6686/01

Edexcel GCE

Statistics

Unit S4

Advanced Subsidiary / Advanced

Time: 1 hour 30 minutes

Materials required for the examination	**Items included with these question papers**
Answer Book (AB04) Graph Paper (GP02) Mathematical Formulae	Nil

Candidates may use any calculator EXCEPT **those with a facility for symbolic algebra, differentiation and/or integration. Thus candidates may** NOT **use calculators such as Texas TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.**

Instructions to Candidates

Full marks may be obtained for answers to ALL questions.

In the boxes on the Answer Book provided, 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 names and signature.

Information for Candidates

A booklet 'Mathematical Formulae including Statistical Formulae and Tables' is provided.

Values from the Statistical Tables should be quoted in full. The answer to each part of a question which requires the use of tables or a calculator should be given to three significant figures, unless otherwise specified.

This paper has 8 questions.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly numbered.

You must show sufficient working to make your methods clear to the Examiner. Answers without working will gain no credit

N1234 R6081

Edexcel
Success through qualifications

1. The random variable X has an F distribution with 5 and 10 degrees of freedom. Find values of a and b such that $\mathrm{P}(a \leq X \leq b) = 0.90$. **(3 marks)**

---

2. A certain brand of mineral water comes in bottles. The amount of water in a bottle, in millilitres, follows a normal distribution with mean μ. The manufacturer claims that $\mu = 125$. The manufacturer takes a sample of 15 bottles and measures the amount of water, x millilitres, in each bottle. The results are summarised by the following statistics.

$$\sum x = 1863, \quad \sum x^2 = 231\,435$$

Test, at the 5% level, whether or not there is evidence that the value of μ is lower than the manufacturer's claim. State your hypotheses clearly. **(8 marks)**

---

3. A group of 10 technology students is assessed by coursework and a written examination. The marks, given as percentages, are given in the table below.

Student	1	2	3	4	5	6	7	8	9	10
Coursework	65	73	62	81	78	74	68	59	76	70
Written exam.	61	76	65	77	72	71	72	42	69	63

(*a*) Use a suitable t-test to determine whether or not the coursework marks are significantly higher than the written examination marks. Use a 5% level of significance. **(8 marks)**

(*b*) State an assumption about the distribution of marks that is needed to make the above test valid. **(1 mark)**

---

4. An ornithologist is monitoring a large flock of a species of bird. The weights, in grams, of the adult birds follow a normal distribution with standard deviation 4.6. If the flock is healthy the mean weight of the adult birds should be 340g. Every two weeks the ornithologist captures a random sample of 12 adult birds and weighs them to test whether or not the mean weight of the flock is below 340g.

(*a*) State suitable null and alternative hypotheses for this test. **(2 marks)**

(*b*) Find the critical region for the sample mean using a 5% level of significance. **(2 marks)**

If the mean weight of the adult birds falls to 335g or less, the health of the flock could be in danger.

(*c*) Find P(Type II error) given that the mean weight of the flock is 335g. **(3 marks)**

Using a 1% level of significance, the critical region for the sample mean $\overline{X}$ is $\overline{X} < 336.9$ and $\text{P(Type II error} \mid \mu = 335) = 0.0751$. If the ornithologist has reason to believe that the mean weight of the adult birds has fallen below 340g he will provide extra food for the birds. If he fails to detect a drop in mean weight he might have to treat the birds at a much higher cost.

(*d*) Suggest, giving a reason, which of the two critical regions the ornithologist should use to keep unnecessary expense to a minimum. **(2 marks)**

---

5. A single observation x is taken from a Poisson distribution with parameter μ.
This observation is to be used to test $H_0 : \mu = 4.5$ against $H_1 : \mu < 4.5$. The critical region is chosen to be $x \leq 1$.

(*a*) Find the size of the critical region. **(1 mark)**

(*b*) Show that the power function for this test is given by $e^{-\mu}(1 + \mu)$. **(2 marks)**

The table below gives the values of the power function to 2 decimal places.

μ	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Power	0.91	s	0.56	0.41	t	0.20	0.14	0.09

(*c*) Calculate the value of s and the value of t. **(2 marks)**

(*d*) On graph paper draw a graph of the power function. **(3 marks)**

(*c*) Find the values of μ for which the test is more likely than not to come to the correct conclusion. **(3 marks)**

---

6. In an experiment to discover the possible error made when using a stopwatch, a student started the watch and stopped it as quickly as she could. The times taken, in hundredths of a second, for 8 such attempts are recorded below:

8, 12, 13, 9, 11, 7, 10, 9

Assuming that the times are normally distributed, find 90% confidence intervals for

(*a*) the mean, **(6 marks)**

(*b*) the variance. **(5 marks)**

---

7. (In this question max(a, b) = the greater of the two values a and b.)
A palaeontologist was attempting to estimate the length of time τ in years, during which a small herbivorous dinosaur existed on earth. She believed from other evidence that the earliest existence of the animal had been at the start of the Jurassic period. Two examples of the animal had been discovered in the fossil record, at times t_1 and t_2, after the start of the Jurassic period. Her model assumed that these times were values of two independent random variables T_1 and T_2, each having continuous uniform distribution on the interval $[0, \tau]$. She considered three estimators for τ:

$$\tau_1 = T_1 + T_2, \quad \tau_2 = \sqrt{3}\,|T_2 - T_1|, \quad \tau_3 = 1.5\max(T_1, T_2).$$

She used appropriate probability theory and calculated the results shown in the table below.

