

ADVANCED SUBSIDIARY GCE MATHEMATICS

4732/01

Probability & Statistics 1

TUESDAY 15 JANUARY 2008

Morning

Time: 1 hour 30 minutes

Additional materials: Answer Booklet (8 pages)

List of Formulae (MF1)

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet.
- Read each question carefully and make sure you know what you have to do before starting your answer.
- Answer all the questions.
- 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.
- You are permitted to use a graphical calculator in this paper.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets [] at the end of each question or part question.
- The total number of marks for this paper is 72.
- You are reminded of the need for clear presentation in your answers.

This document consists of 6 printed pages and 2 blank pages.

© OCR 2008 [K/102/2696]

OCR is an exempt Charity

[Turn over

1	(i) The letters A, B, C, D and E are arranged in a straight line.
	(a) How many different arrangements are possible? [2]
	(b) In how many of these arrangements are the letters A and B next to each other? [3]
	(ii) From the letters A, B, C, D and E, two different letters are selected at random. Find the probability that these two letters are A and B. [2]
2	A random variable T has the distribution $Geo(\frac{1}{5})$. Find
	(i) $P(T=4)$, [2]
	(ii) $P(T > 4)$, [2]
	(iii) $E(T)$.
3	A sample of bivariate data was taken and the results were summarised as follows.
	$n = 5$ $\Sigma x = 24$ $\Sigma x^2 = 130$ $\Sigma y = 39$ $\Sigma y^2 = 361$ $\Sigma xy = 212$
	(i) Show that the value of the product moment correlation coefficient r is 0.855, correct to 3 significant figures. [2]
	(ii) The ranks of the data were found. One student calculated Spearman's rank correlation coefficient r_s , and found that $r_s = 0.7$. Another student calculated the product moment coefficient, R , of these ranks. State which one of the following statements is true, and explain your answer briefly.
	(A) $R = 0.855$
	 (B) R = 0.7 (C) It is impossible to give the value of R without carrying out a calculation using the original data.
	[2]
	(iii) All the values of x are now multiplied by a scaling factor of 2. State the new values of r and r_s . [2]
4	A supermarket has a large stock of eggs. 40% of the stock are from a firm called Eggzact. 12% of the stock are brown eggs from Eggzact.
	An egg is chosen at random from the stock. Calculate the probability that
	(i) this egg is brown, given that it is from Eggzact, [2]
	(ii) this egg is from Eggzact and is not brown. [2]

© OCR 2008 4732/01 Jan08

5 (i) 20% of people in the large town of Carnley support the Residents' Party. 12 people from Carnley are selected at random. Out of these 12 people, the number who support the Residents' Party is denoted by U.

Find

(a)
$$P(U \le 5)$$
, [2]

(b)
$$P(U \ge 3)$$
. [3]

(ii) 30% of people in Carnley support the Commerce Party. 15 people from Carnley are selected at random. Out of these 15 people, the number who support the Commerce Party is denoted by *V*.

Find
$$P(V = 4)$$
. [3]

6 The probability distribution for a random variable Y is shown in the table.

у	1	2	3	
P(Y=y)	0.2	0.3	0.5	

(i) Calculate E(Y) and Var(Y).

[5]

Another random variable, Z, is independent of Y. The probability distribution for Z is shown in the table.

z	z 1		3	
P(Z=z)	0.1	0.25	0.65	

One value of Y and one value of Z are chosen at random. Find the probability that

(ii)
$$Y + Z = 3$$
, [3]

(iii)
$$Y \times Z$$
 is even. [3]

- 7 (i) Andrew plays 10 tennis matches. In each match he either wins or loses.
 - (a) State, in this context, two conditions needed for a binomial distribution to arise. [2]
 - (b) Assuming these conditions are satisfied, define a variable in this context which has a binomial distribution. [1]
 - (ii) The random variable X has the distribution B(21, p), where 0 .

Given that
$$P(X = 10) = P(X = 9)$$
, find the value of p. [5]

8 The stem-and-leaf diagram shows the age in completed years of the members of a sports club.

Male	Female				
8876	1	Female 66677889 1334578899 23347 018 0			
76553321	2	1334578899			
98443	3	23347			
5 2 1	4	018			
90	5	0			

Key: 1 | 4 | 0 represents a male aged 41 and a female aged 40.

- (i) Find the median and interquartile range for the males. [3]
- (ii) The median and interquartile range for the females are 27 and 15 respectively. Make two comparisons between the ages of the males and the ages of the females. [2]
- (iii) The mean age of the males is 30.7 and the mean age of the females is 27.5, each correct to 1 decimal place. Give one advantage of using the median rather than the mean to compare the ages of the males with the ages of the females.

