OCR
RECOGNISING ACHIEVEMENT

Wednesday 23 January 2013 – Morning

A2 GCE MATHEMATICS

4734/01   Probability & Statistics 3

QUESTION PAPER

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.

Turn over

1 The independent random variables X and Y have the distributions N(10, σ^2) and Po(2) respectively. The random variable S is given by $S = 5X - 2Y + c$, where c is a constant.
It is given that $E(S) = \text{Var}(S) = 408$.

    (i) Find the value of c and show that $\sigma^2 = 16$. **[4]**

    (ii) Find $P(X \geqslant E(Y))$. **[2]**

2 A new running track has been developed and part of the testing procedure involves 7 randomly chosen athletes. They each run 100 m on both the old and new tracks.
The results are as follows.

Athlete	1	2	3	4	5	6	7
Time on old track (s)	12.2	10.3	11.5	13.0	11.8	11.7	11.9
Time on new track (s)	11.1	10.5	11.0	12.6	11.0	10.9	12.0

The population mean times on the old and new tracks are denoted by μ_O seconds and μ_N seconds respectively. Stating any necessary assumption, carry out a suitable t-test of the null hypothesis $\mu_O - \mu_N = 0$ against the alternative hypothesis $\mu_O - \mu_N > 0$. Use a 2½% significance level . **[8]**

3 Two reading schemes, A and B, are compared by using them with a random sample of 9 five-year-old children. The children are divided into two groups, 5 allotted to scheme A and 4 to scheme B, and the schemes are taught under similar conditions.
After one year the children are given the same test and their scores, x_A and x_B, are summarised below. With the usual notation,

$$n_A = 5, \; \bar{x}_A = 52.0, \; \Sigma(x_A - \bar{x}_A)^2 = 248,$$

$$n_B = 4, \; \bar{x}_B = 56.5, \; \Sigma(x_B - \bar{x}_B)^2 = 381.$$

It may be assumed that scores have normal distributions.

    (i) Calculate an 80% confidence interval for the difference in population mean scores for the two methods. **[6]**

    (ii) State a further assumption required for the validity of the interval. **[1]**

4 The continuous random variable X has probability density function given by

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

The random variable Y is given by $Y = \dfrac{1}{\sqrt{X}}$.

   (i) Find the (cumulative) distribution function of Y, and hence show that its probability density function is given by

$$g(y) = \frac{3}{y^4},$$

for a set of values of y to be stated. **[7]**

   (ii) Find the value of $E(Y^2)$. **[2]**

5 A constitutional change was proposed for a Golf Club with a large membership. This was to be voted on at the Annual General Meeting. A month before this meeting the secretary asked a random sample of 50 members for their opinions. Out of the 50 members 70% said they approved.

   (i) Calculate an approximate 90% confidence interval for the proportion p of all members who would approve the proposal. **[4]**

   (ii) Explain what is meant by a 90% confidence interval in this context. **[1]**

   (iii) Nearer the date of the meeting the secretary asked a random sample of n members, and, as before, 70% said they approved. This time the secretary calculated an approximate 99% confidence interval for p. It is given that the confidence interval does not include 0.85. Find the smallest possible value of n. **[4]**

6 A large population of plants consists of five species A, B, C, D and E in the proportions p_A, p_B, p_C, p_D and p_E respectively. A random sample of 120 plants consisted of 23, 14, 24, 27 and 32 of A, B, C, D and E respectively. Carry out a test at the 10% significance level of the null hypothesis that the proportions are $p_A = p_B = 0.15$, $p_C = p_D = 0.25$ and $p_E = 0.2$. **[7]**

7 The random variable X has distribution $N(\mu, 1)$. A random sample of 4 observations of X is taken. The sample mean is denoted by $\overline{X}$.

   (i) Find the value of the constant a for which $(\overline{X} - a, \overline{X} + a)$ is a 98% confidence interval for μ. **[3]**

The independent random variable Y has distribution $N(\mu, 9)$. A random sample of 16 observations of Y is taken. The sample mean is denoted by $\overline{Y}$.

   (ii) Write down the distribution of $\overline{X} - \overline{Y}$. **[2]**

   (iii) A 90% confidence interval for μ based on $\overline{Y}$ is given by $(\overline{Y} - 1.234, \overline{Y} + 1.234)$.
Find the probability that this interval does not overlap with the interval in part **(i)**. **[6]**

[Question 8 is printed overleaf.]

8 After contracting a particular disease, patients from a hospital are advised to have their blood tested monthly for a year. In order to test whether patients comply with this advice the hospital management commissioned a survey of 100 patients. A hospital statistician selected the patients randomly from records and asked the patients whether or not they had complied with the advice. The results classified by gender are as follows.

