OCR
RECOGNISING ACHIEVEMENT

Thursday 31 May 2012 – Morning

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

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.

1 A machine fills packets of flour whose nominal weights are 500 g. Each of a random sample of 100 packets was weighed and 14 packets weighed less than 500 g. The population proportion of packets that weigh less than 500 g is denoted by p.

(i) Calculate an approximate 95% confidence interval for p. **[4]**

(ii) The weights of the packets, in grams, are normally distributed with mean μ and variance 50. Assuming that p=0.14, calculate the value of μ. **[3]**

2 Four pairs of randomly chosen twins were each given identical puzzles to solve. The times taken (in seconds) are shown in the following table.

Twin pair	1	2	3	4
Time for first-born	46	38	44	49
Time for second-born	40	41	37	46

Stating any necessary assumption, test at the 10% significance level whether there is a difference between the population mean times of first-born and second-born twins. **[9]**

3 A charity raises money by sending letters asking for donations. Because of recent poor responses, the charity's fund-raiser, Anna, decides to alter the letter's appearance and designs two possible alternatives, one colourful and the other plain. She believes that the colourful letter will be more successful. Anna sends 60 colourful letters and 40 plain letters to 100 people randomly chosen from the charity's database. There were 39 positive responses to the colourful letter and 12 positive responses to the plain letter. The population proportions of positive responses to the colourful and plain letters are denoted by p_C and p_P respectively. Test the null hypothesis $p_C - p_P = 0.15$ against the alternative hypothesis $p_C - p_P > 0.15$ at the $2\frac{1}{2}$% significance level and state what Anna could report to her manager. **[6]**

4 The time interval, T minutes, between consecutive stoppages of a particular grinding machine is regularly measured. T is normally distributed with mean μ.
24 randomly chosen values of T are summarised by

$$\sum_{i=1}^{24} t_i = 348.0 \text{ and } \sum_{i=1}^{24} t_i^2 = 5195.5.$$

(i) Calculate a symmetric 95% confidence interval for μ. **[5]**

(ii) For the machine to be working acceptably, μ should be at least 15.0.
Using a test at the 10% significance level, decide whether the machine is working acceptably. **[6]**

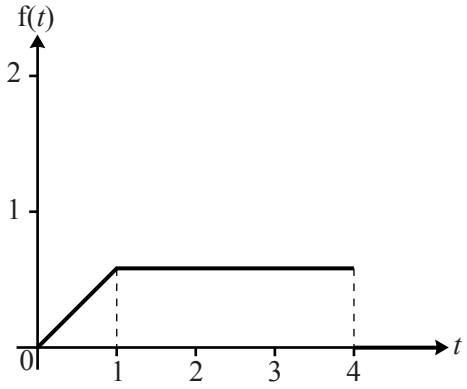
5 The discrete random variables X and Y are independent with $X \sim \text{B}(32, \frac{1}{2})$ and $Y \sim \text{Po}(28)$.

(i) Find the values of $\text{E}(Y - X)$ and $\text{Var}(Y - X)$. **[3]**

(ii) State, with justification, an approximate distribution for $Y - X$. **[3]**

(iii) Hence find $\text{P}(|Y - X| \geqslant 3)$. **[4]**

6



The diagram shows the probability density function f of the continuous random variable T, given by

$$\mathrm{f}(t) = \begin{cases} at & 0 \leqslant t \leqslant 1, \\ a & 1 < t \leqslant 4, \\ 0 & \text{otherwise}, \end{cases}$$

where a is a constant.

(i) Find the value of a. **[2]**

(ii) Obtain the cumulative distribution function of T. **[4]**

(iii) Find the cumulative distribution of Y, where $Y = T^{\frac{1}{2}}$, and hence find the probability density function of Y. **[7]**

7 A study was carried out into whether patients suffering from a certain respiratory disorder would benefit from particular treatments. Each of 90 patients who agreed to take part was given one of three treatments *A, B* or *C* as shown in the table.

