OCR
RECOGNISING ACHIEVEMENT

ADVANCED SUBSIDIARY GCE **4732/01**

MATHEMATICS

Probability & Statistics 1

FRIDAY 23 MAY 2008 Morning

Time: 1 hour 30 minutes

Additional materials (enclosed): None

Additional materials (required):
Answer Booklet (8 pages)
List of Formulae (MF1)

INSTRUCTIONS TO CANDIDATES

- Write your name in capital letters, your 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

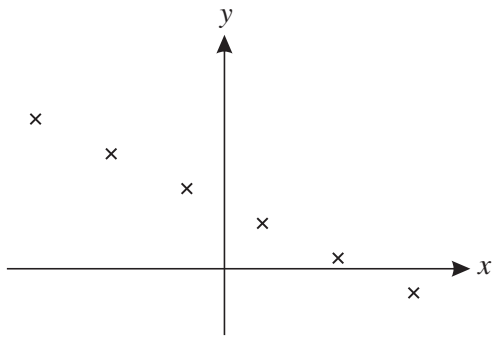
- The number of marks is given in brackets [ ] at the end of each question or part question.
- The total number of marks for this paper is **72**.
- **You are reminded of the need for clear presentation in your answers.**

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

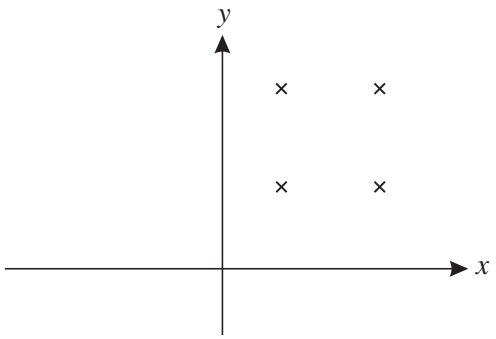
 [Turn over

1 **(i)** State the value of the product moment correlation coefficient for each of the following scatter diagrams. [2]

(a)



(b)



(ii) Calculate the value of Spearman's rank correlation coefficient for the following data. [5]

x	3.8	4.1	4.5	5.3
y	1.4	0.8	0.7	1.2

2 A class consists of 7 students from Ashville and 8 from Bewton. A committee of 5 students is chosen at random from the class.

(i) Find the probability that 2 students from Ashville and 3 from Bewton are chosen. [3]

(ii) In fact 2 students from Ashville and 3 from Bewton are chosen. In order to watch a video, all 5 committee members sit in a row. In how many different orders can they sit if no two students from Bewton sit next to each other? [2]

3 **(i)** A random variable X has the distribution $\mathrm{B}(8, 0.55)$. Find

(a) $\mathrm{P}(X < 7)$, [1]

(b) $\mathrm{P}(X = 5)$, [2]

(c) $\mathrm{P}(3 \leqslant X < 6)$. [3]

(ii) A random variable Y has the distribution $\mathrm{B}\left(10, \frac{5}{12}\right)$. Find

(a) $\mathrm{P}(Y = 2)$, [2]

(b) $\mathrm{Var}(Y)$. [1]

4 At a fairground stall, on each turn a player receives prize money with the following probabilities.

Prize money	£0.00	£0.50	£5.00
Probability	$\frac{17}{20}$	$\frac{1}{10}$	$\frac{1}{20}$

(i) Find the probability that a player who has two turns will receive a total of £5.50 in prize money. [3]

(ii) The stall-holder wishes to make a profit of 20p per turn on average. Calculate the amount the stall-holder should charge for each turn. [4]

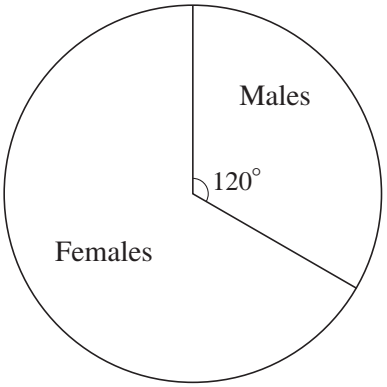
5 **(i)** A bag contains 12 red discs and 10 black discs. Two discs are removed at random, without replacement. Find the probability that both discs are red. [2]

