OCR
RECOGNISING ACHIEVEMENT

ADVANCED SUBSIDIARY GCE
MATHEMATICS 4732
Probability & Statistics 1

Candidates answer on the Answer Booklet

OCR Supplied Materials:
- 8 page Answer Booklet
- List of Formulae (MF1)

Other Materials Required:
None

Wednesday 21 January 2009
Afternoon

Duration: 1 hour 30 minutes

* 4 7 3 2 *

INSTRUCTIONS TO CANDIDATES

- Write your name clearly in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully and make sure that you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- 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.
- **You are reminded of the need for clear presentation in your answers.**
- The total number of marks for this paper is **72**.
- This document consists of **8** pages. Any blank pages are indicated.

1 Each time a certain triangular spinner is spun, it lands on one of the numbers 0, 1 and 2 with probabilities as shown in the table.

Number	Probability
0	0.7
1	0.2
2	0.1

The spinner is spun twice. The total of the two numbers on which it lands is denoted by X.

(i) Show that $P(X = 2) = 0.18$. **[3]**

The probability distribution of X is given in the table.

x	0	1	2	3	4
$P(X = x)$	0.49	0.28	0.18	0.04	0.01

(ii) Calculate $E(X)$ and $Var(X)$. **[5]**

2 The table shows the age, x years, and the mean diameter, y cm, of the trunk of each of seven randomly selected trees of a certain species.

Age (x years)	11	12	20	28	35	45	51
Mean trunk diameter (y cm)	12.2	16.0	26.4	39.2	39.6	51.3	60.6

$[n = 7,\ \Sigma x = 202,\ \Sigma y = 245.3,\ \Sigma x^2 = 7300,\ \Sigma y^2 = 10\,510.65,\ \Sigma xy = 8736.9.]$

(i) (a) Use an appropriate formula to show that the gradient of the regression line of y on x is 1.13, correct to 2 decimal places. **[2]**

(b) Find the equation of the regression line of y on x. **[2]**

(ii) Use your equation to estimate the mean trunk diameter of a tree of this species with age

(a) 30 years, **[1]**

(b) 100 years. **[1]**

It is given that the value of the product moment correlation coefficient for the data in the table is 0.988, correct to 3 decimal places.

(iii) Comment on the reliability of each of your two estimates. **[2]**

3 Erika is a birdwatcher. The probability that she will see a woodpecker on any given day is $\frac{1}{8}$. It is assumed that this probability is unaffected by whether she has seen a woodpecker on any other day.

(i) Calculate the probability that Erika first sees a woodpecker

(a) on the third day, **[3]**

(b) after the third day. **[3]**

(ii) Find the expectation of the number of days up to and including the first day on which she sees a woodpecker. **[1]**

(iii) Calculate the probability that she sees a woodpecker on exactly 2 days in the first 15 days. **[3]**

4 Three tutors each marked the coursework of five students. The marks are given in the table.

Student	*A*	*B*	*C*	*D*	*E*
Tutor 1	73	67	60	48	39
Tutor 2	62	50	61	76	65
Tutor 3	42	50	63	54	71

(i) Calculate Spearman's rank correlation coefficient, r_s, between the marks for tutors 1 and 2. **[5]**

(ii) The values of r_s for the other pairs of tutors, are as follows.

Tutors 1 and 3: $r_s = -0.9$
Tutors 2 and 3: $r_s = \ \ 0.3$

State which two tutors differ most widely in their judgements. Give your reason. **[2]**

5 The stem-and-leaf diagram shows the masses, in grams, of 23 plums, measured correct to the nearest gram.

5	5 6 7 8 8 9
6	1 2 3 5 6 8 9
7	0 0 2 4 5 6 7 8
8	0
9	7

Key : 6 | 2 means 62

(i) Find the median and interquartile range of these masses. **[3]**

(ii) State one advantage of using the interquartile range rather than the standard deviation as a measure of the variation in these masses. **[1]**

(iii) State one advantage and one disadvantage of using a stem-and-leaf diagram rather than a box-and-whisker plot to represent data. **[2]**

(iv) James wished to calculate the mean and standard deviation of the given data. He first subtracted 5 from each of the digits to the left of the line in the stem-and-leaf diagram, giving the following.

