

Wednesday 25 January 2012 – Afternoon

AS GCE MATHEMATICS

4732 Probability and Statistics 1

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:

- Printed Answer Book 4732
- 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**.
- This Printed Answer Book consists of **12** pages. The Question Paper consists of **8** 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 The probability distribution of a random variable X is shown in the table.

x	1	2	3	4
$P(X=x)$	0.1	0.3	$2p$	p

(i) Find p . [2]

(ii) Find $E(X)$. [2]

- 2 In an experiment, the percentage sand content, y , of soil in a given region was measured at nine different depths, x cm, taken at intervals of 6 cm from 0 cm to 48 cm. The results are summarised below.

$$n = 9 \quad \Sigma x = 216 \quad \Sigma x^2 = 7344 \quad \Sigma y = 512.4 \quad \Sigma y^2 = 30\,595 \quad \Sigma xy = 10\,674$$

(i) State, with a reason, which variable is the independent variable. [1]

(ii) Calculate the product moment correlation coefficient between x and y . [3]

(iii) (a) Calculate the equation of the appropriate regression line. [3]

(b) This regression line is used to estimate the percentage sand content at depths of 25 cm and 100 cm. Comment on the reliability of each of these estimates. You are not asked to find the estimates. [3]

- 3 A random variable X has the distribution $B(13, 0.12)$.

(i) Find $P(X < 2)$. [3]

Two independent values of X are found.

(ii) Find the probability that exactly one of these values is equal to 2. [3]

- 4 (a) The table gives the heights and masses of 5 people.

Person	A	B	C	D	E
Height (m)	1.72	1.63	1.77	1.68	1.74
Mass (kg)	75	62	64	60	70

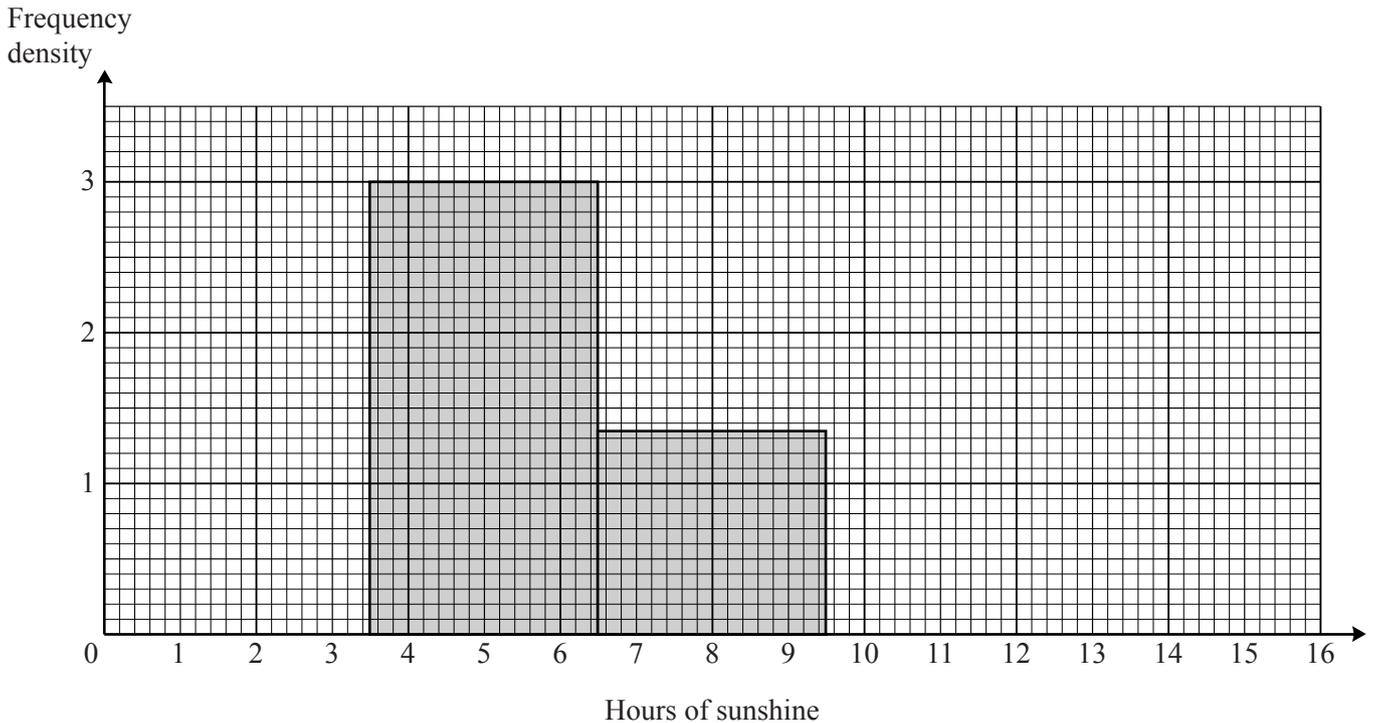
Calculate Spearman's rank correlation coefficient. [5]