Treatment	*A*	*B*	*C*
Number in group	31	25	34

(i) It is claimed that each patient was equally likely to have been given any of the treatments. Test at the 5% significance level whether the numbers given each treatment are consistent with this claim. **[6]**

(ii) After 3 months the numbers of patients showing improvement for treatments *A*, *B* and *C* were 14, 18 and 25 respectively. By setting up a 2 × 3 contingency table, test whether the outcome is dependent on the treatment. Use a 5% significance level. **[8]**

(iii) If one of the treatments is abandoned, explain briefly which it should be. **[2]**

Question			Answer	Marks	Guidance	
1		(i)	$0.14 \pm zs$ $z = 1.96$ $s^2 = 0.14 \times 0.86/100$ (0.072(0), 0.2080)	M1 B1 A1 A1 **[4]**	 Or /99 (0.0716, 0.208(4)) from /99 A0	
		(ii)	Equate $\phi((500 - \mu)/\sqrt{(50)})$ to 0.14 $\Rightarrow (500 - \mu)/\sqrt{(50)} = -1.0803$ $\Rightarrow \mu = 508$ (3SF)	M1 A1 A1 **[3]**		
2			Assumes population of time differences is normal $H_0: \mu_F = \mu_S$, $H_1: \mu_F \neq \mu_S$ Differences 6 -3 7 3 Sample mean = 3.25 Sample variance = 4.5^2 = 20.25 TS = 3.25/("4.5"/2) = 1.44(4) Not $\geq$ 2.353 do not reject H_0 There is insufficient evidence of a difference in the mean times of the two populations	B1 B1 M1 A1 A1 M1 A1 M1 A1 ft **[9]**	Need population. Allow μ_d=0 etc. One tail B1B0M1A1A1M1A1M1(CV=1.638)A0 Comparison with correct t and correct first conclusion ft. Ft TS Not over-assertive.	If 2 sample test used. Popns N,equal var B1 Hypotheses B1 Means 44.25,41 B1 (vars 21.583,14) (Var=(3x21.583+3x14)/6 M0 =17.792 A0) TS=("44.25"- "41")/√"17.792"(¼+¼)M1 1.09 A0 "1.09"<1.943 do not reject NH. M1 Conclusion A1√ Max 6/9

Question			Answer	Marks	Guidance	
3			TS = (39/60 – 12/40 – 0.15)/s	M1	No marks for hyps in this Q.	
			s^2 = 0.65 × 0.35/60 + 0.3 × 0.7/40 (=0.0090417 (5sf))	M1A1	M1 for "variance" involving 40 and 60	eg "pooled" s^2=0.0104125 TS=1.95998 A0A0
			TS = 2.10(3)	A1		
			Compare TS with 1.96 and reject H_o	M1	Or "1.78%"<2.5%	
			Reports that there is evidence (at 5% SL) that. popn resp to coloured more than 15% greater than popn resp to plain	A1ft **[6]**	Ft TS Not over-assertive	
4		(i)	$\bar{t}$ = 348 / 24 (=14.5)	B1		
			s^2 = (5195.5 – 348^2/24)/23(=6.5)	B1		
			348/24 ± t s /√(24)	M1	Allow z=1.96	
			t = 2.069	B1		
			(13.423, 15.557)	A1 **[5]**	Rounding to (13.4, 15.6)	
		(ii)	H_0:μ =15 (Or ≥), H_1:μ < 15	B1 B1		
			α : (14.5 – 15)/√("6.5"/24)	M1	Must have /24	15-14.5 etc M1A0
			= –0.961	A1		
			> –1.319 , do not reject H_0	M1	t-value	"0.961"<1.319 M1 A1 possible
			β: $\bar{x} \le 15 - 1.139\sqrt{6.5/24}$	M1	t-value	
			=14.3(1)	A1		
			< 14.5	M1		
			Insufficient evidence at the 5% SL that the machine is not working acceptably	A1ft **[6]**	ft-TS, not over-assertive	