0	5 6 7 8 8 9
1	1 2 3 5 6 8 9
2	0 0 2 4 5 6 7 8
3	0
4	7

Key : 1 | 2 means 12

The mean and standard deviation of the data in this diagram are 18.1 and 9.7 respectively, correct to 1 decimal place. Write down the mean and standard deviation of the data in the original diagram. **[2]**

6 A test consists of 4 algebra questions, A, B, C and D, and 4 geometry questions, G, H, I and J.

The examiner plans to arrange all 8 questions in a random order, regardless of topic.

(i) (a) How many different arrangements are possible? **[2]**

(b) Find the probability that no two Algebra questions are next to each other and no two Geometry questions are next to each other. **[3]**

Later, the examiner decides that the questions should be arranged in two sections, Algebra followed by Geometry, with the questions in each section arranged in a random order.

(ii) (a) How many different arrangements are possible? **[2]**

(b) Find the probability that questions A and H are next to each other. **[1]**

(c) Find the probability that questions B and J are separated by more than four other questions. **[4]**

7 At a factory that makes crockery the quality control department has found that 10% of plates have minor faults. These are classed as ‘seconds’. Plates are stored in batches of 12. The number of seconds in a batch is denoted by X.

(i) State an appropriate distribution with which to model X. Give the value(s) of any parameter(s) and state any assumptions required for the model to be valid. **[4]**

Assume now that your model is valid.

(ii) Find

(a) $P(X = 3)$, **[2]**

(b) $P(X \geqslant 1)$. **[2]**

(iii) A random sample of 4 batches is selected. Find the probability that the number of these batches that contain at least 1 second is fewer than 3. **[4]**

8 A game uses an unbiased die with faces numbered 1 to 6. The die is thrown once. If it shows 4 or 5 or 6 then this number is the final score. If it shows 1 or 2 or 3 then the die is thrown again and the final score is the sum of the numbers shown on the two throws.

(i) Find the probability that the final score is 4. **[3]**

(ii) Given that the die is thrown only once, find the probability that the final score is 4. **[1]**

(iii) Given that the die is thrown twice, find the probability that the final score is 4. **[3]**

4732 Probability & Statistics 1

Note: "(3 sfs)" means "answer which rounds to ... to 3 sfs". If correct ans seen to ≥ 3sfs, ISW for later rounding. Penalise over-rounding only once in paper.

