OCR
RECOGNISING ACHIEVEMENT

Friday 20 January 2012 – Afternoon

A2 GCE MATHEMATICS

4734 Probability & Statistics 3

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4734
- List of Formulae (MF1)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

4733000112

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.

- The Question Paper will be found in the centre of the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book.** Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Answer **all** the questions.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Give non-exact numerical answers correct to 3 significant figures unless a different degree of accuracy is specified in the question or is clearly appropriate.

INFORMATION FOR CANDIDATES

This information is the same on the Printed Answer Book and the Question Paper.

- The number of marks is given in brackets **[ ]** at the end of each question or part question on the Question Paper.
- **You are reminded of the need for clear presentation in your answers.**
- The total number of marks for this paper is **72**.
- The Printed Answer Book consists of **12** pages. The Question Paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER/INVIGILATOR

- Do not send this Question Paper for marking; it should be retained in the centre or recycled. Please contact OCR Copyright should you wish to re-use this document.

© OCR 2012 [J/102/2706]
DC (NH) 48819/3
OCR is an exempt Charity

1 In a test of association of two factors, A and B, a 2×2 contingency table yielded 5.63 for the value of χ^2 with Yates' correction.

(i) State the null hypothesis and alternative hypothesis for the test. **[1]**

(ii) State how Yates' correction is applied, and whether it increases or decreases the value of χ^2. **[2]**

(iii) Carry out the test at the $2\frac{1}{2}\%$ significance level. **[3]**

2 An investigation in 2007 into the incidence of tuberculosis (TB) in badgers in a certain area found that 42 out of a random sample of 190 badgers tested positive for TB.
In 2010, 48 out of a random sample of 150 badgers tested positive for TB.

(i) Assuming that the population proportions of badgers with TB are the same in 2007 and 2010, obtain the best estimate of this proportion. **[1]**

(ii) Carry out a test at the $2\frac{1}{2}\%$ significance level of whether the population proportion of badgers with TB increased from 2007 to 2010. **[6]**

3 The continuous random variable U has a normal distribution with unknown mean μ and known variance 1. A random sample of four observations of U gave the values

$$3.9,\ 2.1,\ 4.6 \text{ and } 1.4.$$

(i) Calculate a 90% confidence interval for μ. **[3]**

(ii) The probability that the sum of four random observations of U is less than 11 is denoted by p. For each of the end points of the confidence interval in part **(i)** calculate the corresponding value of p. **[5]**

4 X is a continuous random variable with the distribution $N(48.5, 12.5^2)$. The values of X are transformed to standardised values of Y, using the equation $Y = aX + b$, where a and b are constants with $a > 0$.

(i) Find values of a and b for which the mean and standard deviation of Y are 40 and 10 respectively. **[4]**

(ii) State the distribution of Y. **[1]**

Two randomly chosen standardised values are denoted by Y_1 and Y_2.

(iii) Calculate the probability that Y_2 is at least 10 greater than Y_1. **[5]**

5 A statistician suggested that the weekly sales X thousand litres at a petrol station could be modelled by the following probability density function.

$$f(x) = \begin{cases} \frac{1}{40}(2x+3) & 0 \leqslant x < 5, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Show that, using this model, $P(a \leqslant X < a+1) = \frac{a+2}{20}$ for $0 \leqslant a \leqslant 4$. **[3]**

Sales in 100 randomly chosen weeks gave the following grouped frequency table.

x	$0 \leqslant x < 1$	$1 \leqslant x < 2$	$2 \leqslant x < 3$	$3 \leqslant x < 4$	$4 \leqslant x < 5$
Frequency	16	12	18	30	24

(ii) Carry out a goodness of fit test at the 10% significance level of whether f(x) fits the data. **[7]**

6 The continuous random variable Y has probability density function given by

$$f(y) = \begin{cases} -\frac{1}{4}y & -2 \leqslant y < 0, \\ \frac{1}{4}y & 0 \leqslant y \leqslant 2, \\ 0 & \text{otherwise.} \end{cases}$$

Find

(i) the interquartile range of Y, **[4]**

(ii) Var (Y), **[5]**

(iii) E ($|Y|$). **[4]**

7 The manufacturer's specification for batteries used in a certain electronic game is that the mean lifetime should be 32 hours. The manufacturer tests a random sample of 10 batteries made in Factory A, and the lifetimes (x hours) are summarised by

$$n = 10, \sum x = 289.0 \text{ and } \sum x^2 = 8586.19.$$

It may be assumed that the population of lifetimes has a normal distribution.