(b) In an art competition the value of Spearman's rank correlation coefficient, r_s , calculated from two judges' rankings was 0.75. A late entry for the competition was received and both judges ranked this entry lower than all the others. By considering the formula for r_s , explain whether the new value of r_s will be less than 0.75, equal to 0.75, or greater than 0.75. [3]

- 5 At a certain resort the number of hours of sunshine, measured to the nearest hour, was recorded on each of 21 days. The results are summarised in the table.

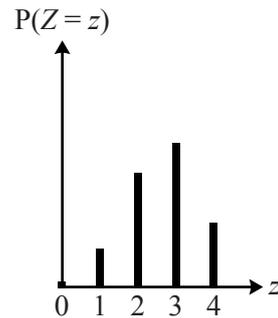
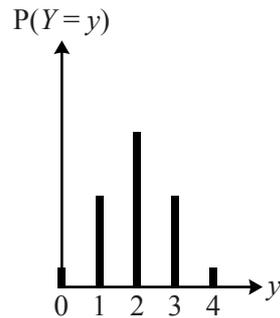
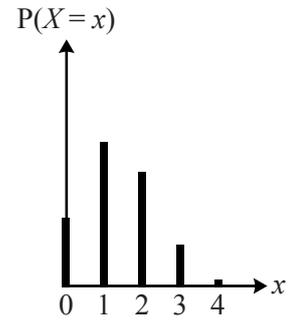
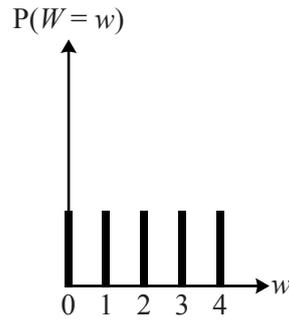
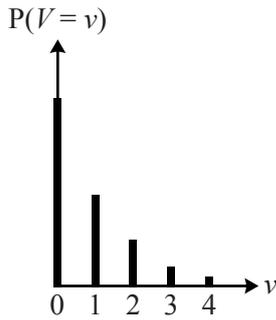
Hours of sunshine	0	1 – 3	4 – 6	7 – 9	10 – 15
Number of days	0	6	9	4	2

The diagram shows part of a histogram to illustrate the data. The scale on the frequency density axis is 2 cm to 1 unit.



- (i) (a) Calculate the frequency density of the 1 – 3 class. [1]
- (b) Fred wishes to draw the block for the 10 – 15 class on the same diagram. Calculate the height, in centimetres, of this block. [2]
- (ii) A cumulative frequency graph is to be drawn. Write down the coordinates of the first two points that should be plotted. You are not asked to draw the graph. [2]
- (iii) (a) Calculate estimates of the mean and standard deviation of the number of hours of sunshine. [5]
- (b) Explain why your answers are only estimates. [1]

- 6 The diagrams illustrate all or part of the probability distributions of the discrete random variables V , W , X , Y and Z .



- (i) One of these variables has the distribution $\text{Geo}(\frac{1}{2})$. State, with a reason, which variable this is. [2]
- (ii) One of these variables has the distribution $\text{B}(4, \frac{1}{2})$. State, with reasons, which variable this is. [3]
- 7 60% of the voters at a certain polling station are women. Voters enter the polling station one at a time. The number of voters who enter, up to and including the first woman, is denoted by X .