<table>
<tr><td></td><td></td><td colspan="2" align="center">Gender</td></tr>
<tr><td></td><td></td><td>Female</td><td>Male</td></tr>
<tr><td>Comply</td><td>Yes</td><td>34</td><td>30</td></tr>
<tr><td></td><td>No</td><td>11</td><td>25</td></tr>
</table>

(i) Test at the 5% significance level whether compliance with the advice is independent of gender. **[8]**

(ii) A manager believed that a greater proportion of female patients than male patients comply with the advice. Carry out an appropriate test of proportions at the 10% significance level. **[7]**

Question			Answer	Marks	Guidance	
1	(i)		$E(S)=50 – 4 + c =408$	M1	Using $E(aX+bY+c)$	
			$=> c = 362$	A1		
			$Var(S)=25\sigma^2+8=408$	M1	Using $Var(aX+bY+c)$	
			$=> \sigma^2 = 16$ AG	A1		
				[4]		
1	(ii)		$P(X \geq 2)=P(Z\geq-8/4)$	M1		
			$=0.9772$	A1	$1 – \Phi(-2)$	
				[2]		
2			Use of paired t test			
			d=old-new:1.1 -0.2 0.5 0.4 0.8 0.8 -0.1	M1		
			$\bar{d}$ =0.471(4), $s^2 = 0.2324$	B1B1	Allow 3.3/7, 33/70; 122/525	Or use $\bar{d}/\sqrt{0.2347/7} > 2.447$ B1
			$(H_o: \mu_d=0, H_1:\mu_d >0)$			
			Test statistic $= 0.4714/(0.2324/7)^{0.5}$	M1	Must be /7	
			$= 2.587$	A1		$\bar{d} > 0.448$
			Critical value $=2.447$	B1		$0.4714>0.448$ so reject H_0 M1
			TS > CV so reject H_0	M1	ft wrong TS, CV	There is etc A1
			There is sufficient evidence that the new	A1	Allow from rounding errors, otherwise cwo.	eg s²=0.233 (from using $\bar{d}$ =0.471)
						TS=2.58(2) M1B1B0M1A0B1M1A1
			track is faster		Or similar in context, not over -assertive	
				[8]		
3	(i)					
			Est $\sigma_A{}^2=62$, Est $\sigma_B{}^2 = 127$	B1	Can be implied eg by 89.86. Allow 248/4	381/3
			PEV=(4×62+3×127)/7=[89.86 or 629/7]]	M1		
			$(\bar{x}_A - \bar{x}_B = -4.5$)		(or $\bar{x}_B - \bar{x}_A = 4.5$)	
			SD=$\sqrt{(89.86(5^{-1}+4^{-1}))}$	B1		
			CI=(-4.5-tSD, -4.5+tSD)	M1	Allow M from incorrect 4.5. Must be t, not z.	
			t=1.415	B1		
			(-13.50, 4.50)	A1	Allow (-4.50, 13.50)	
				[6]		
3	(ii)		Scores resulting from two schemes	B1	Allow popn. variances equal.	NOT 'Variances equal'
			should have equal variances			
				[1]		

Question			Answer	Marks	Guidance	
4	(i)		$F(x)=x^{3/2}$ ($0<x\leq1$)	B1		
			$G(y)=P(Y\leq y)$	M1		
			$\quad= P(1/\sqrt{X}\leq y)$			
			$\quad= P(X\geq1/y^2)$			
			$\quad= 1- F(1/y^2)$	M1		
			$\quad= 1 – 1/y^3$ for $y\geq1$	A1,B1	B1 for $y \geq 1$ seen anywhere.	
			($G(y)=0$ otherwise)			
			$g(y) = G'(y) = 3/y^4$ $y\geq1$, AG	M1A1	M1 for differentiating	
			$\quad(= 0$ otherwise)		Allow from $-1/y^3$	
				[7]		
4	(ii)		(α) $\int_1^\infty \dfrac{3}{y^2}\,dy = 3$	M1A1	Ft range from (i) for M1, but not $\int_0^1 \frac{3}{y^2}dy$	
			(β) $\int_0^1 \dfrac{3}{2\sqrt{x}}\,dx = 3$	M1A1		
				[2]		
5	(i)		$z=1.645$	B1		
			$s^2 =0.7\times0.3/50$	B1	or /49	
			Use $0.7\pm zs$	M1		
			to give (0.593, 0.807)	A1	or (0.592, 0.808) using /49; or %	
				[4]		
5	(ii)		If a large number of independent 90% confidence intervals were found for the proportion p of club members who would vote for the amendment then about 90% of the intervals would contain p	B1	Allow with specific number of CIs NOT *the probability of THE interval containing p is* 0.9	If a candidate uses the word 'probability' oe , it must be clear that (s)he is considering a large number of intervals, in context, not just one.
				[1]		
5	(iii)		Use $0.7 + 2.576\sqrt{(0.21/n)}$	M1*	Any z value.	
			< 0.85	A1	Allow $\leq$ or =, correct z	
			Solve to give $n=62$	*M1A1	Allow $\geq$	
				[4]		

