OCR

General Certificate of Education
Advanced Subsidiary (AS) and Advanced Level

MATHEMATICS

S1

Probability & Statistics 1

Additional materials:
Answer paper
Graph paper
List of Formulae

SPECIMEN PAPER

TIME   1 hour 20 minutes

INSTRUCTIONS TO CANDIDATES

Write your name, Centre number and candidate number in the spaces provided on the answer paper.

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 graphic 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 60.

Questions carrying smaller numbers of marks are printed earlier in the paper, and questions carrying larger numbers of marks later in the paper.

You are reminded of the need for clear presentation in your answers.

This question paper consists of 4 printed pages.

1   Janet and John wanted to compare their daily journey times to work, so they each kept a record of their journey times for a few weeks. Janet's daily journey times, x minutes, for a period of 25 days, were summarised by $\Sigma x = 2120$ and $\Sigma x^2 = 180\,044$. Calculate the mean and standard deviation of Janet's journey times.   [3]

John's journey times had a mean of 79.7 minutes and a standard deviation of 6.22 minutes. Describe briefly, in everyday terms, how Janet and John's journey times compare.   [2]

2   Two independent assessors awarded marks to each of 5 projects. The results were as shown in the table.

Project	A	B	C	D	E
First assessor	38	91	62	83	61
Second assessor	56	84	41	85	62

Calculate Spearman's rank correlation coefficient for the data.   [4]

Show, by sketching a suitable scatter diagram, how two assessors might have assessed 5 projects in such a way that Spearman's rank correlation coefficient for their marks was $+1$ while the product moment correlation coefficient for their marks was not $+1$. (Your scatter diagram need not be drawn accurately to scale.)   [2]

3   Each packet of the breakfast cereal Fizz contains one plastic toy animal. There are five different animals in the set, and the cereal manufacturers use equal numbers of each. Without opening a packet it is impossible to tell which animal it contains. A family has already collected four different animals at the start of a year and they now need to collect an elephant to complete their set. The family is interested in how many packets they will need to buy before they complete their set.

  (i)   Stating any necessary assumption, name an appropriate distribution with which to model this situation. What is the expected number of packets that they family will need to buy?   [3]

  (ii) Find the probability that the family will complete their set with the third packet they buy after the start of the year.   [2]

  (iii) Find the probability that, in order to complete their collection, the family will need to buy more than 4 packets after the start of the year.   [3]

4   A sixth-form class consists of 7 girls and 5 boys. Three students from the class are chosen at random. The number of boys chosen is denoted by the random variable X. Show that
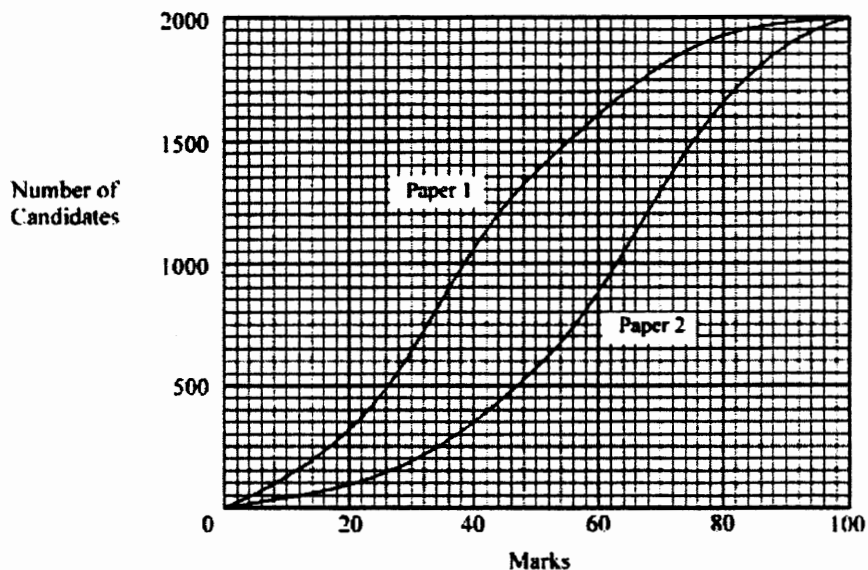
  (i) $P(X = 0) = \frac{7}{44}$,   [2]

  (ii) $P(X = 2) = \frac{7}{22}$.   [3]

The complete probability distribution of X is shown in the following table.

x	0	1	2	3
$P(X = x)$	$\frac{7}{44}$	$\frac{21}{44}$	$\frac{7}{22}$	$\frac{1}{22}$

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

5

The diagram shows the cumulative frequency graphs for the marks scored by the candidates in an examination. The 2000 candidates each took two papers; the upper curve shows the distribution of marks on paper 1 and the lower curve shows the distribution on paper 2. The maximum mark on each paper was 100.