- (i) State a suitable distribution that can be used as a model for X , giving the value(s) of any parameter(s). State also any necessary condition(s) for this distribution to be a good model. [4]

Use the distribution stated in part (i) to find

- (ii) $P(X=4)$, [2]
- (iii) $P(X \geq 4)$. [2]

- 8 On average, half the plants of a particular variety produce red flowers and the rest produce blue flowers.
- (i) Ann chooses 8 plants of this variety at random. Find the probability that more than 6 plants produce red flowers. [3]
- (ii) Karim chooses 22 plants of this variety at random.
- (a) Find the probability that the number of these plants that produce blue flowers is equal to the number that produce red flowers. [2]
- (b) Hence find the probability that the number of these plants that produce blue flowers is greater than the number that produce red flowers. [3]
- 9 A bag contains 9 discs numbered 1, 2, 3, 4, 5, 6, 7, 8, 9.
- (i) Andrea chooses 4 discs at random, without replacement, and places them in a row.
- (a) How many different 4-digit numbers can be made? [2]
- (b) How many different **odd** 4-digit numbers can be made? [3]
- (ii) Andrea's 4 discs are put back in the bag. Martin then chooses 4 discs at random, without replacement. Find the probability that
- (a) the 4 digits include at least 3 odd digits, [4]
- (b) the 4 digits add up to 28. [3]

Note: "(3 sf)" means "answer which rounds to ... to 3 sf". If correct ans seen to ≥ 3 sf, ISW for later rounding
 Penalise over-rounding only once in paper. NB If marking by question and over-rounding is seen, must mark whole paper.

Question		Answer	Marks	Guidance	
1	(i)	$0.1 + 0.3 + 2p + p = 1$ oe $p = 0.2$	M1 A1 [2]		
1	(ii)	Σxp $= 2.7$ oe	M1 A1f [2]	≥ 2 terms correct, FT p	eg $\div 4$: M0A0
2	(i)	x because values (or depths) are fixed (or controlled or chosen or predetermined or manipulated or given oe) because they can be changed or it is changed or because it is not measured ie not "read off" oe or because we change the values ourselves	B1 [1]	Allow "because it goes up in intervals" or "because it is taken at set intervals" Ignore all else NB "x is changed" B1, but "x changes" B0	NOT: x , as values are constant x , as y depends on x x as % sand depends on depth Depth, as not affected by % sand content x , as it is not dependent x , because y is measured x , because it changes y , which is the depth and this is controlled
2	(ii)	$S_{xx} = 7344 - \frac{216^2}{9}$ (= 2160) $S_{yy} = 30595 - \frac{512.4^2}{9}$ (= 1422.36) $S_{xy} = 10674 - \frac{216 \times 512.4}{9}$ (= - 1623.6) $r = \frac{-1623.6}{\sqrt{2160 \times 1422.36}}$ $= -0.926$ (3 sfs)	M1 M1 A1 [3]	correct subst in any S formula correct subst in all S s & in r	