1 (i)	$0.2^2 + 0.7 \times 0.1 \times 2$ = 0.18 **AG**	M2 A1 3	0.2^2 or 0.7×0.1: M1 no errors seen NB $2 \times 0.9 \times 0.1 = 0.18$ M0A0
(ii)	$0.28 + 2\times0.18 + 3\times0.04 + 4\times0.01$ = 0.8 oe $0.28 + 2^2\times0.18 + 3^2\times0.04 + 4^2\times0.01$ - "0.8"^{2} = 0.88 oe	M1 A1 M1 M1 A1 5	≥ 2 terms correct (excl 0×0.49) ÷ 5 (or 4 or 10 etc): M0 ≥ 2 terms correct (excl $0^2\times0.49$) dep +ve result cao $\Sigma(x-\mu)^2$: 2 terms: M1; 5 terms M2 $0.8^2\times0.49+0.2^2\times0.28+1.2^2\times0.18+2.2^2\times0.04+3.2^2\times0.01$ SC Use original table, 0.4:B1 0.44: B1
Total		**8**	
2(i)(a)	$\dfrac{8736.9-\frac{202\times245.3}{7}}{7300-\frac{202^2}{7}}$ or $\dfrac{1658.24}{1470.86}$ = 1.127… (= 1.13 **AG**)	M1 A1 2	correct sub in any correct formula for b eg $\dfrac{236.8921}{210.1249}$ must see 1.127… ; 1.127.. alone: M1A1
(b)	$y - {}^{245.3}/_7 = 1.13(x - {}^{202}/_7)$ $y = 1.1x + 2.5$ (or 2.4) or $y = 1.13x + 2.43$	M1 A1 2	or $a = {}^{245.3}/_7 - 1.13 \times {}^{202}/_7$ 2 sfs suff. (exact: $y = 1.127399..x + 2.50934…$)
(ii)(a)	(1.1(..) × 30 + 2.5(..)) = 35.5 to 36.5	B1f 1	
(b)	(1.1(..) × 100 + 2.5(..)) = 112.4 to 115.6	B1f 1	
(iii)	(a) Reliable (b) Unreliable because extrapolated	B1 B1 2	Both reliable: B1 | (a) more reliable than (b) B1 because (a) within data or (b) outside data B1 Ignore extras
Total		**8**	
3(i)(a)	Geo stated $({}^7/_8)^2({}^1/_8)$ ${}^{49}/_{512}$ or 0.0957 (3 sfs)	M1 M1 A1 3	or impl. by $({}^7/_8)^n({}^1/_8)$ or $({}^1/_8)^n({}^7/_8)$ alone
(b)	$({}^7/_8)^3$ alone ${}^{343}/_{512}$ or 0.670 (3 sfs) allow 0.67	M2 A1 3	or $1-({}^1/_8+{}^7/_8\times{}^1/_8+({}^7/_8)^2\times{}^1/_8)$: M2 one term incorrect, omit or extra: M1 $1-({}^7/_8)^3$ or $({}^7/_8)^2$ alone: M1
(ii)	8	B1 1	
(iii)	Binomial stated or implied ${}^{15}C_2({}^7/_8)^{13}({}^1/_8)^2$ = 0.289 (3 sfs)	M1 M1 A1 3	eg by $({}^7/_8)^a({}^1/_8)^b$ ($a+b = 15$, $a,b \neq 1$), not just ${}^nC_r$
Total		**10**	
4 (i)	1 2 3 4 5 or 5 4 3 2 1 3 5 4 1 2 3 1 2 5 3 Σd^2 (= 32) $1 - {}^{6\times"32"}/_{5(25-1)}$ = - 0.6	M1 A1 M1dep M1dep A1 5	attempt ranks correct ranks S_{xx} or S_{yy}=55–15²/5(=10) or S_{yy}=39–15²/5(= -6) ${}^{-6}/\sqrt{(10\times10)}$

(ii)	1 & 3 Largest neg r_s or large neg r_s or strong neg corr'n or close(st) to -1 or lowest r_s	B1ind B1dep 2	ft if -1 < (i) < -0.9, ans 1 & 2 NOT: furthest from 0 or closest to ±1 little corr'n most disagreement
Total		**7**	

5 (i)	68 75 – 59 = 16	B1 M1 A1 3	 attempt 6^{th} & 18^{th} or 58-60, 74-76 & subtr must be from 75 – 59
(ii)	Unaffected by outliers or extremes (allow less affected by outliers) sd can be skewed by one value	B1 1	NOT: … by anomalies or freaks easier to calculate
(iii)	Shows each data item, retains orig data can see how many data items can find (or easier to read) mode or modal class can find (or easier to read) frequs can find mean Harder to read med (or Qs or IQR) Doesn't show med (or Qs or IQR) B&W shows med (or Qs or IQR) B&W easier to compare meds	 B1 B1 2	NOT: shows freqs shows results more clearly B&W does not show freqs NOT: B&W easier to compare B&W shows spread or variance or skew B&W shows highest & lowest Assume in order: Adv, Disadv, unless told Allow disadv of B&W for adv of S&L & vice versa Ignore extras
(iv)	m = 68.1 NOT by restart sd = 9.7 (or same) NOT by restart	B1 B1 2	 Restart mean or mean & sd: 68.1 or 68.087 & 9.7 or 9.73 B1 only
Total		**8**	

