Paper Reference(s)

6686/01
Edexcel GCE
Statistics S4
Advanced Level
Thursday 23 June 2011 – 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 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 not gain full credit.

P35418A

1. Find the value of the constant a such that

$$P(a < F_{8, 10} < 3.07) = 0.94.$$

(2)

---

2. Two independent random samples $X_1, X_2, ..., X_7$ and Y_1, Y_2, Y_3, Y_4 were taken from different normal populations with a common standard deviation σ.

The following sample statistics were calculated.

$$s_x = 14.67 \qquad s_y = 12.07$$

Find the 99% confidence interval for σ^2 based on these two samples.

(5)

---

3. Manuel is planning to buy a new machine to squeeze oranges in his cafe and he has two models, at the same price, on trial. The manufacturers of machine B claim that their machine produces more juice from an orange than machine A. To test this claim Manuel takes a random sample of 8 oranges, cuts them in half and puts one half in machine A and the other half in machine B. The amount of juice, in ml, produced by each machine is given in the table below.

Orange	1	2	3	4	5	6	7	8
Machine A	60	58	55	53	52	51	54	56
Machine B	61	60	58	52	55	50	52	58

Stating your hypotheses clearly, test, at the 10% level of significance, whether or not the mean amount of juice produced by machine B is more than the mean amount produced by machine A.

(8)

---

4. A proportion p of letters sent by a company are incorrectly addressed and if p is thought to be greater than 0.05 then action is taken.

Using H_0: $p = 0.05$ and H_1: $p > 0.05$, a manager from the company takes a random sample of 40 letters and rejects H_0 if the number of incorrectly addressed letters is more than 3.

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

(2)

(*b*) Find the probability of a Type II error in the case where p is in fact 0.10.

(2)

Table 1 below gives some values, to 2 decimal places, of the power function of this test.

p	0.075	0.100	0.125	0.150	0.175	0.200	0.225
Power	0.35	s	0.75	0.87	0.94	0.97	0.99

Table 1

(*c*) Write down the value of s.

(1)

A visiting consultant uses an alternative system to test the same hypotheses. A sample of 15 letters is taken. If these are all correctly addressed then H_0 is accepted. If 2 or more are found to have been incorrectly addressed then H_0 is rejected. If only one is found to be incorrectly addressed then a further random sample of 15 is taken and H_0 is rejected if 2 or more are found to have been incorrectly addressed in this second sample, otherwise H_0 is accepted.

(*d*) Find the size of the test used by the consultant.

(3)

Figure 1 shows the graph of the power function of the test used by the consultant.

