OCR
RECOGNISING ACHIEVEMENT

Monday 24 June 2013 – Afternoon

A2 GCE MATHEMATICS

4735/01 Probability & Statistics 4

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4735/01
- 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

		S 0	1	2
F	0	$\frac{1}{8}$	$\frac{1}{8}$	0
	1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$
	2	0	$\frac{1}{8}$	$\frac{1}{8}$

An unbiased coin is tossed three times. The random variables F and S denote the total number of heads that occur in the first two tosses and the total number of heads that occur in the last two tosses respectively. The table above shows the joint probability distribution of F and S.

(i) Show how the entry $\frac{1}{4}$ in the table is obtained. **[2]**

(ii) Find $\text{Cov}(F, S)$. **[6]**

2 Two drugs, I and II, for alleviating hay fever are trialled in a hospital on each of 12 volunteer patients. Each received drug I on one day and drug II on a different day. After receiving a drug, the number of times each patient sneezed over a period of one hour was noted. The results are given in the table.

Patient	1	2	3	4	5	6	7	8	9	10	11	12
Drug I	11	34	19	16	10	29	6	17	20	13	4	25
Drug II	12	20	10	18	3	21	9	13	10	19	9	12

The patients may be considered to be a random sample of all hay fever sufferers.
A researcher believes that patients taking drug II sneeze less than patients taking drug I.

Test this belief using the Wilcoxon signed rank test at the 5% significance level. **[7]**

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

$$f(x) = \begin{cases} \frac{1}{4}xe^{-\frac{1}{2}x} & x \geqslant 0, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Show that the moment generating function of X is $(1-2t)^{-2}$ for $t < \frac{1}{2}$, and state why the condition $t < \frac{1}{2}$ is necessary. **[6]**

(ii) Use the moment generating function to find $\text{Var}(X)$. **[3]**

4 The effect of water salinity on the growth of a type of grass was studied by a biologist. A random sample of 22 seedlings was divided into two groups *A* and *B*, each of size 11.
Group *A* was treated with water of 0% salinity and group *B* was treated with water of 0.5% salinity. After three weeks the height (in cm) of each seedling was measured with the following results, which are ordered for convenience.

Group *A*	8.6	9.4	9.7	9.8	10.1	10.5	11.0	11.2	11.8	12.7	12.9
Group *B*	7.4	8.4	8.5	8.8	9.2	9.3	9.5	9.9	10.0	11.1	11.3

Jeffery was asked to test whether the two treatments resulted, on average, in a difference in growth. He chose the Wilcoxon rank sum test.

(i) Justify Jeffery's choice of test. **[1]**

(ii) Carry out the test at the 5% significance level. **[9]**

5 The discrete random variable U has probability distribution given by

$$\mathrm{P}(U=r)=\begin{cases}\frac{1}{16}\binom{4}{r} & r=0,1,2,3,4,\\ 0 & \text{otherwise.}\end{cases}$$

(i) Find and simplify the probability generating function (pgf) of U. **[3]**

(ii) Use the pgf to find $\mathrm{E}(U)$ and $\mathrm{Var}(U)$. **[4]**

(iii) Identify the distribution of U, giving the values of any parameters. **[2]**

(iv) Obtain the pgf of Y, where $Y=U^2$. **[2]**

(v) State, giving a reason, whether you can obtain the pgf of $U+Y$ by multiplying the pgf of U by the pgf of Y. **[1]**

6 The continuous random variable X has mean μ and variance σ^2, and the independent continuous random variable Y has mean 2μ and variance $3\sigma^2$. Two observations of X and three observations of Y are taken and are denoted by X_1, X_2, Y_1, Y_2 and Y_3 respectively.

(i) Find the expectation of the sum of these 5 observations and hence construct an unbiased estimator, T_1, of μ. **[3]**

(ii) The estimator T_2, where $T_2=X_1+X_2+c(Y_1+Y_2+Y_3)$, is an unbiased estimator of μ.
Find the value of the constant c. **[2]**

(iii) Determine which of T_1 and T_2 is more efficient. **[4]**

(iv) Find the values of the constants a and b for which

$$a(X_1^2+X_2^2)+b(Y_1^2+Y_2^2+Y_3^2)$$

is an unbiased estimator of σ^2. **[4]**

7 Each question on a multiple-choice examination paper has n possible responses, only one of which is correct. Joni takes the paper and has probability p, where $0 < p < 1$, of knowing the correct response to any question, independently of any other. If she knows the correct response she will choose it, otherwise she will choose randomly from the n possibilities. The events K and A are 'Joni knows the correct response' and 'Joni answers correctly' respectively.

(i) Show that $\mathrm{P}(A) = \dfrac{q + np}{n}$, where $q = 1 - p$. **[3]**

(ii) Find $P(K|A)$. **[3]**

A paper with 100 questions has $n = 4$ and $p = 0.5$. Each correct response scores 1 and each incorrect response scores -1.