Question			Answer	Marks	Guidance	
5		(i)	(E(X)=16, E(Y)=28) E(Y-X)=12 (Var(X)=8, Var(Y)=28) 28+32x½x½ Var($Y - X$) = 36	B1 M1 A1 **[3]**		
		(ii)	($X \approx$ N(16,8)) since 32 × ½ > 5 and 32 × ½ × ½ > 5 ($Y \approx$ N(28,28)) since E(Y) = 28 is large oe (eg>15) $Y - X \approx$ N(12, 36)	 M1 M1 B1ft **[3]**	 from $np > 5$ and $nq > 5$ or npq>5 N(parameters from (i))	
		(iii)	P($Y - X \leq -2.5$)+P($Y - X \geq 2.5$) = ϕ(–2.417)+ 1 – ϕ(–1.583) 0.0079+0.9433=0.951(2)	M1 A1 M1 A1 **[4]**	 (-2.417,-1.583) Correct use of z and Φ for their value(s) Allow M1A0 if only 1 inerval ,except M1A1 if 0.0079 or 0.9433 seen (SC)	No CC or wrong CC M1 ±3 or ±3.5 instead of ±2.5 (-2.5,-1.5) or (-2.583,-1.417) A1M1 as main scheme A0

Question			Answer	Marks	Guidance	
6		(i)	Total area from $t = 0$ to 4 is 1 and ½ a+ 3a=1 Solve to give a = 2/7	M1 A1 **[2]**	Any method	
6		(ii)	0 t<0 t²/7 0≤t≤1 "1/7"+ $\int_1^t 2dt/7 = 2t/7 - 1/7$ 1 t>4	 B1 M1 A1 B1 **[4[**	 Ft a for B1 and A1 For t<0 and t>4	
		(iii)	G(Y) = P($Y \leq y$) = P($T \leq y^2$) =F(y²) $= \begin{cases} (0 & y<0,) \\ y^4/7 & 0 \leq y \leq 1, \\ (2y^2-1)/7 & 1<y \leq 2, \\ (1 & y \geq 2) \end{cases}$ $g(y) = G'(y) = \begin{cases} 4y^3/7 & 0 \leq y \leq 1 \\ 4y/7 & 1<y \leq 2 \\ (0 & \text{otherwise}) \end{cases}$	M1 A1 M1 A1 ft M1 A1 B1 **[7]**	Allow < Seen or implied Possible to score M0A0,then M1A1M1A1B1 , For using F correctly For correct expressions ft a and F(t) For differentiating Correctly,allow from eg 2y²/7 Correct ranges for y seen	$\int_0^1 2tdt/7 + \int_1^4 2dt/7 = 1$ M1 2ydy=dt oe M1 $(\int 4y^3 dy/7 + \int 4ydy/7 = 1)$ g(y)=4y³/7,4y/7 A1 0 ≤ y ≤1, 1 < y ≤ 2 B1 G(y)=y⁴/7 B1 2y²/7+c and G(1)=1/7 M1 (2y²-1)/7 A1

Question			Answer	Marks	Guidance	
7		(i)	$H_0: p_1 = p_2 = p_{3(}=1/3)$, H_1: Not all equal oe eg in words	B1	p= $\frac{1}{3}$ only insufficient.	
			E-values all 30	B1		
			$\chi^2 = (1^2 + 5^2 + 4^2)/30$	M1		
			= 1.4	A1	Accept 1.39,1.399 etc	
			(Critical value = 5.991)			
			Compare χ^2 with CV and do not reject H_0. Insufficient evidence that groups not randomly chosen.oe	M1 A1ft **[6]**	With valid comparison or Accept that groups randomly chosen.	
		(ii)	(H_0: Outcome indep of treatment)			If incorrect contingency table attempted ,eg 31 25 34 14 18 25
			Treatment A B C			
			Improved 14 18 25 Not Improved 17 7 9	B1	For correct contingency table	M1s available. Max 3/8
			E-values 19.6 15.8 21.5 11.4 9.2 12.5	M1 A1	At least 2 correct FT table All correct CAO	(TS=1.61)
			$\chi^2 = 5.6^2(19.6^{-1} + 11.4^{-1}) + 2.2^2(15.8^{-1} + 9.2^{-1})$ $+ 3.5^2(21.5^{-1} + 12.5^{-1})$	M1 A1	At least 2 correct ft All correct	
			= 6.7	A1		
			> 5.991 and reject H_0	M1		
			Sufficient evidence at the 5% SL that the outcome depends on treatment	A1ft **[8]**	Ft TS if attempt at correct table made.	
		(iii)	Treatment A shows fewer improved than expected, Treatments B and C show more	M1	Or consider proportion(s) improved, 0.45, 0.72, 0.74 (M1)	SC B1 if A chosen with no proportions of successful treatments given.
			So abandon Treatment A	A1 **[2]**		