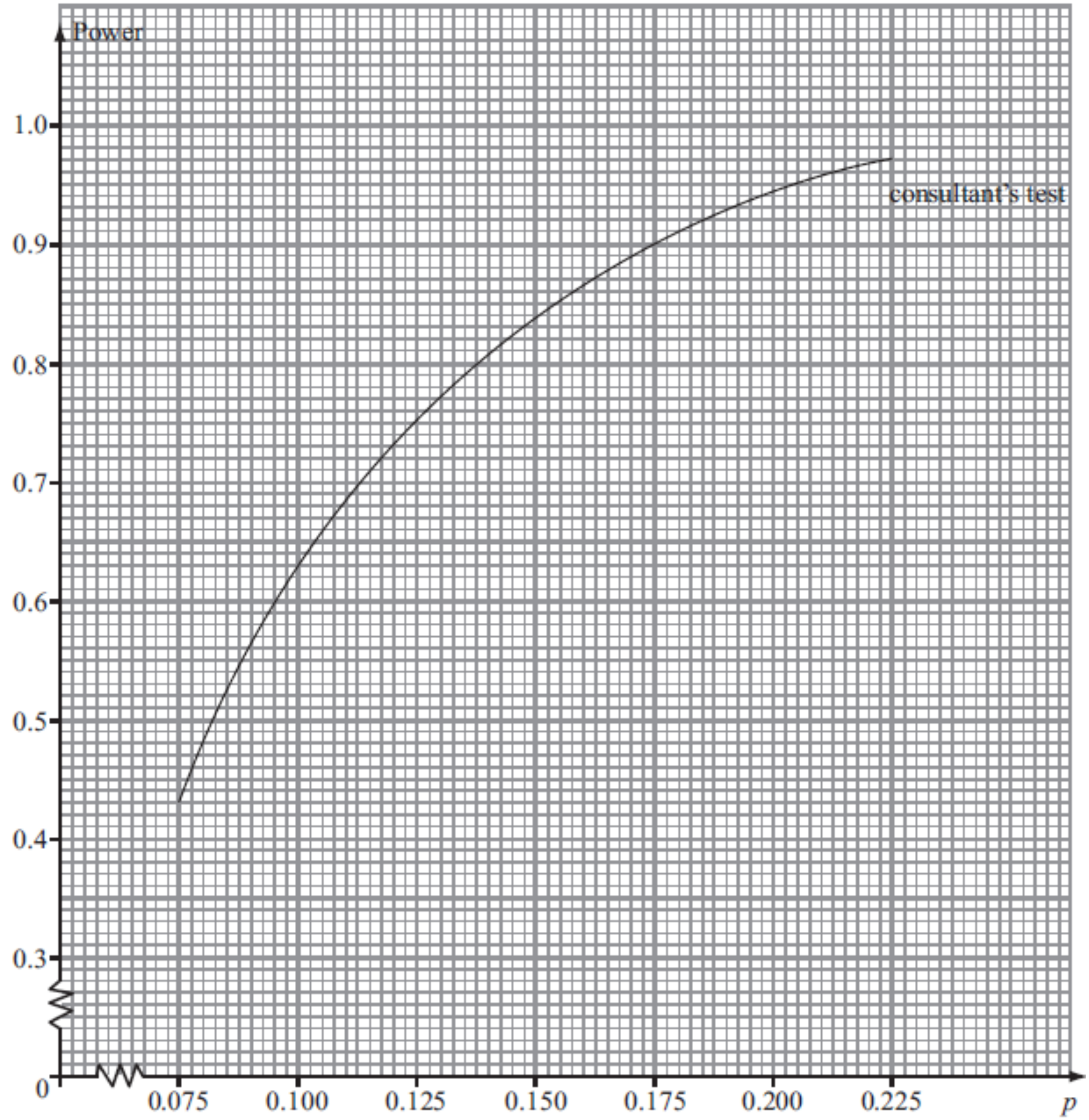

Figure 1

(*e*) On Figure 1 draw the graph of the power function of the manager's test.

(2)

(*f*) State, giving your reasons, which test you would recommend.

(2)

5. The weights of the contents of breakfast cereal boxes are normally distributed.

A manufacturer changes the style of the boxes but claims that the weight of the contents remains the same. A random sample of 6 old style boxes had contents with the following weights (in grams).

512 503 514 506 509 515

The weights, y grams, of the contents of an independent random sample of 5 new style boxes gave

$$\bar{y} = 504.8 \text{ and } s_y = 3.420$$

(*a*) Use a two-tail test to show, at the 10% level of significance, that the variances of the weights of the contents of the old and new style boxes can be assumed to be equal. State your hypotheses clearly.

(5)

(*b*) Showing your working clearly, find a 90% confidence interval for $\mu_x - \mu_y$, where μ_x and μ_y are the mean weights of the contents of old and new style boxes respectively.

(7)

(*c*) With reference to your confidence interval comment on the manufacturer's claim.

(2)

6. A random sample $X_1, X_2, ..., X_n$ is taken from a population where each of the X_i have a continuous uniform distribution over the interval $[0, \beta]$.

The random variable $Y = \max\{X_1, X_2, ..., X_n\}$.

The probability density function of Y is given by

$$f(y) = \begin{cases} \dfrac{n}{\beta^n} y^{n-1} & 0 \le y \le \beta, \\ 0 & \text{otherwise.} \end{cases}$$

(*a*) Show that $E(Y^m) = \dfrac{n}{n+m}\beta^m$.

(3)

(*b*) Write down E(Y).

(1)

(*c*) Using your answers to parts (*a*) and (*b*), or otherwise, show that

$$\text{Var}(Y) = \frac{n}{(n+1)^2(n+2)}\beta^2.$$

(3)

(*d*) State, giving your reasons, whether or not Y is a consistent estimator of β.

(3)

The random variables $M = 2\overline{X}$, where $\overline{X} = \dfrac{1}{n}(X_1 + X_2 + ... + X_n)$, and $S = kY$, where k is a constant, are both unbiased estimators of β.

(*e*) Find the value of k in terms of n.

(1)

(*f*) State, giving your reasons, which of M and S is the better estimator of β in this case.

(3)

Five observations of X are: 8.5 6.3 5.4 9.1 7.6

(*g*) Calculate the better estimate of β.

