Section A (36 marks) **Duration:** 1 hour 30 minutes

1 In a traffic survey, the number of people in each car passing the survey point is recorded. The results are given in the following frequency table.

Number of people	1	2	3	4
Frequency	50	31	16	5

(i) Write down the median and mode of these data. **[2]**

(ii) Draw a vertical line diagram for these data. **[2]**

(iii) State the type of skewness of the distribution. **[1]**

2 There are 14 girls and 11 boys in a class. A quiz team of 5 students is to be chosen from the class.

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

(ii) If the team must include 3 girls and 2 boys, find how many different teams are possible. **[3]**

3 Dwayne is a car salesman. The numbers of cars, x, sold by Dwayne each month during the year 2008 are summarised by

$$n = 12, \qquad \Sigma x = 126, \qquad \Sigma x^2 = 1582.$$

(i) Calculate the mean and standard deviation of the monthly numbers of cars sold. **[3]**

(ii) Dwayne earns £500 each month plus £100 commission for each car sold. Show that the mean of Dwayne's monthly earnings is £1550. Find the standard deviation of Dwayne's monthly earnings. **[3]**

(iii) Marlene is a car saleswoman and is paid in the same way as Dwayne. During 2008 her monthly earnings have mean £1625 and standard deviation £280. Briefly compare the monthly numbers of cars sold by Marlene and Dwayne during 2008. **[2]**

4 The table shows the probability distribution of the random variable X.

r	10	20	30	40
$\mathrm{P}(X = r)$	0.2	0.3	0.3	0.2

(i) Explain why $\mathrm{E}(X) = 25$. **[1]**

(ii) Calculate $\mathrm{Var}(X)$. **[3]**

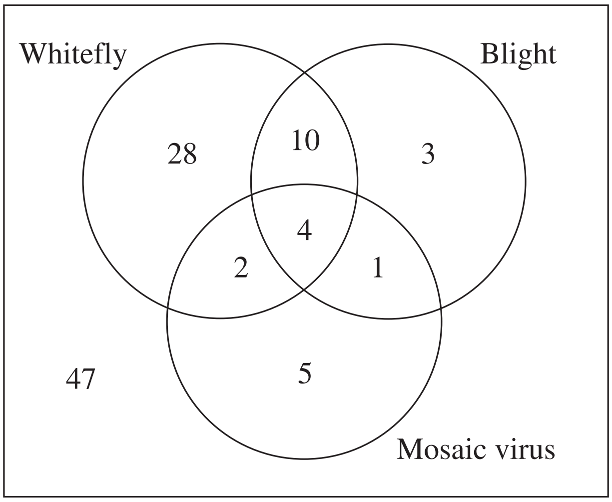
5 The frequency table below shows the distance travelled by 1200 visitors to a particular UK tourist destination in August 2008.

Distance (d miles)	$0 \leqslant d < 50$	$50 \leqslant d < 100$	$100 \leqslant d < 200$	$200 \leqslant d < 400$
Frequency	360	400	307	133

(i) Draw a histogram on graph paper to illustrate these data. **[5]**

(ii) Calculate an estimate of the median distance. **[3]**

6 Whitefly, blight and mosaic virus are three problems which can affect tomato plants. 100 tomato plants are examined for these problems. The numbers of plants with each type of problem are shown in the Venn diagram. 47 of the plants have none of the problems.



(i) One of the 100 plants is selected at random. Find the probability that this plant has

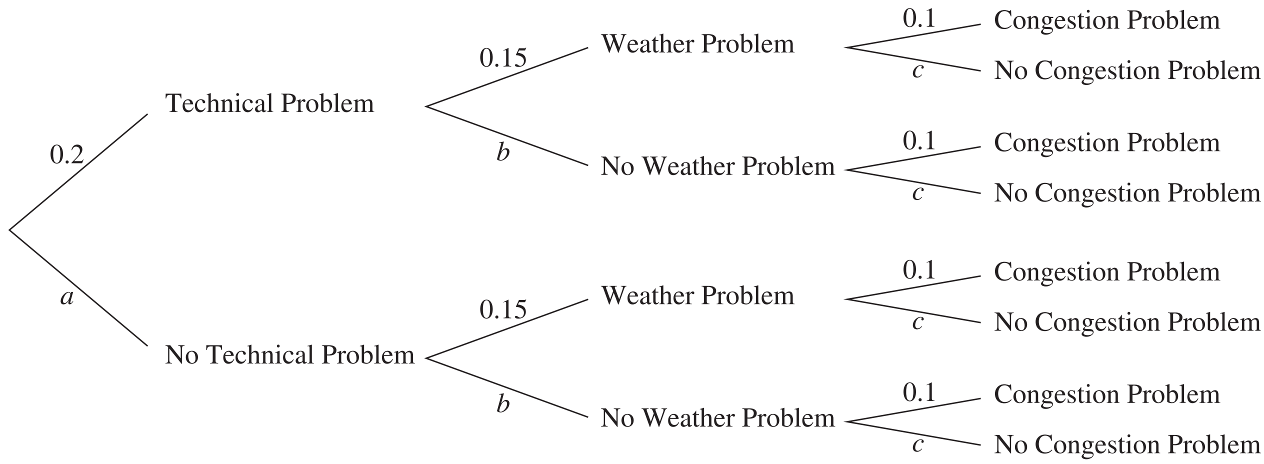
(A) at most one of the problems, **[1]**

(B) exactly two of the problems. **[2]**

(ii) Three of the 100 plants are selected at random. Find the probability that all of them have at least one of the problems. **[3]**

Section B (36 marks)

7 Laura frequently flies to business meetings and often finds that her flights are delayed. A flight may be delayed due to technical problems, weather problems or congestion problems, with probabilities 0.2, 0.15 and 0.1 respectively. The tree diagram shows this information.