(iii) (a) Joni answers all the questions on the paper and scores 40. How many questions did she answer correctly? **[1]**

(b) By finding the distribution of the number of correct answers, or otherwise, find the probability that Joni scores at least 40 on the paper using her strategy. **[6]**

Question			Answer	Marks	Guidance
1	(i)		F = 1, S = 1 requires HTH or THT Probability = $^1/_8 + {}^1/_8 = {}^1/_4$ AG	M1 A1 **[2]**	Clear method – not just multiplication of probs SC $\frac{2}{8}=\frac{1}{4}$ ONLY seen B1. NOT $\frac{1}{2}\times\frac{1}{2}$
1	(ii)		Marginals : 0 1 2 p(S) ¼ ½ ¼ p(F) ¼ ½ ¼ E(S) = 1×½ + 2×¼ = 1 = E(F) E(SF) = $\frac{1}{4}+\frac{2}{8}+\frac{2}{8}+\frac{4}{8}(=1\frac{1}{4})$ Cov(S,F) = E(SF) – E(S)E(F) = ¼	M1 A1 B1 M1* *M1 A1 **[6]**	Correct method, can be implied by (i) Both correct. Can be implied by e.g. E(S) = E(F) = 1 or symmetry.
2			H_0: $m_{II–I}$ = 0, H_1: $m_{II–I}$ < 0 II–I: 1, –14, –9, 2, –7, –8, 3, –4, –10, 6, 5, 13 Rank: 1, –12, –9, 2, –7, –8, 3, –4, –10, 6, 5, 11 P = 1 + 2 + 3 + 6 + 5 = 17 Q = 61 so T = 17 5% CR: $T \le 17$ T is inside CR so reject H_0 There is sufficient evidence at the 5% SL that drug II is associated with fewer sneezes	B1 M1 A1 M1A1 M1 A1 **[7]**	Allow $m_1 > m_2$, $m_d > 0$, etc. or in words, but needs to be in terms of parameters or population ft TS & CV ft TS only. Contextualised, not over-assertive.

Question			Answer	Marks	Guidance
3	(i)		$M(t) = \int_0^\infty \frac{1}{4} x e^{-\frac{1}{2}x(1-2t)} dx$ oe	M1*	From $E(e^{xt})$. Need single exponential term, not nec. correct.
			$= \left[\frac{-xe^{-\frac{1}{2}x(1-2t)}}{2(1-2t)}\right]_0^\infty + \int_0^\infty \frac{e^{-\frac{1}{2}x(1-2t)}}{2(1-2t)} dx$	*M1A1	Integration by parts
			$= \left[\frac{-e^{-\frac{1}{2}x(1-2t)}}{(1-2t)^2}\right]_0^\infty$	A1	$= \left[\frac{-e^{-\frac{1}{2}x(1-2t)}}{4(t-\frac{1}{2})^2}\right]_0^\infty$. Allow without limits.
			= AG $(1-2t)^{-2}$	A1	With evidence, cwo
			Requires $1 - 2t > 0$ for correct limits	B1 **[6]**	Or for convergence of the integral
3	(ii)		$M'(t) = 4(1-2t)^{-3}$ $E(X) = 4$ cwo $M''(t) = 24(1-2t)^{-4}$ $E(X^2) = 24$ cwo Var = 24 – 16 = 8	B1 B1 B1FT **[3]**	or from $1 + 4t$ $+ 12t^2$ provided Var > 0.
4	(i)		Distribution of heights may not be normal/is unknown	B1 **[1]**	Allow "No assumption required", but nothing else Not "groups independent" unless something else as well
4	(ii)		H_0: $m_A = m_B$, H_1: $m_A \neq m_B$	B1	Medians. Allow words in context. Not μ unless "median" stated
			Ranks: A: 4, 8, 10, 11, 14, 15, 16, 18, 20, 21, 22 B: 1, 2, 3, 5, 6, 7, 9, 12, 13, 17, 19	B1	
			$m = n = 11$, $R_m = 159$ or 94	B1	
			Use normal approximation with mean 126.5 [= 253/2]	M1	allow $\frac{1}{2}\times 11\times(11+11+1)$
			Variance 231.92 [= 2783/12]	B1	allow $\frac{1}{12}\times 11\times 11\times(11+11+1)$
			(α) $P(\le 94) = \Phi((94.5 - 126.5)/\sqrt{231.92})$ or $P(\ge 159) = 0.0178$ < 0.025 and reject H_0	M1 A1 M1	Standardising. Allow no/incorrect cc. Value ft TS
			(β) $z = (94.5 - 126.5)/\sqrt{231.92} = -2.101$ < –1.96 so reject H_0	M1A1 M1	Standardising ; value ft TS
			There is evidence that salinity affects growth	A1 **[9]**	Or equivalent in context. ft TS.