(2)

---

7. A machine produces components whose lengths are normally distributed with mean 102.3 mm and standard deviation 2.8 mm. After the machine had been serviced, a random sample of 20 components were tested to see if the mean and standard deviation had changed. The lengths, x mm, of each of these 20 components are summarised as

$$\sum x = 2072 \qquad \sum x^2 = 214\,856$$

(*a*) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not there is evidence of a change in standard deviation.

(7)

(*b*) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not the mean length of the components has changed from the original value of 102.3 mm using

(i) a normal distribution,

(ii) a t distribution.

(9)

(*c*) Comment on the mean length of components produced after the service in the light of the tests from part (*a*) and part (*b*). Give a reason for your answer.

(2)

TOTAL FOR PAPER: 75 MARKS

END

Question Number	Scheme	Marks
1.	$\mathrm{P}\left(F_{8,10} > 3.07\right) = 0.05$	
	So need $\mathrm{P}\left(F_{10,8} > x\right) = 0.01$ so $x = 5.81$	B1
	So $a = \dfrac{1}{5.81} =$ **<u>0.172</u>** awrt 0.172	B1
		2
2.	$s_p{}^2 = \dfrac{6s_x{}^2 + 3s_y{}^2}{9}$ $(= 192.03\ldots)$	M1
	$1.735 < \dfrac{9s_p{}^2}{\sigma^2} < 23.589$	B1M1B1
	So 99% confidence interval is $(73.26\ldots., 996.14\ldots.)$ **awrt (<u>73.3, 996)</u>**	A1
		5
3.	$d = B - A: 1, 2, 3, -1, 3, -1, -2, 2$	M1
	$\bar{d} = 0.875$	M1
	$s_d{}^2 = \dfrac{33 - 8\times 0.875^2}{7} = (3.8392\ldots)$	M1
	$\mathrm{H}_0: \mu_d = 0 \quad \mathrm{H}_1: \mu_d > 0$	B1
	$t_7 = \dfrac{0.875}{\frac{s_p}{\sqrt{8}}} = 1.263\ldots$ **awrt <u>1.26</u>**	M1A1
	$t_7(10\%)$ one tail critical value is **<u>1.415</u>**	B1
	Not significant. There is insufficient evidence to support the claim of manufacturer *B* or machine *B* does not produce more juice (than machine *A*)	A1
		8

Question Number	Scheme	Marks
4. **(a)**	[X = no. of incorrectly addressed letters. X ~B(40,0.05)] $\text{P}(X > 3) = 1 - \text{P}(X \le 3), = 1 - 0.8619 = 0.1381$ **awrt <u>0.138</u>**	M1, A1 (2)
(b)	$\text{P}\left(\text{Type II Error}\right) = \text{P}\left(X \le 3 \middle\| p = 0.10\right)$ $= 0.4231$ **awrt <u>0.423</u>**	M1 A1 (2)
(c)	Power = 1 - P(Type II error) so s = **<u>0.58</u>** (0.5769)	B1 (1)
(d)	Y = no. of incorrectly addressed letters in a sample of 15. Y ~B(15, 0.05) Size = P($Y \ge 2$) + P(Y = 1)$\times$P($Y \ge 2$) = [1 - 0.8290]$\times$[1 + 0.8290 - 0.4633] = 0.23353… **awrt <u>0.23</u>**	M1 A1 A1 (3)
(e)	(use overlay)	B1B1 (2)
(f)	2nd/ consultants test is quicker (since it uses fewer letters) 2nd / consult test is more powerful for p < 0.125 (and values greater than this should be unlikely)	B1 B1 (2) **12**

Question Number	Scheme	Marks
5.		
(a)	$s_x{}^2 = \dfrac{1559691 - 6\times\left(\dfrac{3059}{6}\right)^2}{5} = 22.1666...$	M1
	$\mathrm{H}_0 : \sigma_x{}^2 = \sigma_y{}^2 \quad \mathrm{H} : \sigma_x{}^2 \neq \sigma_y{}^2$	B1
	$\dfrac{s_x{}^2}{s_y{}^2} = 1.895....$	M1
	$F_{5,4} = 6.26$	B1
	$\dfrac{s_x{}^2}{s_y{}^2} = 1.895....$ **awrt <u>1.90</u>** and comment : not significant - variances of **<u>weights</u>** of the two **<u>boxes</u>** can be assumed equal.	A1 (5)
(b)	$\bar{x} = 509.833... \quad \Rightarrow \quad \bar{x} - \bar{y} = 5.03333$	M1
	$s_p{}^2 = \dfrac{5s_x{}^2 + 4s_y{}^2}{9} = 17.513....$ **awrt <u>17.5</u>**	M1A1
	5% two tail t value is $t_9 = 1.833$	B1
	90% confidence interval is $5.03...\pm 1.833\times\sqrt{17.513...}\times\sqrt{\tfrac{1}{6}+\tfrac{1}{5}}$	M1
	(0.388…., 9.6782…) **awrt <u>(0.388, 9.68)</u>**	A1, A1 (7)
(c)	Zero is not in CI, there <u>is</u> evidence to <u>reject</u> the manufacturer's claim Or the weight of the contents of the boxes has changed.	B1ft, B1ft (2) **14**