	Question		Answer	Marks	Guidance	
6			$(H_0: p_A = p_B = 0.15, p_c = p_D = 0.25, p_E = 0.20)$ H_1: all alternatives)			SC 5 difference of sample proportion from population proportion.
			E values 18 18 30 30 24	B1		z= 1.28, -1.02, 1.26, -0.63, 1.83 B1
			$\gamma^2 = 25/18 + 16/18 + 36/30 + 9/30 + 64/32$	M1A1√	M1 for at least one correct, A1 for all correct	Compare with -1.645 AND 1.645 M1
			=6.444	A1	ft E-values	E shoes a sig. diff., others don't. A1cwo
			(α) CV= 7.779	B1		
			6.444<CV so do not reject H_0	M1		
			(β) P$(\gamma^2 > 6.444)= 0.168$	B1		
			0.168 >0.10 so do not reject H_0	M1		NOT Insuff evidence that the data does not fit the proportions.
			There is insufficient evidence at the 10% significance level that the proportions are not as specified	A1 **[7]**	cwo, except rounding errors.	NOR Evidence that data fits the proportions.
7	**(i)**		$a = z\sqrt{(1/4)}$	M1	From $\sigma/\sqrt{n}$	
			z=2.326	B1		
			a=1.163	A1 **[3]**		
7	**(ii)**		Normal with mean 0	B1		
			$\sigma^2 = {}^1/_4 + {}^9/_{16} = 0.8125$ or 13/16	B1 **[2]**	Allow 0.813	
7	**(iii)**		Either $\bar{x} + 1.163 < \bar{y} - 1.234$	M1	Ft their a for M1M1. Can be implied.	
			or $\bar{y} + 1.234 < \bar{x} - 1.163$	M1		
			giving $\bar{x} - \bar{y} < -2.397$ or >2.397	A1		
			P($<$-2.397)=P($Z<$-2.659) oe	M1	Correct use of N for their interval(s).	
			= 0.0039(2)	A1	0.0039 from tables	
			So required probability =0.0078(4)	A1ft **[6]**	ART 0.0078 ft 2× their 0.0039.	

Question			Answer	Marks	Guidance	

8	(i)		(H_0: attributes are independent, H_0: all alts)
			Expected values
			         F M
			   Y 28.8 35.2
			   N 16.2 19.8
			$\gamma^2 = 4.7^2(28.8^{-1}+16.2^{-1}+35.2^{-1}+19.8^{-1})$
			   = 3.87(381..)
			
			 CV= 3.841
			3.87 > 3.841 so reject H_0
			
			There is sufficient evidence (at the 5% SL) that gender and compliance with the advice are not independent.
			
			[8]
				M1
				A1
				M2
				A1
				
				B1
				M1
				
				A1
					Any one correct, e.g. 45×64/100
					All correct
					M2 with Yates and correct form
					M1 without Yates and correct form
					or no modulus.
					
					Ft TS and CV
					
					cwo, other than rounding errors.
						A1 for 4.74 without Yates
						A1 for 4.81 , no modulus.

8	(ii)		H_0: $p_F = p_M$, H_1:p_F>p_M p_s = (34+30)/(45+55) $z=\dfrac{\frac{34}{45}-\frac{30}{55}}{\sqrt{0.64\times0.36(45^{-1}+55^{-1})}}$ =2.18 >1.282 Reject H_0, there is sufficient evidence at the 2 ½ % significance level that more women than men comply with the advice. **[7]**	B1 B1 M1 B1 A1 M1 A1	Or equivalent Variance must contain 0.64 and 1-0.64 or denominators 45 and 55. For denominator Or P(z>2.18)=0.015< 0.10 ft TS. SR if p_s not used. B1B0M1B0A1(for z=2.26(4)) M1A1 cwo, in context, not over-assertive.	eg $\sqrt{\dfrac{0.64\times0.36}{100}}$ M1B0A0 Max(5/7)