Question			Answer	Marks	Guidance
5	(i)		$G(t) = E(t^X) = {}^1/_{16}(1 + 4t + 6t^2 + 4t^3 + t^4)$ $= {}^1/_{16}(1 + t)^4$	M1A1 A1 **[3]**	Correct form ; correct coefficients allow $(\frac{1}{2}+\frac{1}{2}t)^4$.
5	(ii)		$G'(t) = ¼(1 + t)^3$ $E(U) = G'(1)$ $\mathbf{= 2}$ $G''(t) = {}^3/_4(1 + t)^2$ $Var(U) = G''(1) + G'(1) – (G'(1))^2$ $= 3 + 2 – 4$ $\mathbf{= 1}$	M1 A1 M1 A1 **[4]**	or expanded form. No marks from part (iii) Finding G" and formula correct
5	(iii)		B(4, ½)	B1 B1 **[2]**	Binomial Parameters
5	(iv)		$Y = 0 \quad 1 \quad 4 \quad 9 \quad 16$ $G_Y(t)= {}^1/_{16} + {}^1/_4t + {}^3/_8t^4 + {}^1/_4t^9 + {}^1/_{16}t^{16}$	B1 B1 **[2]**	Values of Y
5	(v)		No, U and Y are not independent	B1 **[1]**	
6	(i)		$E(T_1) = 2E(X) + 3E(Y)$ $= 8\mu$ Unbiased estimate = $(X_1 + X_2 + Y_1 + Y_2 + Y_3)/8$	M1 A1 A1 **[3]**	NOT $\frac{2x+3y}{8}$
6	(ii)		$E(T_2) = 2\mu + 6c\mu = \mu$ $\Rightarrow c = -{}^1/_6$	M1 A1 **[2]**	Setting up an equation

Question			Answer	Marks	Guidance
6	(iii)		$\text{Var}(T_1) = (2\sigma^2+9\sigma^2)/64$ $= {}^{11}/_{64}\sigma^2$ $\text{Var}(T_2) = \sigma^2 + \sigma^2 + {}^1/_{36}(3\sigma^2 + 3\sigma^2 + 3\sigma^2) = {}^9/_4\sigma^2$ T_1 has the smaller variance so is more efficient	M1 A1 A1 A1ft **[4]**	Using var of sum = sum of var
6	(iv)		$E(T_3) = a(2\sigma^2 + 2\mu^2) + b(9\sigma^2 + 12\mu^2) = \sigma^2$ Coefficient of $\mu^2 = 0$ gives $2a + 12b = 0$ Coefficient of $\sigma^2 = 1$ gives $2a + 9b = 1$ Solve to give $a = 2$ and $b = -{}^1/_3$	M1A1 B1 A1 **[4]**	$(\text{Var}(X) =)\ E(X^2) - [E(X)]^2$ seen or implied: M1 either equation.
7	(i)		$P(A) = P(K)\times 1 + P(K')\times{}^1/_n$ $= p + (1-p)/n$ $= \dfrac{q+np}{n}$ AG	M1 A1 B1 **[3]**	 allow $p+\dfrac{q}{n}$
7	(ii)		$P(K \cap A) = p$ $P(K\|A) = \dfrac{p}{\frac{q+np}{n}}$ $= \dfrac{np}{q+np}$	B1 M1 A1 **[3]**	 AEF
7	(iii)	(a)	If X answers are correct $100 - X$ are incorrect so score $= 2X - 100 = 40$ giving $X = 70$	B1 **[1]**	70 seen

Question			Answer	Marks	Guidance
7	(iii)	(b)	P(A) = 5/8		
			(α) E(X) = $100 \times {}^5/_8 = 62.5$	B1	
			Var(X) = $s^2 = 100 \times {}^5/_8 \times {}^3/_8$ (= 23.4375) ($= \frac{375}{16}$)	M1A1	Allow M1 from wrong p
			P($X \geq 70$) = $1 - \Phi[(69.5 - 62.5)/s]$	M1A1	Normal approximation. Allow M1 from 40/70 or wrong p
			= 0.0741	A1	Standardise M1 only if no or wrong cc, A1 for 0.0607
			(β) E($2X - 100$) = 25	B1	
			Var($2X - 100$) = 93.75	M1A1	
			P($2X - 100 \geq 40$) = $1 - \Phi[(39 - 25)/\sqrt{(93.75)}]$	M1A1	Standardise, M1 only for no or wrong cc, A1 for 0.0671
			= 0.0741	B1	
			(γ) Score per question = S		
			E(S) = $1 \times {}^5/_8 - 1 \times {}^3/_8 = ¼$	B1	
			Var(S) = $1^2 \times {}^5/_8 + 1^2 \times {}^3/_8 - ({}^1/_4)^2$	M1A1	
			Total, $T \sim N(25, 93.75)$		
			P($T \geq 40$) = $1 - \Phi[39 - 25)/\sqrt{(93.75)}]$	M1A1	As for β
			= 0.0741	B1	
				[6]	