Question Number	Scheme	Marks
6. **(a)**	$\mathrm{E}\left(Y^m\right)=\frac{n}{\beta^n}\int y^m \times y^{n-1}\ \mathrm{d}y$ =, $\left[\frac{n}{\beta^n}\times\frac{1}{m+n}\times y^{m+n}\right]_0^\beta$	M1, A1
	$=\frac{n}{\cancel{\beta^n}}\times\frac{1}{m+n}\times\beta^{m+\cancel{n}}=\frac{n}{m+n}\beta^m$ **(*)**	A1cso
		(3)
(b)	$\mathrm{E}\left(Y\right)=\frac{n}{n+1}\beta$	B1
		(1)
(c)	$\mathrm{E}\left(Y^2\right)=\frac{n}{n+2}\beta^2$, $\mathrm{Var}\left(Y\right)=\mathrm{E}\left(Y^2\right)-\left[\mathrm{E}(Y)\right]^2$	B1,M1
	$\mathrm{Var}\left(Y\right)=\frac{n}{n+2}\beta^2-\frac{n^2}{(n+1)^2}\beta^2=\frac{n}{(n+1)^2(n+2)}\beta^2$ **(*)**	A1cso
		(3)
(d)	As $n\to\infty$ $\mathrm{E}(Y)\to\beta$, $\mathrm{Var}(Y)\to 0$	M1,A1
	So Y <u>is</u> a consistent estimator for β.	A1
		(3)
(e)	$k=\frac{n+1}{n}$	B1
		(1)
(f)	$\mathrm{Var}\left(M\right)=4\mathrm{Var}(\bar{X})=4\frac{\sigma^2}{n}=\frac{4}{n}\times\frac{\beta^2}{12}=\frac{\beta^2}{3n}$	B1
	$\frac{(n+1)^2}{n^2}\times\frac{n}{(n+1)^2(n+2)}\beta^2=\frac{\beta^2}{n(n+2)}<\frac{\beta^2}{3n}$ so S is better ($n>1$)	M1A1
		(3)
(g)	Max = 9.1, $s=\frac{6}{5}\times 9.1=$ <u>**10.9(2)**</u>	M1A1
		(2) **16**

Question Number	Scheme	Marks
7.		
(a)	$s_x{}^2 = \dfrac{214856 - 20\times\left(\dfrac{2072}{20}\right)^2}{19} = 10.357\ldots$ **awrt <u>10.4</u>**	B1
	$H_0: \sigma = 2.8$ (or $\sigma^2 = \ldots$) $H_1: \sigma \neq 2.8$ (or $\sigma^2 \neq \ldots$)	B1
	$\dfrac{(n-1)s^2}{\sigma^2} \sim \chi^2_{19}$ test statistic = 25.102… **awrt <u>25.1</u>**	M1A1
	$\chi^2_{19}(0.025) = 32.852$, $\chi^2_{19}(0.975) = 8.907$	B1B1
	Not significant so no evidence of a change in standard deviation	A1
		(7)
(b) (i)	$H_0: \mu = 102.3$ $H_1: \mu \neq 102.3$	B1
	$z = \dfrac{\frac{2072}{20} - 102.3}{\frac{2.8}{\sqrt{20}}} = 2.0763\ldots$ **awrt <u>2.08</u>**	M1A1
	Critical value is z =1.96 or awrt 0.019 < 0.025	B1
	So a significant result, there is evidence of a change in mean length	A1ft
(ii)	$t = \dfrac{\frac{2072}{20} - 102.3}{\sqrt{\frac{10.357\ldots}{20}}} = 1.8064\ldots$ **awrt <u>1.81</u>**	M1A1
	Critical value of $t_{19} = 2.093$	B1
	Not significant, there is insufficient evidence of a change in mean length	A1
		(9)
(c)	(a) suggests that σ is unchanged so can use σ = 2.8 so normal test can be used	B1ft
	So using (i) conclude that there is evidence of an increase in mean length	B1ft
		(2) **18**