(i) Write down the values of the probabilities a, b and c shown in the tree diagram. **[2]**

One of Laura's flights is selected at random.

(ii) Find the probability that Laura's flight is not delayed and hence write down the probability that it is delayed. **[4]**

(iii) Find the probability that Laura's flight is delayed due to just one of the three problems. **[4]**

(iv) Given that Laura's flight is delayed, find the probability that the delay is due to just one of the three problems. **[3]**

(v) Given that Laura's flight has no technical problems, find the probability that it is delayed. **[3]**

(vi) In a particular year, Laura has 110 flights. Find the expected number of flights that are delayed. **[2]**

8 The Department of Health 'eat five a day' advice recommends that people should eat at least five portions of fruit and vegetables per day. In a particular school, 20% of pupils eat at least five a day.

(i) 15 children are selected at random.

(A) Find the probability that exactly 3 of them eat at least five a day. **[3]**

(B) Find the probability that at least 3 of them eat at least five a day. **[3]**

(C) Find the expected number who eat at least five a day. **[2]**

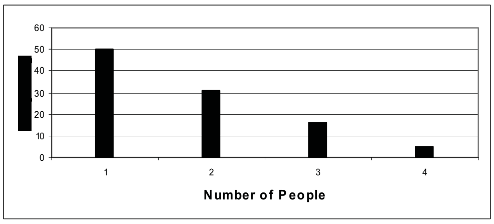
A programme is introduced to encourage children to eat more portions of fruit and vegetables per day. At the end of this programme, the diets of a random sample of 15 children are analysed. A hypothesis test is carried out to examine whether the proportion of children in the school who eat at least five a day has increased.

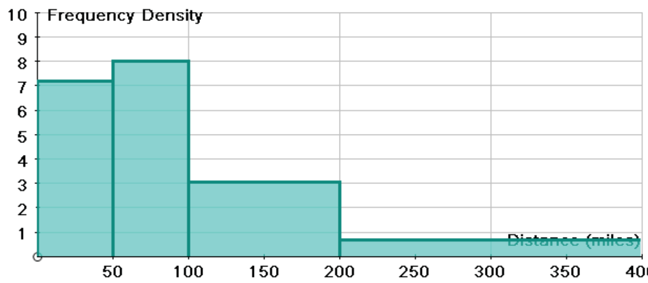
(ii) (A) Write down suitable null and alternative hypotheses for the test.

(B) Give a reason for your choice of the alternative hypothesis. **[4]**

(iii) Find the critical region for the test at the 10% significance level, showing all of your calculations. Hence complete the test, given that 7 of the 15 children eat at least five a day. **[6]**

4766 Statistics 1

Q1 (i)	Median = 2 Mode = 1	B1 CAO B1 CAO	**2**
(ii)		S1 labelled linear scales on both axes H1 heights	**2**
(iii)	Positive	B1	**1**
		TOTAL	**5**
Q2 (i)	$\binom{25}{5}$ different teams = 53130	M1 for $\binom{25}{5}$ A1 CAO	**2**
(ii)	$\binom{14}{3}\times\binom{11}{2}=364\times55=20020$	M1 for either combination M1 for product of both A1 CAO	**3**
		TOTAL	**5**
Q3 (i)	Mean = $\frac{126}{12}$ = 10.5 Sxx = $1582-\frac{126^2}{12}$ = 259 s = $\sqrt{\frac{259}{11}}$ = 4.85	B1 for mean M1 for attempt at *Sxx* A1 CAO	**3**
(ii)	New mean = 500 + 100 $\times$10.5 = 1550 New s = 100 $\times$4.85 = 485	B1 <u>ANSWER GIVEN</u> M1A1FT	**3**
(iii)	On average Marlene sells more cars than Dwayne. Marlene has less variation in monthly sales than Dwayne.	E1 E1FT	**2**
		TOTAL	**8**