(ii) Another bag contains 7 green discs and 8 blue discs. Three discs are removed at random, without replacement. Find the probability that exactly two of these discs are green. [3]

(iii) A third bag contains 45 discs, each of which is either yellow or brown. Two discs are removed at random, without replacement. The probability that both discs are yellow is $\frac{1}{15}$. Find the number of yellow discs which were in the bag at first. [4]

[Turn over

6 **(i)** The numbers of males and females in Year 12 at a school are illustrated in the pie chart. The number of males in Year 12 is 128.

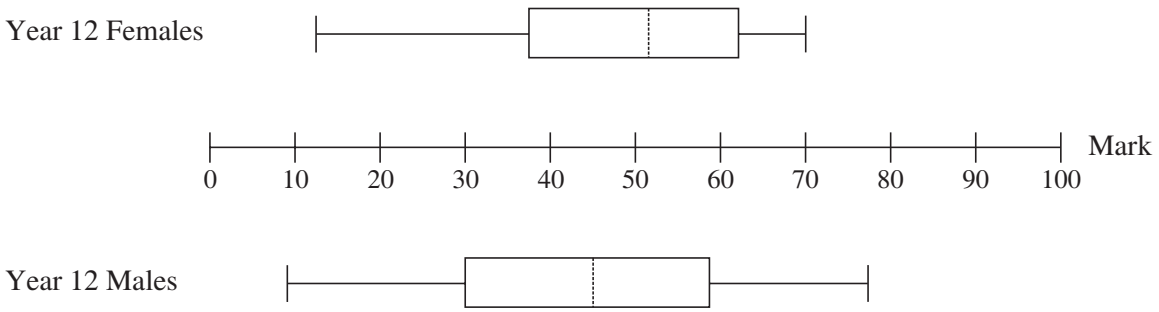


Year 12

(a) Find the number of females in Year 12. [1]

(b) On a corresponding pie chart for Year 13, the angle of the sector representing males is 150°. Explain why this does not necessarily mean that the number of males in Year 13 is more than 128. [1]

(ii) All the Year 12 students took a General Studies examination. The results are illustrated in the box-and-whisker plots.



(a) One student said "The Year 12 pie chart shows that there are more females than males, but the box-and-whisker plots show that there are more males than females."

Comment on this statement. [1]

(b) Give two comparisons between the overall performance of the females and the males in the General Studies examination. [2]

(c) Give one advantage and one disadvantage of using box-and-whisker plots rather than histograms to display the results. [2]

(iii) The mean mark for 102 of the male students was 51. The mean mark for the remaining 26 male students was 59. Calculate the mean mark for all 128 male students. [3]

7 Once each year, Paula enters a lottery for a place in an annual marathon. Each time she enters the lottery, the probability of her obtaining a place is 0.3. Find the probability that

(i) the first time she obtains a place is on her 4th attempt, [3]

(ii) she does not obtain a place on any of her first 6 attempts, [2]

(iii) she needs fewer than 10 attempts to obtain a place, [3]

(iv) she obtains a place exactly twice in her first 5 attempts. [3]

8 A city council attempted to reduce traffic congestion by introducing a congestion charge. The charge was set at £4.00 for the first year and was then increased by £2.00 each year. For each of the first eight years, the council recorded the average number of vehicles entering the city centre per day. The results are shown in the table.

Charge, £x	4	6	8	10	12	14	16	18
Average number of vehicles per day, y million	2.4	2.5	2.2	2.3	2.0	1.8	1.7	1.5

[$n = 8$, $\Sigma x = 88$, $\Sigma y = 16.4$, $\Sigma x^2 = 1136$, $\Sigma y^2 = 34.52$, $\Sigma xy = 168.6$.]