(i) Use the diagram to estimate the median mark for each of paper 1 and paper 2, and the interquartile range for paper 1. **[6]**

(ii) State with a reason which of the two papers you think was the easier one. **[2]**

(iii) The candidates' marks for the two papers could also be illustrated by means of a pair of box-and-whisker plots. Give two brief comments on any advantages or disadvantages in using cumulative frequency graphs and box-and-whisker plots to represent the data. **[2]**

6 Items from a production line are examined for any defects. The probability that any item will be found to be defective is 0.15, independently of all other items.

(i) A batch of 16 items is inspected. Using tables of cumulative binomial probabilities, or otherwise, find the probability that

(a) at least 4 items in the batch are defective, **[2]**

(b) exactly 4 items in the batch are defective. **[2]**

(ii) Five batches, each containing 16 items, are taken.

(a) Find the probability that at most 2 of these 5 batches contain at least 4 defective items. **[4]**

(b) Find the expected number of batches that contain at least 4 defective items. **[2]**

7 An experiment was conducted to see whether there was any relationship between the maximum tidal current, $y\,\mathrm{cm\,s^{-1}}$, and the tidal range, x metres, at a particular marine location. [The *tidal range* is the difference between the height of high tide and the height of low tide.] Readings were taken over a period of 12 days, and the results are shown in the following table.

x	2.0	2.4	3.0	3.1	3.4	3.7	3.8	3.9	4.0	4.5	4.6	4.9
y	15.2	22.0	25.2	33.0	33.1	34.2	51.0	42.3	45.0	50.7	61.0	59.2

$$[\Sigma x = 43.3, \Sigma y = 471.9, \Sigma x^2 = 164.69, \Sigma y^2 = 20\,915.75, \Sigma xy = 1837.78.]$$

The scatter diagram below illustrates the data.
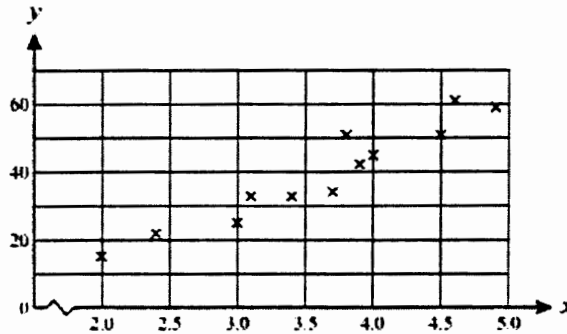
(i)   Calculate the product moment correlation coefficient for the data, and comment briefly on your answer with reference to the appearance of the scatter diagram. [4]

(ii) Calculate the equation of the regression line of maximum tidal current on tidal range. [3]

(iii) Estimate the maximum tidal current on a day when the tidal range is 4.2 m, and indicate briefly how reliable an estimate you consider your answer to be. [3]

(iv) It is suggested that the equation found in part (ii) could be used to predict the maximum tidal current on a day when the tidal range is 15 m. Comment briefly on the validity of this suggestion. [1]

1	Mean is 84.8 minutes	B1	
	Standard deviation $= \sqrt{\dfrac{180044}{25} - 84.8^2}$	M1	May be implied
	$= 3.27$ minutes	A1 **3**	
	John's average time is about 5 mins less than Janet's	B1✓	
	John's times are more variable than Janet's	B1✓ **2**	

2	Ranks are: $\begin{matrix}1 & 5 & 3 & 4 & 2\\ 2 & 4 & 1 & 5 & 3\end{matrix}$	B1	Or with ranks reversed
	Values of d are $-1, 1, 2, -1, -1$	M1	Or reversed, or values of d^2
	$r_s = 1 - \dfrac{6 \times 8}{5 \times 24} = 0.6$	M1	Correct formula for Spearman used
		A1 **4**	Correct answer (fraction or decimal)
	(e.g.)	B2 **2**	For 5 points, showing any non-linear 'increasing' relationship

3	(i) Each packet is equally likely to contain any of the 5 animals, independently of other packets	B1	Allow either 'equally likely' or 'independent'
	Geometric distribution	B1	No need to state $p = \frac{1}{5}$ here
	Expected number is 5	B1 **3**	
	(ii) $\left(\frac{4}{5}\right)^2 \times \left(\frac{1}{5}\right) = \frac{16}{125}$ or 0.128	M1	Any numerical '$q^n p$' calculation
		A1 **2**	Correct answer
	(iii) $\left(\frac{4}{5}\right)^4$ or $1 - \left\{\frac{1}{5} + \left(\frac{4}{5}\right)\left(\frac{1}{5}\right) + \left(\frac{4}{5}\right)^2\left(\frac{1}{5}\right) + \left(\frac{4}{5}\right)^3\left(\frac{1}{5}\right)\right\}$	M1	Allow M mark even if there is an error of 1 in the number of terms
		A1	For correct expression for the answer
	$\frac{256}{625}$ or 0.4096 or 0.410	A1 **3**	Correct fraction or decimal

