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]**

1(i) $P(X=2) = P(0,2) + P(1,1) + P(2,0)$

$= (0.7 \times 0.1) + (0.2 \times 0.2) + (0.1 \times 0.7)$

$= 0.07 + 0.04 + 0.07$

$= \underline{\underline{0.18}}$ ✓

(iii)

x	0	1	2	3	4
$P(X=x)$	0.49	0.28	0.18	0.04	0.01
px	0	0.28	0.36	0.12	0.04
x^2	0	1	4	9	16
px^2	0	0.28	0.72	0.36	0.16

$$\sum px = 0+0.28+0.36+0.12+0.04$$
$$= 0.8$$

$$\underline{\underline{\mu = 0.8}} \checkmark$$

$$\sum px^2 = 1.52$$

$$\sigma^2 = \sum px^2 - \mu^2$$
$$= 1.52 - 0.8^2$$
$$= \underline{\underline{0.88}} \checkmark$$

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]**

2(i)(a) $b = \dfrac{S_{xy}}{S_{xx}}$

$$S_{xy} = \Sigma xy - \frac{\Sigma x \Sigma y}{n}$$

$$S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

$$S_{xy} = 8736.9 - \frac{(202 \times 245.3)}{7}$$

$$= 1658.2428571$$

$$S_{xx} = 7300 - \frac{202^2}{7}$$

$$= 1470.857142$$

$$b = \frac{1658.2428571}{1470.857142} = 1.12739899$$

$$= 1.13 \text{ (3sf)} \checkmark$$

(b) $y = a + bx$

$a = \bar{y} - b\bar{x}$

$\bar{y} = \frac{245.3}{7} = 35.0428571$

$\bar{x} = \frac{202}{7} = 28.857142$

$a = 2.4342857$

$= 2.43$ (3sf)

$\underline{\underline{y = 2.43 + 1.13x}}$ ✓

2(ii)(a) $y = 2.43 + 1.13x$

$y = 2.43 + (1.13 \times 30)$

$= \underline{36.33\text{cm}}$ ✓

(b) $y = 2.43 + (1.13 \times 100)$

$= \underline{115.43\text{cm}}$ ✓

(iii) The higher value suggests that the estimate for y when $x = 30$ would be reasonable ✓

when $x = 100$ this is outside the range of values collected so any estimate based on this would be inaccurate (extrapolation) ✓

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]**

3(i) $X \sim Geo(\frac{1}{8})$

(a) $P(X=x) = \left(\frac{7}{8}\right)^2 \times \frac{1}{8} = \frac{49}{512}$ ✓

(b) $P(X>3) = \left(\frac{7}{8}\right)^3 = \frac{343}{512}$ ✓

(ii) $E(X) = \frac{1}{p} = \frac{1}{1/8} = \underline{\underline{8}}$ ✓

(iii) $Y \sim B(15, \frac{1}{8})$

$$P(Y=2) = \binom{15}{2} \times \left(\frac{1}{8}\right)^2 \times \left(\frac{7}{8}\right)^{13}$$

$= 0.2891438866$

$= \underline{\underline{0.289}}$ (3sf) ✓

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]**

(i)

STUDENT	T1	T2	R1	R2	d	d^2
A	73	62	1	3	-2	4
B	67	50	2	5	-3	9
C	60	61	3	4	-1	1
D	48	76	4	1	3	9
E	39	65	5	2	3	9
						32

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 32}{5(24)}$$

$$= -0.6$$

(ii) Tutors 1 and 3 disagree most
This is shown by the negative
value of $r_s = -0.9$ (very close
to -1)

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)
6	1 2 3 5 6 8 9	(7)
7	0 0 2 4 5 6 7 8	(8)
8	0	(1)
9	7	(1)
		23

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]**

5(i) median position $= \frac{n+1}{2} = \frac{23+1}{2} = \frac{24}{2} = 12^{th}$

median = 68 ✓

lower quartile position $= \frac{n+1}{4} = \frac{24}{4} = 6^{th}$

lower quartile = 59

upper quartile position $= \frac{3(n+1)}{4} = \frac{72}{4} = 18^{th}$

upper quartile = 75

IQR = UQ − LQ

= 75 − 59

IQR = 16 ✓

5(ii) The interquartile range is not affected by outlier values ✓

(iii) An advantage of using a stem-and-leaf diagram is that the original data is preserved. ✓

A disadvantage of using a stem-and-leaf diagram is that it is harder to find the median and quartiles ✓ in a stem-and-leaf and, thus, make comparisons to other data sets.

5(iv) Every value in the CODED stem-and-leaf has been reduced by 50.

The mean is therefore $18.1 + 50 = \underline{\underline{68.1}}$ ✓

The standard deviation is UNAFFECTED by CODING.