(i) Calculate the product moment correlation coefficient for these data. [3]

(ii) Explain why x is the independent variable. [1]

(iii) Calculate the equation of the regression line of y on x. [4]

(iv) **(a)** Use your equation to estimate the average number of vehicles which will enter the city centre per day when the congestion charge is raised to £20.00. [2]

(b) Comment on the reliability of your estimate. [2]

(v) The council wishes to estimate the congestion charge required to reduce the average number of vehicles entering the city per day to 1.0 million. Assuming that a reliable estimate can be made by extrapolation, state whether they should use the regression line of y on x or the regression line of x on y. Give a reason for your answer. [2]

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)	(a) -1 (b) 0	B1 B1 2	allow ≈ -1 or close to -1 not "strong corr'n", not -0.99 allow ≈ 0 or close to 0 not "no corr'n"
(ii)	4 3 2 1 or 1 2 3 4 1 3 4 2 4 2 1 3 Σd^2 (= 14) $1 - \frac{6\Sigma d^2}{4(4^2-1)}$ = -0.4 oe	M1 A1 M1 M1 A1 5	Ranks attempted, even if opp Dep M1 or $S_{xy} = 23-{}^{100}/_{4}$ or $S_{xx} = S_{yy} = 30-{}^{100}/_{4}$ Dep 2nd M1 $S_{xy}/\sqrt{(S_{xx}S_{yy})}$
Total		**7**	
2(i)	$\frac{{}^{7}C_2 \times {}^{8}C_3}{{}^{15}C_5}$ = ${}^{56}/_{143}$ or ${}^{1176}/_{3003}$ or 0.392 (3sfs)	M1 M1 A1 3	${}^{7}C_2 \times {}^{8}C_3$ or 1176 : M1 (Any C or P)/${}^{15}C_5$: M1 (dep < 1) or $\frac{7}{15}\times\frac{6}{14}\times\frac{8}{13}\times\frac{7}{12}\times\frac{6}{11}$ or 0.0392: M1 $\times {}^{5}C_2$ or × 10 : M1 (dep ≥4 probs mult) if 2↔3, treat as MR max M1M1
(ii)	3! x 2! or ${}^{3}P_3 \times {}^{2}P_2$ not in denom = 12	M1 A1 2	BABAB seen: M1 120-12: M1A0 NB ${}^{4!}/_{2!}$ = 12: M0A0
Total		**5**	
3(i)(a)	0.9368 or 0.937	B1 1	
(b)	0.7799 – 0.5230 or ${}^{8}C_5 \times 0.45^3 \times 0.55^5$ = 0.2569 or 0.2568 or 0.257	M1 A1 2	Allow 0.9368 – 0.7799
(c)	0.7799 seen – 0.0885 (not 1 – 0.0885) = 0.691 (3 sfs)	M1 M1 A1 3	${}^{8}C_5 \times 0.45^3 \times 0.55^5 + {}^{8}C_4 \times 0.45^4 \times 0.55^4 + {}^{8}C_3 \times 0.45^5 \times 0.55^3$: M2 1 term omitted or wrong or extra: M1
(ii)(a)	${}^{10}C_2 \times ({}^{7}/_{12})^8 \times ({}^{5}/_{12})^2$ seen = 0.105 (3 sfs)	M1 A1 2	or 0.105 seen, but not ISW for A1
(b)	$2{}^{31}/_{72}$ or ${}^{175}/_{72}$ or 2.43 (3 sfs)	B1 1	NB ${}^{12}/_{5}$ = 2.4: B0
Total		**9**	
4(i)	${}^{1}/_{20} \times {}^{1}/_{10}$ or ${}^{1}/_{200}$ or 0.005 x 2 = ${}^{1}/_{100}$ or 0.01	M1 M1dep A1 3	
(ii)	$E(X) = 0+50\times{}^{1}/_{10}+500\times{}^{1}/_{20}$ or $0+0.5\times{}^{1}/_{10}+5\times{}^{1}/_{20}$ = 30p = £0.30 or ${}^{3}/_{10}$ Charge "30p" + 20p or 0.3 + 0.2 = 50p or 0.50 or 0.5	M1 A1 M1 A1 4	or eg 20 goes: 2 × £0.50 + £5.00 = £6.00 ("£6.00" + 20 × £0.20) ÷ 20 condone muddled units eg 0.3 + 20 x = 20, 70, 520 : M1A1 $20\times{}^{17}/_{20} + 70\times{}^{1}/_{10} + 520\times{}^{1}/_{20}$: M1 = 50 A1 x, $(x-50)$, $(x-500)$: M1A1 $x\times{}^{17}/_{20}+(x-50)\times{}^{1}/_{10}+(x-500)\times{}^{1}/_{20} = 20$: M1 x = 50 : A1 Ignore "£" or "p"
Total		**7**	