4	(i) $P(X = 0) = \binom{7}{3} \div \binom{12}{3} = \frac{35}{220} = \frac{7}{44}$	M1	For ratio of relevant $\binom{n}{r}$ terms, or equivalent
		A1 **2**	Given answer correctly shown
	(ii) $P(X = 2) = P(2 \text{ boys and 1 girl})$	B1	May be implied
	$= \binom{7}{1} \times \binom{5}{2} \div \binom{12}{3} = \frac{7 \times 10}{220} = \frac{7}{22}$	M1	For use of relevant $\binom{n}{r}$ terms, or complete alternative multiplication/addition of probs
		A1 **3**	Given answer correctly shown
	$E(X) = 0 \times \frac{7}{44} + 1 \times \frac{21}{44} + 2 \times \frac{7}{22} + 3 \times \frac{1}{22} = \frac{5}{4}$	M1	For correct calculation process
		A1	Correct answer (fraction or decimal form)
	$E(X^2) = 0 \times \frac{7}{44} + 1 \times \frac{21}{44} + 4 \times \frac{7}{22} + 9 \times \frac{1}{22} = \frac{95}{44}$	B1	For correct numerical expression for $\Sigma x^2 p$
	$\text{Var}(X) = \frac{95}{44} - \left(\frac{5}{4}\right)^2 = \frac{105}{176}$ or 0.597 (to 3dp)	M1	For correct overall method for variance
		A1✓ **5**	For correct answer

5	(i)	Medians correspond to 1000 candidates	M1		Reading off at 1000; may be implied	
		$m_1 = 38$, $m_2 = 63$	A1		Correct value for either median	
			A1		For both correct	
		Quartiles correspond to 1500 and 500 candidates	M1		Reading off at either; may be implied	
		$q_3 = 56$, $q_1 = 26$	A1		Both correct	
		IQR $= 30$	A1✓	6		

(ii)	Paper 2 was easier	B1			
	Marks were higher on paper 2	B1	2	Or similar statement, e.g. 'higher median'	

(iii)	Possible valid comments include:			
	Box plots give quick direct comparisons of medians and IQRs			
	Box plots don't include all the information that CF graphs do			
	CF graphs can be used to read off values both ways round	B1		For any one valid comment
	etc	B1	2	Any other valid comment

6	(i)	(a)	$1 - 0.7899 = 0.210(1)$	M1		Complement of relevant tabular value
				A1✓	2	Correct answer
		(b)	$0.9209 - 0.7899 = 0.131$	M1		Subtracting relevant tabular values
				A1	2	Correct answer
	(ii)	(a)	$0.790^5 + 5 \times 0.790^4 \times 0.210 + 10 \times 0.790^3 \times 0.210^2$	M1		Any use of B(5, 0.210)
				M1		Correct 3 cases identified
				A1✓		Correct numerical expression for required probability, with their value from (i)(a)
			$= 0.934$	A1✓	4	
		(b)	$5 \times 0.210 = 1.05$	M1		For relevant use of np
				A1	2	For correct answer

7	(i)	$r = \dfrac{1837.78 - \dfrac{43.3 \times 471.9}{12}}{\sqrt{\left(164.69 - \dfrac{43.3^2}{12}\right)\left(20915.75 - \dfrac{471.9^2}{12}\right)}}$	M1		Or equivalent; may be implied
		$= 0.956$	A1		
		The value is close to $+1$,	B1		For relating the value to 1
		and the points in the diagram lie (fairly) close to a straight line with positive gradient	B1	4	For a reasonable comment about linearity
	(ii)	Gradient of regression line is			
		$\dfrac{1837.78 - \dfrac{43.3 \times 471.9}{12}}{164.69 - \dfrac{43.3^2}{12}} = 15.9789$	M1		May be implied if calculator routine used
		$y - \dfrac{471.9}{12} = 15.9789\left(x - \dfrac{43.3}{12}\right)$	M1		May similarly be implied
		$y = 16.0x - 18.3$	A1	3	
	(iii)	$y = 16.0 \times 4.2 - 18.3$	M1		
		Current is $48.8\,\text{cms}^{-1}$	A1✓		Units required in answer
		Diagram indicates some uncertainty, e.g. $\pm 5\,\text{cms}^{-1}$	B1	3	Allow any reasonable comment
	(v)	The prediction would be (very) unreliable because of the extrapolation involved	B1	1	For conclusion and idea of extrapolation