Question		Answer	Marks	Guidance
2	(iii) (a)	$b = \frac{-1623.6}{2160}$ or $-0.75\dots$ or $-\frac{451}{600}$ $y - \frac{512.4}{9} = "-0.75\dots"(x - \frac{216}{9})$ $y = -0.75x + 75(.0)$ (2 sf) or $y = -\frac{451}{600}x + \frac{5623}{75}$	M1 M1 A1 [3]	ft S_{xy} & S_{xx} from (ii) or $a = \frac{512.4}{9} - 0.75\dots \times (-\frac{216}{9})$ or $\frac{5623}{75}$ 2 sf is enough Allow $y = -0.75x + (-75)$ If ans to (i) is y , & x on y found here: $b' = \frac{-1623.6}{1422.36}$ (= -1.14) M1 $x - \frac{216}{9} = "-1.14"(y - \frac{512.4}{9})$ M1 $x = -1.14y + 89(.0)$ A1 If ans to (i) is x , but x on y found here: B1 only for $x = -1.14y + 89(.0)$
2	(iii) (b)	r close to -1 (or high or strong), $ r $ close to 1 25 within range of data oe, so reliable 100 outside range of data oe, so unreliable Must give reasons Allow "accurate" instead of "reliable"	B1 B1 B1 [3]	Allow strong or good or high corr'n or rel'nship etc or so more reliable or so less reliable If (ii) $ r < 0.7$: poor corr'n oe B1f 25 unreliable B1f 100 unreliable B1f or strong neg corr'n. Award this mark even if comment linked to 100 instead of linked to 25. BUT: " r close to -1 , so unreliable": B0 Can still score next marks if mention "within" and "outside range" or 100 gives neg %age "Reliable because r near -1 " B1B0B0 "Small sample so unreliable" B0B0B0 Ignore all else
3	(i)	$(1 - 0.12)^{13}$ or $13 \times (1 - 0.12)^{12} \times 0.12$ $(1 - 0.12)^{13} + 13 \times (1 - 0.12)^{12} \times 0.12$ $= 0.526$ (3 sf)	M1 M1 A1[3]	Either seen Fully correct method 1 – correct terms: M1M0A0
3	(ii)	${}^{13}C_2 \times 0.12^2 \times (1 - 0.12)^{11}$ $2 \times "0.275275" \times (1 - "0.275275")$ $= 0.399$ (3 sf)	M1 M1 A1 [3]	or $0.275(\dots)$ Correct method except allow omit " $2 \times$ " Allow if \times or $+$ something NB unlike 2 nd M1 in (i) which is for fully correct method NB $2 \times 0.12 \times 0.88$: M0M0A0

Question		Answer	Marks	Guidance	
5	(ii)	(3.5, 6) (0.5, 0) or (6.5, 15)	B1 B1 [2]	Ignore incorrect (6, 3.5) AND (15, 6.5): B1	
5	(iii) (a)	$\frac{\Sigma xf}{21}$ = 5.43 (3 sf) or $\frac{114}{21}$ or $\frac{38}{7}$ oe $\frac{\Sigma x^2 f}{21}$ or $\frac{817.5}{21}$ or 38.9... – “5.43” ² or = 9.46 or 9.4592.... ($\sqrt{9.4592\dots}$) = 3.08 (3 sfs)	M1 A1 M1 M1 A1 [5]	Allow x within classes, incl end pts then $\div 5$: M0A0 Allow x within class, incl end pt $\div 5$: M0 dep +ve result; done before $\sqrt{\quad}$; not $-(\bar{x}^2 \div \dots)$	≥ 2 non-zero terms correct ft their x ≥ 2 non-zero terms correct ft their x Calc 4 values of $(x - \bar{x})^2$ or $(x - \bar{x})^2 f$ or (11.8, 0.184, 6.61, 50) or (70.5, 1.65, 26.4, 100) or 199 M1 $\frac{\Sigma(x-\bar{x})^2 f}{21}$ fully correct method M1
5	(iii) (b)	Actual values or exact hours unknown oe Don't have raw data. oe or measured to nearest hour oe	B1 [1]	or Data given in classes or grouped oe or Data evenly distributed in classes oe	Mid-points or medians or averages of class boundaries used oe

Question		Answer	Marks	Guidance
6	(i)	V because [probs or values or geometric or etc] decreasing or halving or Highest prob is 1st Allow if word “decreasing” or “halving” or “sloping downwards” or any equivalent seen NOT “Positive skew”	B1 B1 [2]	X because mode = 1 oe or Highest prob is $P(1)$ oe B2 Z because $P(0) = 0$ or variable can't be 0 oe Allow “Geo distr'n cannot be zero” oe B2 “None of them”: Ignore any reason given. B2 For answer V the first B1 is indep, but not for other answers, ie: V with no reason or incorrect reason scores B1B0, but Z or X or any other letter with no reason or incorrect reason scores B0B0. In all cases, once mark(s) have been scored, ignore all other comments.
6	(ii)	Y. Peaks at 2 Y. Like normal, peak at 2 Y. Highest prob is middle one (or is at 2) Y. $P(X = 2)$ is max Y. Increase to 2 then decr Y. 1 4 6 4 1 alone or with $0.5^4 \times$ Y. 0.0625, 0.25, 0.375, 0.25, 0.0625 Y. $P(1) = P(3)$ and $P(2)$ is greater/different or equiv of any of the above ----- If none of the above applies: Any implication that values not all equal eg: Not uniform or values increase (then decrease) or there is a peak Symmetrical or mirror image oe or ${}^4C_0 = {}^4C_4$ or 2nd = 4th or similar or mean = 2, or $E(X) = 2$, or 2 is hi'est prob, or peak at 2, or peak is middle value Y	B1B1B1 B1 B1 B1 [3]	Ignore all else ${}^4C_0, {}^4C_1, {}^4C_2$, etc indep indep indep If values of <u>some</u> probs listed: 2 to 4 values: B1 Y: B1 For 3 rd B1 must link list with Y diag, eg “symmetrical” or “peak in middle” or “peak at 2” or “1 st = last” or “2 nd = 4 th ” “same shape as Y diag”. etc etc