5(i)	$^{12}/_{22}$ x $^{11}/_{21}$ $= {}^{2}/_{7}$ oe or 0.286 (3 sfs)	M1 A1 2	or $^{12}C_2$ / $^{22}C_2$
(ii)	$^{7}/_{15}$ x $^{6}/_{14}$ x $^{8}/_{13}$ or $^{8}/_{65}$ oe × 3 oe $= {}^{24}/_{65}$ or 0.369 (3 sfs)	M1 M1 A1 3	Numerators any order $^{7}C_2$ x $^{8}C_1$:M1 3 x prod any 3 probs (any C or P)/$^{15}C_3$:M1 (dep <1) $1-(^{8}/_{15}x^{7}/_{14}x^{6}/_{13}+3\times{}^{8}/_{15}x^{7}/_{14}x^{7}/_{13}+{}^{7}/_{15}x^{6}/_{14}x^{5}/_{13})$: M2 one prod omitted or wrong: M1
(iii)	$\frac{x}{45}\times\frac{x-1}{44}=\frac{1}{15}$ oe $x^2 - x - 132 = 0$ or $x(x-1) = 132$ $(x - 12)(x + 11) = 0$ or $x = \frac{1 \pm \sqrt{(1^2 - 4\times(-132))}}{2}$ No. of Ys = 12	M1 A1 M1 A1 4	not $\frac{x}{45}\times\frac{x}{44}=\frac{1}{15}$ or $\frac{x}{45}\times\frac{x}{45}=\frac{1}{15}$ or $\frac{x}{45}\times\frac{x-1}{45}=\frac{1}{15}$ oe ft 3-term QE for M1 condone signs interchanged allow one sign error Not $x = 12$ or -11 ans 12 from less wking, eg $12\times11=132$ or T & I: full mks Some incorrect methods: $\frac{x}{45}\times\frac{x-1}{44}=\frac{1}{15}$ oe M1 $x^2 + x = 132$ A0 $x = 11$ M1A0 $12 \times 11 = 132$ M1A1M1 $x = 12$ and (or "or") 11 A0 NB 12 from eg 12.3 rounded, check method
Total		**9**	

6(i)(a)	256	B1 1	
			(i)(b) & (ii)(abc): ISW ie if correct seen, ignore extras
(b)	Total unknown or totals poss diff or Y13 may be smaller or similar or size of pie chart may differ	B1 1	pie chart shows only proportions oe or no. of students per degree may differ not "no. of F may be less" not "Y13 may be larger"
(ii)(a)	B&W does not show frequencies oe	B1 1	or B&W shows spread or shows mks or M lger range
(b)			1 mk about overall standard, based on median or on F's IQR being "higher" 1 mk about spread (or range or IQR) or about skewness. must be overall, not indiv mks must be comparison, not just figures Examples:
	F generally higher or median higher F higher on average or F better mks F IQR is above M IQR	B1	not F higher mean
	F more compact M wide(r) range or gter IQR or gter variation or gter variance or more spread or less consistent M evenly spread or F skewed	B1 2	not M have hiest and lowest mks condone F +ve skew
(c)	<u>Advantage</u>: B&W shows med or Qs or IQR or range or hiest & lowest or key values	B1	not B&W shows skewness not B&W shows info at a glance not B&W easier to compare data sets not B&W shows mean not B&W shows spread not B&W easier to calculate or easier to read
	<u>Disadvantage</u>: B&W loses info' B&W shows less info' B&W not show freqs B&W not show mode B&W: outlier can give false impression hist shows more info hist shows freqs or fds hist shows modal class (allow mode) hist shows distribution better can calc mean from hist	B1 2	not B&W does not give indiv (or raw) data not B&W does not show mean not hist shows freq for each mark not hist shows all the results not hist shows total allow adv of hist as disadv of B&W
(iii)	102 x 51 + 26 x 59 ÷ 128 = 52.6 (3 sfs)	M1 M1dep A1 3	or 5202 + 1534 or 6736
Total		**10**	