The standard deviation remains $\underline{\underline{9.7}}$ ✓

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]**

(i) (a) $8! = 40320$ ✓

(b) A G A G A G A G

$4! \times 4! \times 2$

G A G A G A G A

$\frac{4! \times 4! \times 2}{8!} = \frac{1}{35}$

$\frac{4! \times 4! \times 2}{8!} = \frac{24 \times 24 \times 2}{40320} = \frac{1152}{40320} = \frac{1}{35}$ ✓

(ii) (a) Algebra permutations $= 4! = 24$

Geometry permutations $= 4! = 24$

$24 \times 24 = \underline{\underline{576}}$ ✓

(b) _ _ _ A   H _ _ _

ALGEBRA   GEOMETRY

There are 3! ways that A can be the last question in the ALGEBRA section

There are 3! ways that H can be the first question in GEOMETRY section

$$\frac{3!}{24} \times \frac{3!}{24}$$

$$\frac{6}{24} \times \frac{6}{24}$$

$$\frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$$ ✓

6(ii)(c)

To be separated by <u>MORE</u> than 4 questions there has to be at least <u>5</u> questions between B and J

possible positions

B × × × × × J J

ALGEBRA GEOMETRY

$$\frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$$

_ B × × × × × J ← only possible position

$$\frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$$

$$\frac{1}{8} + \frac{1}{16} = \frac{3}{16} \checkmark$$

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]**

7(i) Binomial ✓ $X \sim B(12, 0.1)$ ✓

The probability of a plate having a fault must be constant. ✓

The probability of any plate being faulty must be independent from any other plate being faulty. ✓

7(ii)(a)

$$P(X=3) = \binom{12}{3} \times 0.1^3 \times 0.9^9$$

$$= 220 \times 0.001 \times 0.387420489$$

$$= 0.08523250758$$

$$= 0.0852 \text{ (3sf)}$$

(ii)(b) $P(x \geq 1) = 1 - P(x \leq 0)$

$= 1 - P(x = 0)$

$= 1 - 0.2824$

$= 0.7176$ ✓

$= 0.718$

7 (iii) Set up another binomial distribution

From the previous part $P(X \geq 1) = 0.7176$

$$Y \sim B(4, 0.7176)$$

$$P(Y \leq 2) = P(Y=0) + P(\cancel{X}=1) + P(Y=2)$$

$\binom{4}{2}$

$$= 0.2824^4 + (4 \times 0.2824^3 \times 0.7176) + (4C2 \times 0.2824^2 \times 0.7176^2)$$

$$= 0.3174079749$$

$$= 0.317 \text{ (3sf)}$$

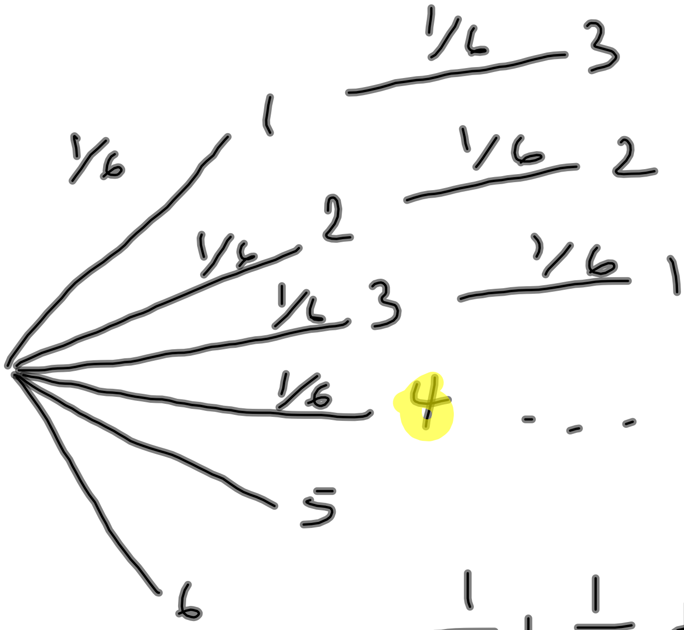
8 A game uses an unbiased die with faces numbered 1 to 6. The die is thrown once. If it shows 4 or 5 or 6 then this number is the final score. If it shows 1 or 2 or 3 then the die is thrown again and the final score is the sum of the numbers shown on the two throws.

(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]**

8(i)



$\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

$\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

$\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

. . . . . . $\frac{6}{36}$

$$\frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{6}{36}$$

$$= \frac{9}{36} = \frac{1}{4}$$ ✓

8(ii) Must have rolled either 4, 5 or 6
so $P(4 | 1 \text{ throw}) = 1/3$ ✓

(iii)

1st throw

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

2nd throw

$$P(4 | 2 \text{throws}) = \frac{3}{18} = \frac{1}{6}$$ ✓