A record was kept of the number of hours, X, spent by each member at the club in a year. The results were summarised by

$$n = 49$$
, $\Sigma(x - 200) = 245$, $\Sigma(x - 200)^2 = 9849$.

(iv) Calculate the mean and standard deviation of X. [6]

© OCR 2008 4732/01 Jan08

9 It is thought that the pH value of sand (a measure of the sand's acidity) may affect the extent to which a particular species of plant will grow in that sand. A botanist wished to determine whether there was any correlation between the pH value of the sand on certain sand dunes, and the amount of each of two plant species growing there. She chose random sections of equal area on each of eight sand dunes and measured the pH values. She then measured the area within each section that was covered by each of the two species. The results were as follows.

	Dune	A	В	С	D	Ε	F	G	Н
	pH value, x	8.5	8.5	9.5	8.5	6.5	7.5	8.5	9.0
Area, y cm ² ,	Species P	150	150	575	330	45	15	340	330
covered	Species Q	170	15	80	230	75	25	0	0

The results for species P can be summarised by

$$n = 8$$
, $\Sigma x = 66.5$, $\Sigma x^2 = 558.75$, $\Sigma y = 1935$, $\Sigma y^2 = 711275$, $\Sigma xy = 17082.5$.

- (i) Give a reason why it might be appropriate to calculate the equation of the regression line of y on x rather than x on y in this situation. [1]
- (ii) Calculate the equation of the regression line of y on x for species P, in the form y = a + bx, giving the values of a and b correct to 3 significant figures. [4]
- (iii) Estimate the value of y for species P on sand where the pH value is 7.0. [2]

The values of the product moment correlation coefficient between x and y for species P and Q are $r_P = 0.828$ and $r_Q = 0.0302$.

- (iv) Describe the relationship between the area covered by species Q and the pH value. [1]
- (v) State, with a reason, whether the regression line of y on x for species P will provide a reliable estimate of the value of y when the pH value is

(vi) Assume that the equation of the regression line of y on x for species Q is also known. State, with a reason, whether this line will provide a reliable estimate of the value of y when the pH value is 8.

© OCR 2008 4732/01 Jan08

4732 Probability & Statistics 1

Note: "(3 sfs)" means "answer which rounds to ... to 3 sfs". If correct ans seen to \geq 3 sfs, ISW for later rounding

Penalise over-rounding only once in paper.

	ver-rounding only once in <u>paper</u> .	1	1
1ia	$5! \text{ or } {}^5P_5$	M1	
	= 120	A1 2	
b	4! or ⁴ P ₄ seen	M1	or $2 \times 3!$ or $2! \times 3!$ or $2! \times {}^{3}P_{3}$
U	4! × 2	M1dep	$\begin{array}{c} 3! \times 3! \times 4 \\ 2 \times 3! \times 4 \end{array}$
	48	A1 3	2 ^ 3! ^ 4
			3.41.0.35
ii	$^{1}/^{5}C_{2} \text{ or } ^{1}/_{5} \times ^{1}/_{4} \times 2 \text{ or } 0.4 \times 0.25 \text{ or } ^{2}/_{5P2}$	M1	Allow M1 for ${}^{5}C_{2}$ or ${}^{1}/_{5}$ x ${}^{1}/_{4}$ or ${}^{1}/_{20}$
			or $\frac{1}{5} \times \frac{1}{5} \times 2$ or $\frac{2}{25}$ oe
	$= \frac{1}{10}$	A1 2	
Total		7	
2i	$(^{4}/_{5})^{3} \times (^{1}/_{5})$ oe	M1	Allow M1 for $({}^{4}/_{5})^{4}$ x $({}^{1}/_{5})$
21	$= \frac{64}{625} \text{ or } 0.102 \text{ (3 sfs)}$		7 thow 1411 for (75) X (75)
		A1 2	4/3 4/3
ii	$(4/5)^4$ alone		Allow $(^4/_5)^3$ or $(^4/_5)^5$; not $1 - (^4/_5)^4$
	or $1 - (\frac{1}{5} + \frac{4}{5}x^{\frac{1}{5}} + (\frac{4}{5})^{2}x^{\frac{1}{5}} + (\frac{4}{5})^{3}x^{\frac{1}{5}})$	M1	Allow one term omitted or wrong
			or "correct" extra
	$= {}^{256}/_{625}$ or 0.410 (3 sfs)	A1 2	Allow 0.41
iii	5	B1 1	
Total	24.00	5	240 240 240
3i	$212 - \frac{24 \times 39}{}$		$\frac{24.8}{\sqrt{14.8 \times 56.8}}$ or $\frac{24.8}{\sqrt{840.64}}$ or $\frac{24.8}{3.85 \times 7.54}$ or $\frac{24.8}{29}$
	5		V14.8×30.8 V840.04 3.03×7.34 27
	$r = \frac{1}{\sqrt{2a^2 + 2a^2}}$	B2 2	
	$(130 - \frac{24^2}{361 - \frac{39^2}{39}})$		B2 for correct subst in r
	$r = \frac{212 - \frac{24 \times 39}{5}}{\sqrt{(130 - \frac{24^2}{5})(361 - \frac{39^2}{5})}}$		B1 for correct subst in any S
ii	R = 0.7 or (B)	B1	(A) and (B) true: B0B0
11		B1 2	
	Definition of r_s is PMCC for ranks		dep 1 st B1
iii	r = 0.855	B1	
	$r_{\rm s} = 0.7$	B1 2	or "unchanged": B1B1
			Interchanged: B1
Total		6	
4i	$0.4 \times p = 0.12$ or $0.12/0.4$ or $12/40$ oe	M1	
	p = 0.3 oe	A1 2	
ii			an 0.4
11	0.4 x (1 – their 0.3) oe eg $^{40}/_{100} \times ^{28}/_{40}$	M1	or 0.4 – 0.12 or 0.28 or 28 seen
			Not 0.4×0.88 unless ans to (i) is 0.12
	0.28 or 28% oe	A1ft 2	
Total		4	
5ia	Binomial stated or implied	B1	by use of tables or $0.2^{a} \times 0.8^{b}$, $a+b=12$
	0.9806	B1 2	19 11 11 11 11 11 11 11 11 11 11 11 11 1
b	0.5583 seen	M1	add 10 corr terms or 1-(add 3 corr terms):
υ			` '
	1 - 0.5583	M1	M2
			or 1–0.7946 or 0.205 or 1-0.6774 or 0.323
	= 0.442 (3 sfs)	A1 3	or 1-0.3907 or 0.609
	, ´		or add 9 terms or 1-(add 2 or 4
			terms): M1
::	150 - 24 - 0 711	NAO	15C = 0.211 = 0.74 · M1
ii	$^{15}C_4 \times 0.3^4 \times 0.7^{11}$	M2	¹⁵ C ₄ x 0.3 ¹¹ x 0.7 ⁴ : M1
	= 0.219 (3 sfs)	A1 3	
Total		8	
	L	1	

