambda}$).

(3)

A random sample of n pieces of this material, each of size 4 m^2, was taken. The number of faults on each piece, Y, was recorded.

(*c*) Show that $\frac{1}{4}\overline{Y}$ is an unbiased estimator of λ.

(2)

(*d*) Find Var ($\frac{1}{4}\overline{Y}$).

(3)

(*e*) Find the minimum value of n for which $\frac{1}{4}\overline{Y}$ becomes a better estimator of λ than $\hat{\lambda}$.

(2)

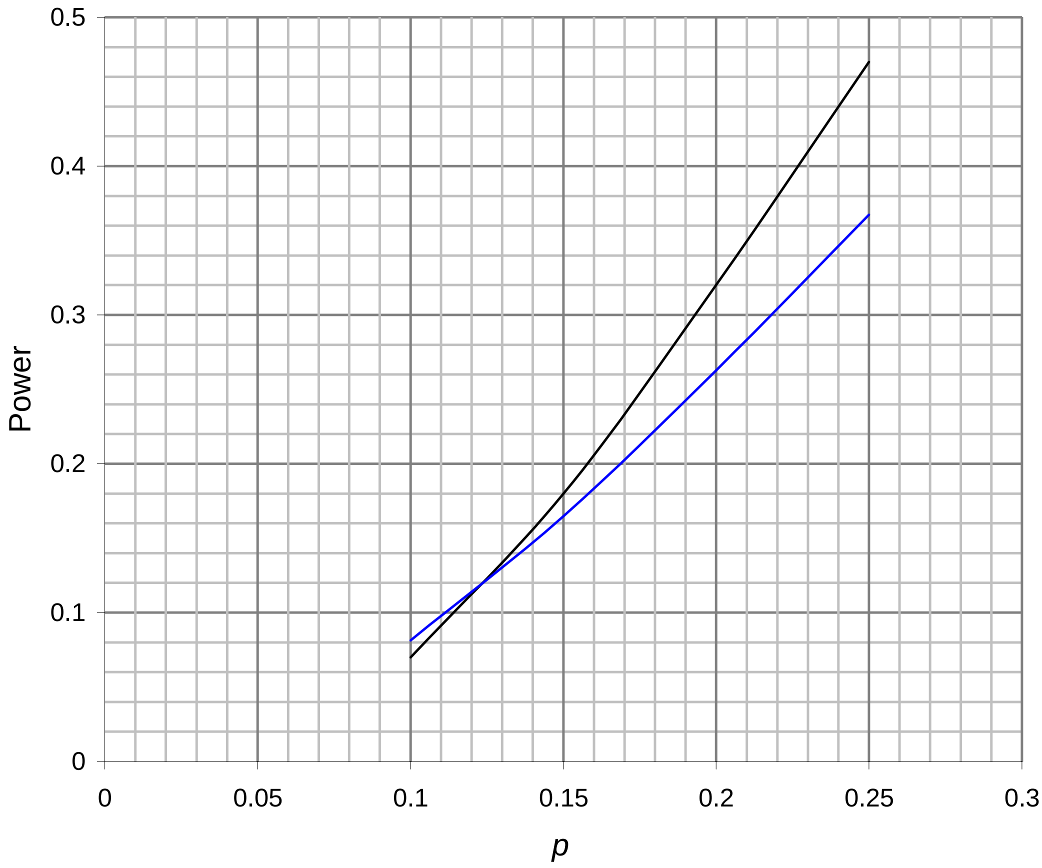
TOTAL FOR PAPER: 75 MARKS

END

June 2010
Statistics S4 6686
Mark Scheme

Question Number	Scheme	Marks
Q1 (a)	$H_0 : \sigma_1^2 = \sigma_2^2,\ H_1 : \sigma_1^2 \neq \sigma_2^2$	B1
	critical values $F_{6,7} = 3.87 \left(\frac{1}{F_{6,7}} = 0.258 \right)$	B1
	$\frac{s_2^2}{s_1^2} = \frac{5.2^2}{4.1^2}; = 1.61 \quad \left(\frac{s_1^2}{s_2^2} = \frac{4.1^2}{5.2^2} = 0.622 \right)$	M1; A1
	Since 1.61 (0.622) 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)	$Sp^2 = \frac{7 \times 4.1^2 + 6 \times 5.2^2}{7+6} = 21.53\ldots$	M1A1
	t_{13}=3.012	B1
	99% CI = (17.9 – 15.9) $\pm\, 3.012 \times \sqrt{21.53} \times \sqrt{\frac{1}{8}+\frac{1}{7}}$	M1A1ft
	= $\pm$(9.23, -5.233), [or accept: [0,9.23] or [-9.23,0]] awrt 9.23, -5.23	A1A1 (7)
(c)	a person will be quicker at the task second time through/ times not independent/ familiar with the task/groups are not independent	B1 (1)
		[13]
	Notes B1 Allow $\sigma_1 = \sigma_2$ and $\sigma_1 \neq \sigma_2$ B1 must match their F M1 for $\frac{s_2^2}{s_1^2}$ or other way up A1 awrt 1.61(0.622) M1 A1 Sp^2 may be seen in part a B1 3.012 only M1 for (17.9 – 15.9) $\pm$ t value $\times \sqrt{S_p^{\ 2}} \times \sqrt{\frac{1}{8}+\frac{1}{7}}$ A1ft their Sp^2 A1 awrt 9.23/-9.23 A1 awrt -5.23/5.23 (c) B1 any correct sensible comment	