Variable	Expectation	Variance
T_1	$\frac{\tau}{2}$	$\frac{\tau^2}{12}$
$\lvert T_2 - T_1 \rvert$	$\frac{\tau}{3}$	$\frac{\tau^2}{18}$
max(T_1, T_2)	$\frac{2\tau}{3}$	$\frac{\tau^2}{18}$

Using these results,

(*a*) determine the bias of each of these estimators, **(4 marks)**

(*b*) find the variance of each of these estimators. **(3 marks)**

Using your results from parts (*a*) and (*b*) state, giving a reason,

(*c*) which estimator is the best of the three, **(2 marks)**

(*d*) which estimator is the worst. **(2 marks)**

---

8. A company undertakes investigations to compare fuel consumption x, in miles per gallon, of two different cars the *Relaxant* and the *Elegane*, with a view to purchasing a number of cars. A random sample of 13 *Relaxants* and an independent random sample of 7 *Eleganes* were taken and the following statistics calculated.

Car	Sample size n	Sample mean $\bar{x}$	Sample variance s^2
Relaxant	13	32.31	14.48
Elegane	7	28.43	35.79

The company assumes that fuel consumption for each make of car follows a normal distribution.

(*a*) Stating your hypotheses clearly test, at the 10% level of significance, whether or not the two distributions have the same variance. **(4 marks)**

(*b*) Stating your hypotheses clearly test, at the 5% level of significance, whether or not there is a difference in mean fuel consumption between the two types of car. **(6 marks)**

(*c*) Explain the importance of the conclusion to the test in part (*a*) in justifying the use of the test in part (*b*). **(1 mark)**

(*d*) State two factors which might be considered when undertaking an investigation into fuel consumption of two models of car to ensure that a fair comparison is made. **(2 marks)**

END

EDEXCEL FOUNDATION
Stewart House 32 Russell Square London WC1B 5DN
Telephone 0171-393 4444 Fax 0171-753 4558

Specimen Paper

Advanced Subsidiary/Advanced Level
General Certificate of Education

Subject **STATISTICS** Paper No. **S4**

Question number	Scheme	Marks
1.	Tables: $P(F_{5,10} \le 3.33) = 0.05 \Rightarrow b = 3.33$	B1
	$P(F_{10,5} \le 4.74) = 0.05 \Rightarrow P\left(F_{5,10} \ge \frac{1}{4.74}\right) = 0.05$ $\therefore a = 0.2109\ldots$	M1
	AWRT 0.21	A1 (3)
2.	$H_0 : \mu = 125$ $H_1 : \mu < 125$	B1
	$\bar{x} = \frac{1863}{15} = 124.2$, $s^2 = \frac{231435 - 15 \times 124.2^2}{14} = 3.6$	B1,M1A1
	test statistic: $t = \frac{124.2 - \mu}{\frac{s}{\sqrt{15}}} = -1.63299\ldots$ AWRT (-) 1.633	M1A1
	$t_{14}(5\%)$ c.v. is $(-)1.761$	B1
	$\therefore$ not significant, insufficient evidence that value of μ is lower than claimed	A1 $\sqrt{}$ (8)
3. (a)	d=coursework – written: 4, -3, -3, 4, 6, 3, -4, 17, 7, 7	M1
	$\bar{d} = \frac{38}{10} = 3.8$, $s_d^2 = \frac{498 - 10\bar{d}^2}{9} = 39.2\dot{8}$	A1A1
	test statistic: $t = \frac{3.8}{\frac{s_d}{\sqrt{10}}} = 1.917\ldots$ AWRT 1.92	M1A1
	$H_0 : \mu_d = 0$ $H_1 : \mu_d > 0$	B1
	$t_9(5\%)$ c.v. is $(\pm)1.833$; $\therefore$ significant – there is evidence coursework marks are higher	B1 A1$\sqrt{}$ (8)
(b)	The difference between the marks follows a normal distribution.	B1 (1)