6i	Σγρ	M1	$>$ 2 terms added \div 3 or \div 6 etc: M0
	=2.3	A1	
	$\Sigma y^2 p$ (= 5.9)	M1	\geq 2 terms added \div 3 or \div 6 etc: M0
	$-(\Sigma yp)^2$	M1	dep +ve result
	= 0.61 oe	A1 5	•
			$(-1.3)^2 \times 0.2 + (-0.3)^2 \times 0.3 + 0.7^2 \times 0.5$: M2
			one term correct: M1
			Use of Z: MR, lose last A1 (2.55, 0.4475)
ii	$0.2 \times 0.25 + 0.3 \times 0.1$ or $0.05 + 0.03$ alone	M2	M1 for one product eg correct×2: M1
			or clearly ident (1,2), (2,1): M1
	= 0.08 oe	A1 3	
iii	$0.3 \times 0.1 + 0.3 \times 0.25 + 0.3 \times 0.65$		
	$+ 0.25 \times 0.2 + 0.25 \times 0.5$ alone		M1: any 3, 4 of these prods alone
	or $0.03 + 0.075 + 0.195 + 0.05 + 0.125$	M2	or these 5 prods plus 1 extra or repeat
			or (ii) + prod
			or $0.3 + \text{prod or } 0.25 + \text{prod}$
			or clearly identify
	10		(1,2) (3,2) (2,2) (2,1) (2,3)
	$= 0.475 \text{ or }^{19}/_{40} \text{ oe}$	A1 3	
			M2 for $0.3 + (0.2 + 0.5) \times 0.25$
			or $0.25 + (0.1 + 0.65) \times 0.3$
			or $0.3 + 0.25 - 0.3 \times 0.25$
			or $1 - (0.2 + 0.5)(0.1 + 0.65)$
			N1 6 (0.2 + 0.5)(0.1 + 0.65)
TD 4.1		11	M1 for (0.2+ 0.5)(0.1+0.65)
Total	D 14 1 1 1	11	11 (6 ' 22 ' 1 24 (6 ' 1 22 ' 1
7ia	Results or matches are indep	B1	allow "wins" indep; not "trials" indep not "success"
:1.	Prob of winning is constant	B1 2 B1 1	not success
ib	No of wins (or losses)	B1 1	on (1 m) for a leadlest and three leat
ii		M1	or $(1-p)$ for q & allow omit bracket or $352716 p^{10} q^{11} = 293930 p^9 q^{12}$
	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	M1	or $352/16p^{44} = 293930p^{4}q^{15}$ M1 for $^{12}/_{10}$ or $^{6}/_{5}$ or 1.2 or $^{5}/_{6}$ or 0.833
	$\frac{12}{10}p = q$ or $\frac{12}{10}p(1-p)^{-1} = 1$ or similar	M1M1	
	10 10		M1 for <i>p</i> & <i>q</i> cancelled correctly
	1.2p = 1 - p oe eg $p = 0.833(1-p)$	M1	or equiv equn in p or q (cancelled)
	or $352716p = 293930(1-p)$	1711	nos not nec'y cancelled; not alg denom
	01 332/10p - 233330(1-p)		nos not nee y cancened, not arg denom
	$p = \frac{5}{11}$ or 0.455 (3 sfs) oe	A1 5	
Total	F 11 52 51 52 (2 525) 55	8	
101111		U	