Question		Answer	Marks	Guidance
7	(i)	Geo(0.6) or G(0.6) or Geo with $p = 0.6$ P(woman) const or chance of woman const Each voter has same prob Whether one voter is a woman is indep of whether any other is a woman	B1B1 B1 B1 [4]	Allow Geo(60%) B1B1 or %age of women is constant Allow: “voter(s) independent”, “Men & women are independent” “P(woman) is indep” “Each woman is indep” Bin(..., 0.6) B0B1. Can still score comment marks In context Allow “vote indep” In context (EACH comment must be in context) Ignore all else
7	(ii)	$0.4^3 \times 0.6$ $= \frac{24}{625}$ or 0.0384	M1 A1f [2]	ft their Geo(p) from (i) ft their Geo(p) from (i) Allow $0.3^3 \times 0.6$ (but no other $q^3 \times 0.6$) eg if $p = 0.4$, ans 0.0864 M1A1f
7	(iii)	0.4^3 alone, or $(0.4^4 + 0.4^3 \times 0.6)$ or $(0.4^4 +$ (ii)) $= \frac{8}{125}$ or 0.064	M1 A1f [2]	$1 - (0.6 + 0.4 \times 0.6 + 0.4^2 \times 0.6)$ (allow extra term $0.4^3 \times 0.6)$ ft their Geo(p) from (i) ft their Geo(p) from (i) Allow M1 for 0.4^4 alone (= 0.0256) M0 for $0.4^r \times 0.6$ and for $1 - 0.4^3$ eg if $p = 0.4$, ans 0.216 M1A1f
8	(i)	Binomial stated $1 - 0.9648$ $= 0.0352$ (3 sfs) or $\frac{9}{256}$	M1 M1 A1 [3]	or implied by $C \times 0.5^r$ or use of table or ${}^8C_7 \times 0.5^7 \times 0.5 + 0.5^8$ fully correct method or $0.5^7 \times 0.5 + 0.5^8$ or $0.5^8 + 0.5^8$ $1 - (0.5^8 + 8 \times 0.5^8 + {}^8C_2 0.5^8 \dots)$ all correct
8	(ii)	(a) ${}^{22}C_{11} \times 0.5^{11} \times 0.5^{11}$ $= 0.168$ (3 sfs)	M1 A1 [2]	Fully correct method. Not ISW eg 0.168^2 or 2×0.168 or $1 - 0.168$: M0A0