Question number	Scheme	Marks
4.	$W \sim \mathrm{N}\left(\mu, 4.6^2\right)$	
(a)	H_0: $\mu = 340$ $\qquad$ H_1: $\mu < 340$	B1B1 (2)
(b)	Reject $\mu = 340$ if $\dfrac{\overline{X} - 340}{\frac{4.6}{\sqrt{12}}} < -1.6449 \Rightarrow \overline{X} < \underline{337.8}$	M1A1 (2)
(c)	P(Type II error) $= \mathrm{P}\left(\overline{X} \geq 337.8 \ldots \mid \mu = 335\right)$	M1
	$= \mathrm{P}(Z \geq 2.12\ldots) \quad = \underline{0.0170}$	M1A1 (3)
(d)	Type II errors are the more expensive.	M1
	So choose first option (i.e. 5%) since P(Type II error) is lower.	A1 (2)
5.	$H_0 : \mu = 4.5 \qquad H_1 : \mu < 4.5 \qquad X \sim \mathrm{P_o}(4.5)$	
(a)	Size = P($X \leq 1$) = 0.0611	B1 (1)
(b)	Power $\mathrm{P}(X \leq 1 \mid \mu) = \mathrm{e}^{-\mu}, + \dfrac{\mathrm{e}^{-\mu}.\mu^1}{1!} = \underline{\mathrm{e}^{-\mu}(1+\mu)}$	M1,A1 (2)
(c)	$s = 2\mathrm{e}^{-1} = \underline{0.74}$; $t = 3.5\mathrm{e}^{-2.5} = \underline{0.29}$	B1B1 (2)
(d)	See next page $\qquad$ Graph	B1B1B1 (3)
(e)	If $\mu < 4.5$ need to reject H_0 so P($X \leq 1$) > 0.5, i.e. $0 < \mu < 1.7$	M1A1 √
	If $\mu = 4.5$ need to accept H_0 so P($X > 1$) = 0.94, $\therefore\ \mu = 4.5$	B1 (3)
6. (a)	$\bar{x} = \dfrac{79}{8} = \underline{9.875}$, $\qquad s_x = 2.031 \left(= \sqrt{\dfrac{809 - 8.\bar{x}^2}{7}} \right)$	B1B1
	90% C.I. is $9.875 \pm t_7(5\%)\dfrac{2.031}{\sqrt{8}}$ $\qquad$ $t_7(5\%) = 1.895$	B1 M1
	$= (8.514\ldots, 11.235\ldots)$ $\qquad$ AWRT $\underline{8.51,\ 11.23 \sim 11.24}$	A1A1 (6)
(b)	$2.167 < \dfrac{7 \times s_x^2}{\sigma^2} < 14.067$	B1M1B1
	ie $\qquad \sigma^2 > 2.0526\ldots$ $\qquad$ AWRT $\underline{2.05}$	A1
	and $\qquad \sigma^2 < 13.3248\ldots$ $\qquad$ AWRT $\underline{13.3}$	A1 (5)

Question number	Scheme	Marks
7. **(a)**	$E(\tau_1)=\frac{\tau}{2}+\frac{\tau}{2}=\tau$, Bias = 0	B1
	$E(\tau_2)=\sqrt{3}\times\frac{\tau}{3}=\frac{\tau}{\sqrt{3}}$, Bias $=\tau\left(\frac{1}{\sqrt{3}}-1\right)$	B1, B1
	$E(\tau_3)=1.5\times 2\frac{\tau}{3}=\tau$, Bias = 0	B1 (4)
(b)	$\text{Var}(\tau_1)=\frac{\tau^2}{12}+\frac{\tau^2}{12}=\frac{\tau^2}{6}$	B1
	$\text{Var}(\tau_2)=\left(\sqrt{3}\right)^2.\frac{\tau^2}{18}=\frac{\tau^2}{6}$	B1
	$\text{Var}(\tau_3)=\frac{9}{4}.\frac{\tau^2}{18}=\frac{\tau^2}{8}$	B1 (3)
(c)	τ_3 is the best ∵, smallest variance and unbiased	M1A1 (2)√
(d)	τ_2 is the worst ∵, larger variance and unbiased	M1A1 (2)√

Question number	Scheme	Marks
8. **(a)**	$H_0 : \sigma_R^2 = \sigma_E^2$ $\qquad$ $H_1 : \sigma_R^2 \neq \sigma_E^2$	B1
	$F_{6,\,12}$ (5%) $_{1\text{ tail}}$ cv = 3.00, $\dfrac{s_E^2}{s_R^2} = \dfrac{35.79}{14.48} = 2.4716...$	B1, M1
	Not significant so do not reject – insufficient evidence to suspect $\sigma_R^2 \neq \sigma_E^2$	A1 √ (4)
(b)	$H_0 : \mu_R = \mu_E$ $\qquad$ $H_1 : \mu_R \neq \mu_E$	B1
	$s^2 = \dfrac{6 \times 35.79 + 12 \times 14.48}{18} = 21.58\dot{3}$	M1
	Test statistic: $\quad t = \dfrac{32.31 - 28.43}{s\sqrt{\frac{1}{13} + \frac{1}{7}}} = 1.78146...$ $\qquad$ AWRT 1.78	M1A1
	$t_{18}(5\%)_{2\text{ tail}}$ cv = 2.101	B1
	∴Not significant: insufficient evidence of difference in mean performance	A1 √ (6)
(c)	Test in (b) requires $\sigma_1^2 = \sigma_2^2$	B1 (1)
(d)	eg same: type of driving roads and journey length weather driver } Any two sensible suggestions	B1B1 (2)