(i) Carry out a one-tail test at the 5% significance level of whether the specification is being met. **[7]**

(ii) Justify the use of a one-tail test in this context. **[1]**

Batteries made with the same specification are also made in Factory B. The lifetimes of these batteries are also normally distributed. A random sample of 12 batteries from this factory was tested. The lifetimes are summarised by

$$n = 12, \sum x = 363.0 \text{ and } \sum x^2 = 11\,290.95.$$

(iii) (a) State what further assumption must be made in order to test whether there is any difference in the mean lifetimes of batteries made at the two factories.
Use the data to comment on whether this assumption is reasonable. **[3]**

(b) Carry out the test at the 10% significance level. **[7]**

Question			Answer	Marks	Guidance
1	(i)		H_0: *A* and *B* are not associated H_1: *A* and *B* are associated	B1 **[1]**	For both. Allow indpt., not indpt.
1	(ii)		Yates $\chi^2 = \sum(\lvert O - E\rvert - 0.5)^2/E$ which decreases the value	B1 B1 **[2]**	Dep '-0.5' seen.
1	(iii)		CV 5.024 seen 5.63 > CV and reject H_0 There is evidence at the $2\frac{1}{2}$% SL of an association between *A* and *B*	B1 M1 A1 **[3]**	Ft their CV Allow B1 if correct conclusion, but comparison not shown. CWO (ie from 5.024)
2	(i)		Est(p) = (48+42)/(190+150) (=9/34 or 0.265)	B1 **[1]**	Allow consistent use of '-' instead of '+' throughout.
2	(ii)		H_0: $p_{10} = p_{07}$, H_1: $p_{10} > p_{07}$ where the *p*s denote relevant population proportions Test statistic = (48/150 – 42/190)/ sd sd = $\sqrt{(^9/_{34} \times {}^{25}/_{34}(190^{-1} + 150^{-1}))}$ TS = 2.053 Compare with 1.96 and reject H_0; and conclude that there is evidence that the proportion of badgers in the area with TB has increased	B1 M1 A1 A1 M1 A1ft **[6]**	If eg x,y used – must be defined. sd involving 190 and 150 and addition of 'variances' Allow A1 ft for sd $\sqrt{(48\text{x}102/150^3)+(42\text{x}148/190^3)}$ or TS=2.038 SC Allow for correct comparison with 2.24 for 2-tail test. Conclusion, contextualised, not too assertive.
3	(i)		Use 3 ± (*z or t*)$\sqrt{(1/4)}$ With *z* = 1.645 (2.1775, 3.8225)	M1 B1 A1 **[3]**	Sample mean intended ART (2.18, 3.82)
3	(ii)		$S \sim N(4\mu, 4)$ *Use* z=([11 – 4μ]/"2") ; 1.14/5 & -2.14/5 ϕ(z) = 0.016 & 0.874	M1 M1 A1 M1 A1 **[5]**	Or use distribution of sample mean with 2.75