6 (i) (a)	8! = 40320	M1 A1 2	Allow $^4P_4$ & $^3P_3$ instead of 3! & 4! thro'out Q6
(b)	$^4/_8 \times {^4/_7} \times {^3/_6} \times {^3/_5} \times {^2/_4} \times {^2/_3} \times {^1/_2}$ $\times 2$ $= {^1/_{35}}$ or 0.0286 (3 sfs)	M1 M1dep A1 3	$4! \times 4! \div 8!$ $\times 2$ | $4! \times 4! + 4! \times 4!$ $\div 8!$ allow 1 – above for M1 only oe, eg $^{1152}/_{40320}$
(ii)(a)	$4! \times 4!$ = 576	M1 A1 2	allow $4! \times 4! \times 2$: M1
(b)	$^1/_{16}$ or 0.0625	B1 1	
(c)	Separated by 5 or 6 qus stated or illus $^1/_4 \times {^1/_4} \times 3$ or $^1/_{16} \times 3$ ($^1/_4 \times {^1/_4}$ or $^1/_{16}$ alone or ×(2 or 6): M1) $^3/_{16}$ or 0.1875 or 0.188	M1 M2 A1 4	allow 5 only or 6 only or (4, 5 or 6) can be impl by next M2 or M1 $3! \times 3! \times 3$ ($3! \times 3!$ alone or ×(2 or 6); or $(3! + 3!) \times 3$: M1) ($\div 576$) correct ans, but clearly B, J sep by 4: M0M2A0 1- P(sep by 0, 1, 2, 3, (4)) M1 $1-({^1/_4}+{^1/_4}+{^1/_4}\times{^3/_4}+{^1/_4}\times{^1/_2})$ or $1-({^1/_4}\times{^1/_4}+{^1/_2}\times{^1/_4}+{^3/_4}\times{^1/_4}+1\times{^1/_4}+{^3/_4}\times{^1/_4})$ M2 (one omit: M1)
Total		**12**	

7 (i)	Binomial $n = 12, p = 0.1$ Plates (or seconds) independent oe Prob of fault same for each plate oe	B1 B1 B1 B1 4	 B(12, 0.1) : B2 NOT: batches indep Comments must be in context Ignore incorrect or irrelevant
(ii)(a)	0.9744 – 0.8891 or $^{12}C_3 \times 0.9^9 \times 0.1^3$ = 0.0852 or 0.0853 (3 sfs)	M1 A1 2	
(b)	1 – 0.2824 or $1 - 0.9^{12}$ =0.718 (3 sfs)	M1 A1 2	allow 1 – 0.6590 or $1 - 0.9^{11}$
(iii)	"0.718" and 1 – "0.718" used $(1\text{-}0.718)^4 + 4(1\text{-}0.718)^3 \times 0.718$ $+ {^4C_2}(1\text{-}0.718)^2 \times 0.718^2$ = 0.317 (3 sfs)	B1 M2 A1 4	ft (b) for B1M1M1 M1 for any one term correct (eg opp tail or no coeffs) 1 – P(3 or 4) follow similar scheme M2 or M1 1 – correct wking (= 0.623) B1M2 cao
Total		**12**	

8 (i)	$^1/_6 + 3 \times (^1/_6)^2$ $= ^1/_4$	M2 A1 3	or $3 \times (^1/_6)^2$ or $^1/_6 + (^1/_6)^2$ or $^1/_6 + 2(^1/_6)^2$ or $^1/_6 + 4(^1/_6)^2$ M1
(ii)	$^1/_3$	B1 1	
(iii)	3 routes clearly implied out of 18 possible (equiprobable) routes $^1/_6$	M1 M1 A1 3	or $^1/_3 \times ^1/_6 \times 3$ M2 or $^1/_3 \times ^1/_6$ or $^1/_6 \times ^1/_6 \times 3$ or $^1/_3 \times ^1/_3 \times 3$ or $^1/_4 - ^1/_6$ M1 but $^1/_6 \times ^1/_6 \times 2$ M0 $\frac{(\frac{1}{6})^2 \times 3}{\frac{1}{2}}$ or $\frac{\frac{1}{4}-\frac{1}{6}}{\frac{1}{2}}$ or $\frac{\frac{1}{2}\times\frac{1}{6}}{\frac{1}{2}}$ oe M2 or $\frac{\text{P(4\&twice)}}{\text{P(twice)}}$ stated or $\frac{\text{prob}}{\frac{1}{2}}$ M1 Whatever 1st, only one possibility on 2nd M2 $^1/_6$, no wking M1M1A1 $^1/_{12}$, no wking M0
Total		**7**	

Total 72 marks