Question			Answer	Marks	Guidance
8	(ii)	(b)	$1 - \text{"0.168"}$ $\frac{1}{2}(1 - \text{"0.168"})$ $= 0.416$ (3 sfs)	M1 M1 A1 [3]	or $0.5^{22}({}^{22}C_{12} + {}^{22}C_{13} + {}^{22}C_{14} + \dots + 22 + 1)$ All 11 correct terms seen, or correct ans: M2 or $P(X = 12, 13, \dots, 21, 22)$ stated or implied with ≥ 2 terms shown or one extra term M1 or $1 - ({}^{22}C_{12} + {}^{22}C_{13} + {}^{22}C_{14} + \dots + 22 + 1)$ $1 -$ all 12 correct terms M2 or similar marks for $P(X = 10, 9, 8, \dots, 0)$
		(a)	9P_4 or ${}^9!/5!$ or ${}^9C_4 \times 4!$ $= 3024$	M1 A1 [2]	alone oe eg ${}^9C_1 \times {}^8C_1 \times {}^7C_1 \times {}^6C_1$ or $9 \times 8 \times 7 \times 6$
9	(i)	(b)	8P_3 or $8 \times 7 \times 6$ oe or ${}^8C_3 \times 3!$ $\times 5$ (or 5C_1) $= 1680$	M1 M1 A1 [3]	Allow $\times \dots$ or $\div \dots$ Correct $\times 5$ or ${}^8C_3 \times 5$ (or 5C_1) Not ISW, eg $1680/3024$: M1M1A0
					SC: consistent use of with replacement in (i) (or if only (a) or (b) attempted) (ia) M0A0 (ib) 999×5 or 4995 M1 M0A0

Question			Answer	Marks	Guidance	
9	(ii)	(a)	${}^5C_3 \times {}^4C_1$ or 5C_4 oe ${}^5C_3 \times {}^4C_1 + {}^5C_4$ oe correct method so far (= 45) $\div {}^9C_4$ Allow anything $\div {}^9C_4$ $= \frac{5}{14}$ or 0.357 (3 sfs) oe, eg $\frac{35}{98}$ or $\frac{45}{126}$	M1 M1 M1 A1 [4]	${}^5P_3 \times 4 \times 4$ or $5!$ (or 5P_4) $960 + 120$ oe correct method so far $\div {}^9P_4$ [must involve any P or any !] $\div {}^9P_4$ Marks must come from one method, not mixture of two methods	$\frac{5}{9} \times \frac{4}{8} \times \frac{3}{7} \times \frac{4}{6}$ Allow \times or $+$... $\times 4$ correct method so far $\frac{5}{9} \times \frac{4}{8} \times \frac{3}{7} \times \frac{2}{6}$ Allow \times or $+$... or: $\frac{5}{9} \times \frac{4}{8} \times \frac{3}{7} \times \frac{4}{6}$ or $\frac{5}{9} \times \frac{4}{8} \times \frac{3}{7}$ M1 $\frac{5}{9} \times \frac{4}{8} \times \frac{3}{7} \times \frac{4}{6} \times 3 + \frac{5}{9} \times \frac{4}{8} \times \frac{3}{7}$ M1 NB $\frac{5}{9} \times \frac{4}{8} \times \frac{3}{7} \times 3 = \frac{5}{14}$ M0M0M0A0
9	(ii)	(b)	9, 8, 7, 4 or 9, 8, 6, 5 No mark yet 2 $\div {}^9C_4$ oe Must be (1 or 2 or 4) $\div {}^9C_4$ $= \frac{1}{63}$ oe or 0.0159 (3 sfs)	M1 M1 A1 [3]	$\frac{1}{9} \times \frac{1}{8} \times \frac{1}{7} \times \frac{1}{6}$ or $\frac{4}{9} \times \frac{3}{8} \times \frac{2}{7} \times \frac{1}{6}$ Allow \times or $+$... $\times 4! \times 2$ or $\times 2$ fully correct method NB Marks from one method only, not mixed methods SC: consistent use of with replacement in (ii), (or if only (a) or (b) attempted) (iia) $(\frac{5}{9})^4$ M1 $+ {}^4C_3 (\frac{5}{9})^3 (\frac{4}{9})$ (= 0.400) M1 M0A0 (iib) $(\frac{1}{9})^4$ (=0.000152) M1 attempt find no of gps M1A0	$4! + 4!$ or $2 \times 4!$ oe $\div {}^9P_4$ or \div (i)(a) oe Must be (96 or 48 or 24) $\div {}^9P_4$ $\frac{2}{9} \times \frac{2}{8} \times \frac{1}{7} \times \frac{1}{6}$ allow \times or $+$... M1 $\times 4! / 4 \times 2$ fully correct method M1 $1 - ((\frac{4}{9})^4 + 4(\frac{4}{9})^3(\frac{5}{9}) + {}^4C_2(\frac{4}{9})^2(\frac{5}{9})^2)$ M2 One term missing or extra or wrong M1