Question Number	Scheme	Marks
Q2 (a)	The differences in the mean heart rates are normally distributed.	B1 (1)
(b)	D = standing up – lying down	
	H_0: $\mu_D = 5$ H_1: $\mu_D > 5$	B1
	d: 9, 6, 4, 2, 8, 9, 3, 5, 7, 7	M1
	$\bar{d} = 6$; $s_d = \sqrt{\dfrac{414 - 10\times 36}{9}} = 2.45$	M1;M1
	$t_9 = \dfrac{6-5}{2.45/\sqrt{10}} = 1.29$	M1A1
	$t_9(5\%) = 1.833$	B1
	insignificant. There is no evidence to suggest that heart rate rises by more than 5 beats when standing up.	A1 ft (8)
		[9]
	Notes must have "The differences in (mean heart rate) are normally distributed) B1 both correct allow μ_D -5 >0 (μ_D = -5 H_1: μ_D < -5) M1 finding differences M1 finding $\bar{d}$ M1 $\sqrt{\dfrac{\sum d^2 - 10\times(\bar{d})^2}{9}}$ o.e M1 $\pm\left(\dfrac{6-5}{s_d/\sqrt{10}}\right)$ need to see full expression with numbers in A1 awrt ± 1.29 . B1 ± 1.833 only A1 ft their CV and t. Need context. Heart rate and 5 beats	

Question Number	Scheme	Marks
Q3 (a)	$X \sim B(5, p)$	
	Size = P(reject H_0/ p = 0.05)	
	= P(X > 1/ p = 0.05)	
	= 1 – 0.9774	M1
	= 0.0226	A1 (2)
(b)	Power = 1 – P(0) – P(1)	M1
	$= 1 - (1-p)^5 - 5(1-p)^4 p$	M1
	$= 1 - (1-p)^4 (1 - p + 5p)$	
	$= 1 - (1-p)^4 (1 + 4p)$	A1cso
		(3)
(c)	$Y \sim B(10, p)$	
	P (Type I error) = P(Y> 2/ p = 0.05)	M1
	= 1 – 0.9885	
	= 0.0115	A1 (2)
(d)	s = 0.18	B1 (1)
(e)		B1ft (1)

Question Number	Scheme	Marks
(f)	**i** intersection 0.12 – 0.13 "their graphs intersection"	B1ft
	ii if $p > 0.12$ the deputy's test is more powerful.	B1 (2)
(g)	More powerful for $p < 0.12$ and p unlikely to be above 0.12	
	Allow it would cost more/take longer/more to sample	B1 (1)
		[12]
	Notes (a) M1 for finding P (X>1) A1 awrt 0.0226 (b) M1 for 1-P(0) – P(1) M1 for $1 - (1 - p)^5 - 5(1 - p)^4 p$ A1 cso (a) M1 for finding P(Y > 2) A1 awrt0.0115 (b) B1 0.18 cao (c) B1 graph. ft their value of s (d) B1 ft their intersection. B1 deputy test more powerful o.e. (e) If give first statement they must suggest p unlikely to be above 0.12	

Question Number	Scheme	Marks
Q4 (a)	$\bar{x} = \dfrac{291}{15} = 19.4 \quad s = \sqrt{\dfrac{5968 - 15\bar{x}^2}{14}} = 4.800$	M1M1
	i $t_{14} = 2.145$	B1
	95% CI = $19.4 \pm 2.145 \times \dfrac{4.800}{\sqrt{15}}$	M1 A1ft
	$= (16.7, 22.1)$	A1A1
	ii 95% CI is given by	
	$\dfrac{14 \times 4.800^2}{26.119} < \sigma^2 < \dfrac{14 \times 4.800^2}{5.629}$	M1 B1B1
	(12.4, 57.3) accept 12.3	A1A1 (12)
(b)	Require P(X > 23) = P$\left(Z > \dfrac{23-\mu}{\sigma}\right)$ to be as large as possible OR $\dfrac{23-\mu}{\sigma}$ to be as small as possible; both imply highest σ and μ. $\dfrac{23-22.1}{\sqrt{57.3..}} = 0.124$	M1M1
	P(Z > 0.124) = 1 – 0.5478	M1
	= 0.4522	A1 (4)
		[16]
	Notes (a)(i) M1 $\dfrac{291}{15}$ M1 $\sqrt{\dfrac{5968-15\bar{x}^2}{14}}$ B1 2.145 M1 $(19.4) \pm t \times \dfrac{\text{"their s"}}{\sqrt{15}}$ A1ft $19.4 \pm 2.145 \times \dfrac{\text{"their s"}}{\sqrt{15}}$ A1 awrt 16.7 A1 awrt 22.1 (ii) M1 $\dfrac{14 \times s^2}{\chi^2}$ B1 26.119 B1 5.629 A1 awrt12.4/12.3 A1 awrt 57.3 (b) M1 use of highest mean and sigma M1 standardising using values of mean and sigma from intervals M1 finding 1 – P(z > their value) A1 awrt 0.45	