	т		T
8i	m = 26.5	B1	
	LQ = 22 or 21.5 or 21.75		
	UQ = 39 $00 = 21.75$ $00 = 21.75$ $00 = 21.75$ $00 = 21.75$	M1	M1 for either LQ or UQ
	IQR = 17	A1 3	A1 must be consistent LQ, UQ & IQR
ii	Ave or overall or med or "it" similar	B1f	or F med (or ave) higher or F mean less
			or M & F both have most in 20s
	Male spread greater or M more varied oe	B1f 2	or male range greater
	1 5		or more younger F or more older M
iii	Med less (or not) affected by extreme(s) or	B1 1	oe; not "anomalies"
111		וים	
 	Mean (more) affected by extreme(s)	 	ignore eg "less accurate"
iv			must consistently decode last or first
ĺ	Decode last		
ĺ	245/49	M1	
ĺ	= 5	A1	
	mean = 205	B1f	200 + "5"
	$\sqrt{(9849/49 - (^{245}/_{49})^2)}$	M1	dep √+ve
	$\sqrt{(9849/49 - (749))}$ = 13.3 (3sfs) or 4 $\sqrt{11}$	A1	Sep 1110
			don M1 on one 176
	$sd = 13.3 \text{ or } 4\sqrt{11}$	B1f 6	dep M1 or ans 176; award if not +200
	Decode first		
	$245 + 200 \times 49 \text{ or } 10045$		
	$\frac{10045}{49}$ M1		allow 445/49 or 9.08 seen
	= 205 M1		, 49 or 7.00 boom
	$\Sigma x^2 = 9849 + 400 \times 10045 - 49 \times 40000$		
	or 2067849 B1		
	$\int_{\Sigma}^{\Sigma} x^{2} = x = 2 = 0$,
ĺ	$\sqrt{\frac{"\Sigma x^2"}{49}} - "\overline{x}^2" $ M1		dep √+ve
			Σx^2 must be: attempt at Σx^2
			>9849
			not involve 9849 ²
			not $(\Sigma x)^2$ eg 10045^2 , 445^2
			\bar{x} must be decoded attempt, eg 9.08
L	$= 13.3 \text{ or } 4\sqrt{11}$ A1	L	<u></u>
Total		12	
9i	Because growth may depend on pH oe	B1 1	In context. Not <i>x</i> is controlled or indep
	or expt is investigating if y depends on x		
**		<u> </u>	
ii	$S_{xy} = 17082.5 - 66.5 \times 1935/8 (= 997.8125)$		
	$S_{xx} = 558.75 - 66.5^2/8$ (= 5.96875)		
	$b = S_{xy}/S_{xx}$	M1	Correct sub into any correct <i>b</i> formula
	= 167 (3 sfs)	A1	
	y - 1935/8 = "167"(x - 66.5/8)	M1	or $a = 1935/8 - 167$ x 66.5/8
	y = 1953/8 = 107 (x = 00.5/8) y = -1150 + 167x	A1 4	$\frac{\text{of } a - 1933/8 - 107 \times 00.3/8}{\text{cao NB 3 sfs}}$
:::		·	
iii	$y = -1150 + 167 \times 7$	M1	ft their eqn for M1 only
	= 19 to 23	A1 2	
iv	No (or little) relationship or correlation	B1 1	or weak or small corr'n.
			Not "agreement"
va	Reliable as <i>r</i> high oe	B1 1	Allow without "interpolation" oe,
			but must include r high
L	Unraliable as sytume leti	D1 1	
<u>b</u>	Unreliable as extrapolation oe	B1 1	or unreliable as gives a neg value
vi	Unreliable (or No) because r near 0	B1 1	or No because Q values vary widely
	or because little (or no or small) corr'n		for pH = 8.5
	(or rel'n)		
Total		11	

Total 72 marks