7(i)	Geo stated 0.7^3 x 0.3 $^{1029}/_{10000}$ oe or 0.103 (3 sfs)	M1 M1 A1 3	or implied by 0.7^rx0.3 or 0.3^rx0.7 Allow 0.7^4 x 0.3
(ii)	0.7^6 alone = 0.118 (3 sfs)	M1 A1 2	1-(0.3+0.3×0.7+...+ 0.3×0.7^5) not 1-0.7^6
(iii)	0.7^9 $1 - 0.7^9$ 0.960 (3 sfs)	M1 M1 A1 3	not 0.3 × 0.7^9 allow $1 - 0.7^{10}$ or 0.972 for M1 allow 0.96, if no incorrect wking seen 0.3 + 0.7 x 0.3 + …. + 0.7^8 x 0.3: M2 1 term omitted or wrong or "correct" extra: M1
(iv)	Bin stated $^5C_2$ x 0.7^3 x 0.3^2 or 0.8369 – 0.5282 = 0.3087 or 0.309 (3 sfs)	M1 M1 A1 3	or implied by table or ^nC_r or $0.7^3 \times 0.3^2$ or 0.0309
Total		**11**	
8(i)	$\dfrac{168.6 - \frac{88\times16.4}{8}}{\sqrt{(1136 - \frac{88^2}{8})(34.52 - \frac{16.4^2}{8})}}$ = -0.960 (3 sfs)	M2 A1 3	$(= \dfrac{-11.8}{\sqrt{168\times0.9}})$ M1: correct subst in any correct *S* formula M2: correct substn in any correct *r* formula allow -0.96, if no incorrect wking seen
(ii)	must refer to, or imply, external constraint on *x* e.g *x* is controlled or values of *x* fixed or chosen allow *x* is fixed	B1 1	not *x* is not random not *x* affects *y* not *x* not affected by *y* not *x* goes up same amount each time not charge affects no. of vehicles not *x* not being measured
(iii)	$\dfrac{168.6 - \frac{88\times16.4}{8}}{1136 - \frac{88^2}{8}}$ = -0.0702 (3 sfs) or $-^{59}/_{840}$ or $-^{11.8}/_{168}$ $y - {}^{16.4}/_8 =$ "-0.0702"$(x - {}^{88}/_8)$ y = -0.07x + 2.8 or better	M1 A1 M1 A1 4	ft their S_{xy} and S_{xx} incl $^{168.6}/_{1136}$ if used in (i) or -0.07 if no incorrect wking or $a = {}^{16.4}/_8 -$ ("-0.0702") x $^{88}/_8$ or $^{2371}/_{840}$ oe eg $y = -{}^{59}/_{840}x + {}^{2371}/_{840}$
(iv)(a)	"-0.07" x 20 + "2.8" = 1.4(2) million (2 sfs)	M1 A1 2	 no ft
(b)	*r* close to -1 or corr'n is high just outside given data, so reliable	B1 B1 2	or good corr'n or pts close to line but not if "close to -1, hence unreliable" if *r* low in (i), ft: "*r* low" or "poor corr'n" etc or outside given data so unreliable not "reliable as follows trend" not "reliable as follows average" no ft from (iv)(a)
(v)	*y* on *x* *x* is indep	B1 B1 2	 or *x* controlled or *y* depends on *x* or *y* not indep dep on not "*x* on *y*" *r* close to -1 so makes little difference: B2
Total		**14**	