Question			Answer	Marks	Guidance
4	(i)		$E(Y) = aE(X) + b \Rightarrow 40 = 48.5a + b$ $Var(Y) = a^2 Var(X)$ OR $SDY = aSDX$, $(10=12.5a)$ $\therefore a=0.8$ $b=1.2$	M1 M1 M1 A1 **[4]**	 If Var(Y)=…+b used allow final M1 for a=0.794 b=1.49 from GC Solve simultaneously
4	(ii)		N(40, 100)	B1 **[1]**	Allow 10^2
4	(iii)		Use $Y_2 - Y_1 \sim N(0, \sigma^2)$ $\sigma^2 = 200$ $P(Y_2 - Y_1 \geq 10) =$ $=1 - \Phi[10/\sqrt{(\text{“}200\text{”})}]$ $= 0.2399$	M1 A1 A1 M1 A1 **[5]**	Allow use of X instead of Y. $\sigma^2=312.5$ 10/a (=12.5) [“12.5”/$\sqrt{\text{”312.5”}}$] ART 0.240
5	(i)		$\int_a^{a+1} \frac{2x+3}{40} dx = \left[\frac{x^2+3x}{40}\right]_a^{a+1}$ $= (a + 2)/20$ AG	M1 A1 A1 **[3]**	Or correct use of F(x) or $(2x+3)^2/160$ Properly obtained
5	(ii)		(H_o:f(x) fits data, H_1:f(x) does not fit data) E-values: 10 15 20 25 30 $= 36/10 + 9/15 + 4/20 + 25/25 + 36/30$ $= 6.6$ Compare with 7.779 and do not reject H_o There is insufficient evidence (at the 10%SL) that f(x) does not fit the data	 B2 M1 A1 B1 M1 A1ft **[7]**	 B1 for 2 correct E-values, B2 for all correct M1 for 2 correct χ^2 values ft CAO B1 for 7.779 Allow M1 for consistent conclusion from wrong CV. Dep 7.779 used. Allow f(x) fits data NOT Data fits f(x)
6	(i)		$Q_3 : \int_{Q_3}^{2} \frac{1}{4} y dy = 1/4 \Rightarrow \left[\frac{1}{8}y^2\right]_{Q_2}^{2} = 1/4 \Rightarrow Q_3 = \sqrt{2}$ oe for LQ $Q_1 = -\sqrt{2}$, $IQR = 2\sqrt{2}$	M1 A1 A1 A1 **[4]**	For 1 correct integral with limits For correct equation For Q_1 or Q_3 allow 1.41,-1.41 For IQR allow 2.83

Question			Answer	Marks	Guidance
6	(ii)		$E(Y^2) = \int_{-2}^{0} -\frac{y^3}{4}dy + \int_{0}^{2} \frac{y^3}{4}dy$ $= [-y^4/16] + [y^4/16]$ $= 2$ $E(Y) = 0$; $Var(Y) = 2$	M1 A1 A1 B1 A1 **[5]**	Allow final A1 even if E(Y)=0 missing, but not if ² missing from $Var(Y)=E(Y^2)-[E(Y)]^2$
6	(iii)		$\int_{-2}^{0}(y^2/4)dy + \int_{0}^{2}(y^2/4)dy$ = 2/3 + 2/3 = 4/3	M1 A1 A1 A1 **[4]**	Allow M1A0 for one of these integrals correct and the other incorrect or missing or correct and not added. If f(|Y|)used , B1 for y/2, M1A1 for $\int(y^2/2)dy$,with limits, AEF SC M1A1 for 4/3 obtained from incorrect use of modulus in negative y case
7	(i)		H_0: μ=32, H_1: μ < 32 $s_1^2 = (8586.19 - 289^2/10)/9 = (26.01)$ TS = $(28.9 - 32)/(s/\sqrt{10})$; $s=\sqrt{26.01}$ $= -1.922$ Compare with –1.833 and reject H_0 There is evidence at the 5% SL that spec not met	B1 B1 M1 A1 A1 M1 A1ft **[7]**	Or in words. Need population. [and 7(iii)(b)] AEF eg s=5.1 Allow M1A1 for (32-28.9) etc, but A0 for TS=1.922 Allow 1.922>1.833 (and A1 if earned.) Not too assertive.
7	(ii)		No complaints if mean > spec	B1 **[1]**	Sensible reason
7	(iii)	(a)	Equal population variances required for lives of batteries made in two factories $s_2^2 = (11290.95 - 363^2/12)/11$ =28.2,sample variances are close, so assumption valid	B1 M1 A1 **[3]**	Do not insist on ‘population’ Do not allow if 28.2 first seen in (b)

Question			Answer	Marks	Guidance
7	(iii)	(b)	$H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$	B1	Allow consistent use of '+' instead of '-' throughout
			$s^2 = (9 \times 26.01 + 11 \times 28.2)/20 = 27.2145$	M1 A1	
			$TS = (28.9 - 363/12)/[s(10^{-1} + 12^{-1})^{1/2}]$	M1	Or equivalent method. Allow from "26.01"/10+"28.2"/12. Allow M1 if 9 and/or 11 used instead of 10 and/or 12.Pooled or not.
			$= -0.6044$	A1	Allow A1ft for -0.607 from unpooled sample.
			Compare correctly with –1.725 and do not reject H_0	M1	SC Allow for correct comparison with -1.325 for 1 tail test.
			There is insufficient evidence (at the 10% sig level) that there is a difference in mean life of the batteries.	A1ft	Must have used +/- 1.725
				[7]	