Question Number	Scheme	Marks
Q5 (a)	$H_0: \mu = 70$ [accept ≤ 70], $H_1: \mu > 70$	B1
	$t = \dfrac{71.2-70}{3.4/\sqrt{20}} = 1.58$	M1A1
	critical value $t_{19}(5\%) = 1.729$	B1
	not significant, insufficient evidence to confirm manufacturer's claim	A1 ft (5)
(b)	$H_0: \sigma^2 = 16$, $H_1: \sigma^2 \ne 16$	B1
	test statistic $\dfrac{(n-1)s^2}{\sigma^2}$ =, $\dfrac{219.64}{16} = 13.7..$	M1 A1
	critical values $\chi^2_{19}(5\%)\text{ upper tail} = 32.852$ $\chi^2_{19}(5\%)\text{ lower tail} = 8.907$ not significant	B1 B1
	Insufficient evidence to suggest that the variance of the miles per gallon of the panther is different from that of the Tiger.	A1ft (6)
		[11]
	Notes (a) B1 both hypotheses using μ M1 $\dfrac{71.2-70}{3.4/\sqrt{20}}$ A1 awrt 1.73 A1 correct conclusion ft their t value and CV (b) B1 both hypotheses and 16. accept $\sigma = 4$ and $\sigma \ne 4$ M1 $\dfrac{(19)\times 3.4^2}{16}$ allow $\dfrac{(19)\times 3.4^2}{4}$ A1 awrt 13.7 B1 32.852 B1 8.907 A1 correct contextual comment NB those who use $\sigma^2 = 4$ throughout can get B0 M1 A0B1 B1 A1	

Question Number	Scheme	Marks
Q6 (a)	$X_1 \sim \text{Po}(3\lambda)$ $X_2 \sim \text{Po}(7\lambda)$ $X_3 \sim \text{Po}(10\lambda)$	M1
	$\text{E}(\hat{\lambda}) = k\,[\text{E}(X_1) + \text{E}(X_2) + \text{E}(X_3)]$ $= 20\lambda k$	M1
	$\hat{\lambda}$ unbiased therefore $20\lambda k = \lambda$	M1
	$k = \frac{1}{20}$	A1 (4)
(b)	$\text{Var}(\hat{\lambda}) = \frac{1}{20^2}\text{Var}(X_1 + X_2 + X_3)$	M1
	$= \frac{1}{20^2}(3\lambda + 7\lambda + 10\lambda)$	M1
	$= \frac{\lambda}{20}$	A1ft (3)
(c)	$Y \sim \text{Po}(4\lambda)$ $\text{E}\left(\frac{1}{4}\overline{Y}\right) = \frac{1}{4} \times 4\lambda = \lambda$ therefore unbiased	M1 A1 (2)
(d)	$\text{Var}\left(\frac{1}{4}\overline{Y}\right) = \frac{1}{16} \times \frac{4\lambda}{n}$	M1 B1
	$= \frac{\lambda}{4n}$	A1 (3)
(e)	$\frac{\lambda}{4n} < \frac{\lambda}{20}$	M1
	$n > 5$ therefore $n = 6$	A1 (2)
		[14]

Question Number	Scheme	Marks
Q6	Notes (a) M1 all 3 needed. Poisson and mean M1 adding their means M1 putting their E($\hat{\lambda}$) = λ A1 cao (b) M1 use of $k^2\text{Var}(X_1 + X_2 + X_3)$ M1 using their means from part(a) as Variances and adding together A1 cao (c) M1 use of 4λ A1 cso plus conclusion. Accept working out bias to = 0 (d) M1 $\frac{1}{16} \times \text{Var}\bar{Y}$ B1 for $\text{Var}\bar{Y} = \frac{4\lambda}{n}$ A1 cao (e) M1 for $\text{Var}\left(\frac{1}{4}\bar{Y}\right) < \text{Var}(\hat{\lambda})$ A1 n = 6	