Q4 (i)	E(X) = 25 because the distribution is symmetrical. Allow correct calculation of Σrp	E1 <u>ANSWER GIVEN</u>	**1**
(ii)	E(X^2) = $10^2 \times 0.2 + 20^2 \times 0.3 + 30^2 \times 0.3 + 40^2 \times 0.2$ = 730 Var(X) = 730 – 25^2 = 105	M1 for $\Sigma r^2 p$ (at least 3 terms correct) M1dep for – 25² A1 CAO	**3**
		TOTAL	**4**
Q5 (i)	Distance / freq / width / f dens: 0- / 360 / 50 / 7.200 50- / 400 / 50 / 8.000 100- / 307 / 100 / 3.070 200-400 / 133 / 200 / 0.665 	M1 for fds A1 CAO Accept any suitable unit for fd such as eg freq per 50 miles. L1 linear scales on both axes and label W1 width of bars H1 height of bars	**5**
(ii)	Median = 600th distance Estimate = 50 + $^{240}/_{400} \times 50$ = 50 + 30 = 80	B1 for 600th M1 for attempt to interpolate A1 CAO	**3**
		TOTAL	**8**
Q6 (i)	(*A*) P(at most one) = $\dfrac{83}{100} = 0.83$ (*B*) P(exactly two) = $\dfrac{10+2+1}{100} = \dfrac{13}{100} = 0.13$	B1 aef M1 for (10+2+1)/100 A1 aef	**1** **2**
(ii)	P(all at least one) = $\dfrac{53}{100} \times \dfrac{52}{99} \times \dfrac{51}{98} = \dfrac{140556}{970200} = 0.145$	M1 for $\frac{53}{100}\times$ M1*dep* for product of next 2 correct fractions A1 CAO	**3**
		TOTAL	**6**

Q7 (i)	$a = 0.8$, $b = 0.85$, $c = 0.9$.	B1 for any one B1 for the other two	**2**
(ii)	P(Not delayed) = $0.8 \times 0.85 \times 0.9 = 0.612$ P(Delayed) = $1 - 0.8 \times 0.85 \times 0.9 = 1 - 0.612 = 0.388$	M1 for product A1 CAO M1 for 1 – P(delayed) A1FT	**4**
(iii)	P(just one problem) $= 0.2\times0.85\times0.9 + 0.8\times0.15\times0.9 + 0.8\times0.85\times0.1$ $= 0.153 + 0.108 + 0.068 = 0.329$	B1 one product correct M1 three products M1 sum of 3 products A1 CAO	**4**
(iv)	P(Just one problem | delay) $= \dfrac{\text{P(Just one problem and delay)}}{\text{P(Delay)}} = \dfrac{0.329}{0.388} = 0.848$	M1 for numerator M1 for denominator A1FT	**3**
(v)	P(Delayed | No technical problems) *Either* $= 0.15 + 0.85 \times 0.1 = 0.235$ *Or* $= 1 - 0.9 \times 0.85 = 1 - 0.765 = 0.235$ *Or* $= 0.15 \times 0.1 + 0.15 \times 0.9 + 0.85 \times 0.1 = 0.235$ *Or (using conditional probability formula)* $\dfrac{\text{P(Delayed and no technical problems)}}{\text{P(No technical problems)}}$ $= \dfrac{0.8\times0.15\times0.1+0.8\times0.15\times0.9+0.8\times0.85\times0.1}{0.8}$ $= \dfrac{0.188}{0.8} = 0.235$	M1 for 0.15 + M1 for second term A1CAO M1 for product M1 for 1 – product A1CAO M1 for all 3 products M1 for sum of all 3 products A1CAO M1 for numerator M1 for denominator A1CAO	**3**
(vi)	Expected number = $110 \times 0.388 = 42.7$	M1 for product A1FT	**2**
		TOTAL	**18**

Q8 (i)	X ~ B(15, 0.2) (*A*) $\text{P}(X = 3) = \binom{15}{3} \times 0.2^3 \times 0.8^{12} = 0.2501$ OR from tables $0.6482 - 0.3980 = 0.2502$ (*B*) $\text{P}(X \geq 3) = 1 - 0.3980 = 0.6020$ (*C*) E(*X*) = *np* = 15 × 0.2 = 3.0	M1 $0.2^3 \times 0.8^{12}$ M1 $\binom{15}{3} \times p^3 q^{12}$ A1 CAO OR: M2 for 0.6482 – 0.3980 A1 CAO M1 P(X≤2) M1 1-P(X≤2) A1 CAO M1 for product A1 CAO	**3** **3** **2**
(ii)	(*A*) Let *p* = probability of a randomly selected child eating at least 5 a day H_0: $p = 0.2$ H_1: $p > 0.2$ (*B*) H_1 has this form as the proportion who eat at least 5 a day is expected to <u>increase</u>.	B1 for definition of *p* in context B1 for H_0 B1 for H_1 E1	**4**
(iii)	Let $X \sim \text{B}(15, 0.2)$ $\text{P}(X \geq 5) = 1 - \text{P}(X \leq 4) = 1 - 0.8358 = 0.1642 > 10\%$ $\text{P}(X \geq 6) = 1 - \text{P}(X \leq 5) = 1 - 0.9389 = 0.0611 < 10\%$ So critical region is {6,7,8,9,10,11,12,13,14,15} 7 lies in the critical region, so we reject null hypothesis and we conclude that there is evidence to suggest that the proportion who eat at least five a day has increased.	B1 for 0.1642 B1 for 0.0611 M1 for at least one comparison with 10% A1 CAO for critical region *dep* on M1 and at least one B1 M1 *dep* for comparison A1 *dep* for decision and conclusion **in context**	**6**
		TOTAL	**18**