mock papers 1

1. A personnel manager wants to find out if a test carried out during an employee's interview and a skills assessment at the end of basic training is a guide to performance after working for the company for one year.

The table below shows the results of the interview test of 10 employees and their performance after one year.

Employee	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Interview test, x %.	65	71	79	77	85	78	85	90	81	62
Performance after one year, y %.	65	74	82	64	87	78	61	65	79	69

[You may use $\sum x^2 = 60\ 475, \sum y^2 = 53\ 122, \sum xy = 56\ 076$]

(a) Showing your working clearly, calculate the product moment correlation coefficient between the interview test and the performance after one year.

(5)

The product moment correlation coefficient between the skills assessment and the performance after one year is –0.156 to 3 significant figures.

(b) Use your answer to part (a) to comment on whether or not the interview test and skills assessment are a guide to the performance after one year. Give clear reasons for your answers.

(2)

Question 1 continued

Question 1 continued

Question 1 continued

Q1

(Total 7 marks)

2. Cotinine is a chemical that is made by the body from nicotine which is found in cigarette smoke. A doctor tested the blood of 12 patients, who claimed to smoke a packet of cigarettes a day, for cotinine. The results, in appropriate units, are shown below.

Patient	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*
Cotinine level, x	160	390	169	175	125	420	171	250	210	258	186	243

[You may use $\sum x^2 = 724\ 961$]

(a) Find the mean and standard deviation of the level of cotinine in a patient's blood.

(4)

(b) Find the median, upper and lower quartiles of these data.

(3)

A doctor suspects that some of his patients have been smoking more than a packet of cigarettes per day. He decides to use $Q_3+1.5(Q_3-Q_1)$ to determine if any of the cotinine results are far enough away from the upper quartile to be outliers.

(c) Identify which patient(s) may have been smoking more than a packet of cigarettes a day. Show your working clearly.

(4)

Research suggests that cotinine levels in the blood form a skewed distribution.

One measure of skewness is found using $\dfrac{(Q_1 - 2Q_2 + Q_3)}{(Q_3 - Q_1)}$.

(d) Evaluate this measure and describe the skewness of these data.

(3)

Question 2 continued

Question 2 continued

Question 2 continued

Q2

(Total 14 marks)

3. The histogram in Figure 1 shows the time taken, to the nearest minute, for 140 runners to complete a fun run.

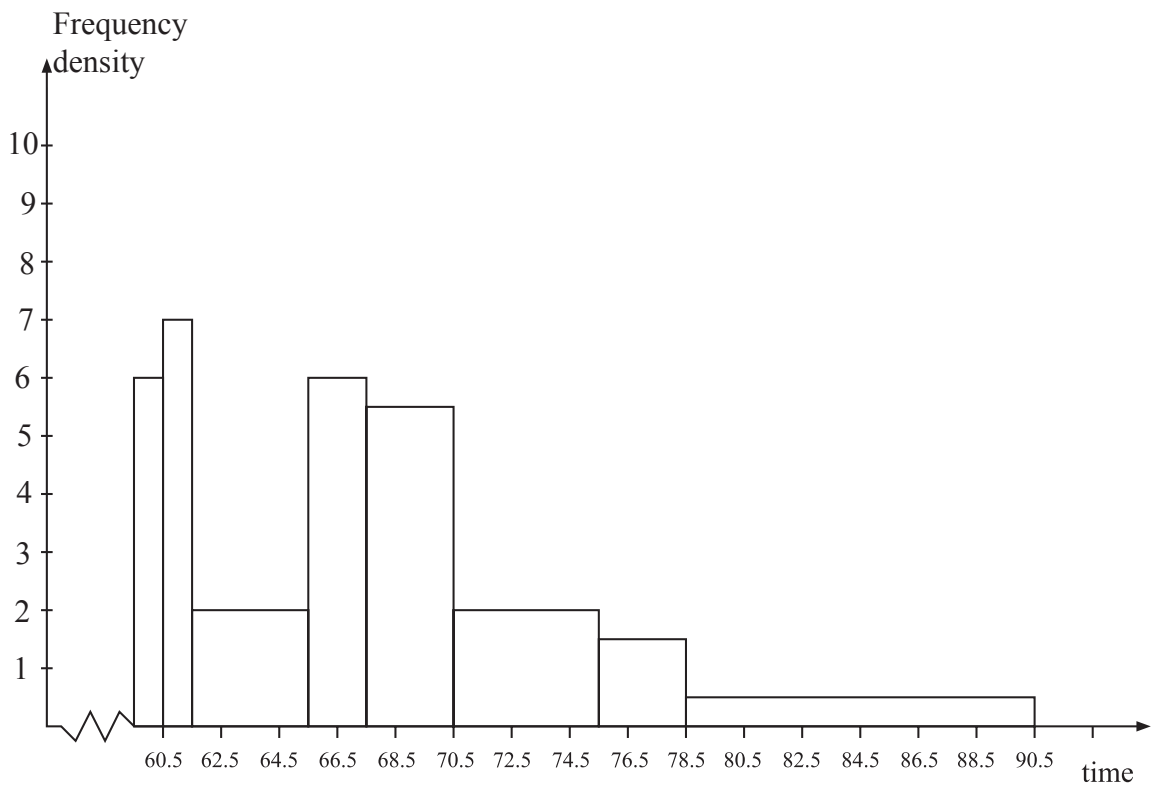


Figure 1

Use the histogram to calculate the number of runners who took between 78.5 and 90.5 minutes to complete the fun run.

(5)

Question 3 continued

Q3

(Total 5 marks)

4. A second hand car dealer has 10 cars for sale. She decides to investigate the link between the age of the cars, x years, and the mileage, y thousand miles. The data collected from the cars are shown in the table below.

Age, x (years)	2	2.5	3	4	4.5	4.5	5	3	6	6.5
Mileage, y (thousands)	22	34	33	37	40	45	49	30	58	58

[You may assume that $\sum x = 41, \sum y = 406, \sum x^2 = 188, \sum xy = 1818.5$]

(a) Find S_{xx} and S_{xy}.

(3)

(b) Find the equation of the least squares regression line in the form $y=a+bx$. Give the values of a and b to 2 decimal places.

(4)

(c) Give a practical interpretation of the slope b.

(1)

(d) Using your answer to part (b), find the mileage predicted by the regression line for a 5 year old car.

(2)

Question 4 continued

Question 4 continued

Question 4 continued

Q4

(Total 10 marks)

5. The following shows the results of a wine tasting survey of 100 people.

96 like wine *A*,
93 like wine *B*,
96 like wine *C*,
92 like *A* and *B*,
91 like *B* and *C*,
93 like *A* and *C*,
90 like all three wines.

(a) Draw a Venn Diagram to represent these data.

(6)

Find the probability that a randomly selected person from the survey likes

(b) none of the three wines,

(1)

(c) wine *A* but not wine *B*,

(2)

(d) any wine in the survey except wine *C*,

(2)

(e) exactly two of the three kinds of wine.

(2)

Given that a person from the survey likes wine *A*,

(f) find the probability that the person likes wine *C*.

(3)

Question 5 continued

Question 5 continued

Question 5 continued

(Total 16 marks)

Q5

6. The weights of bags of popcorn are normally distributed with mean of 200 g and 60% of all bags weighing between 190 g and 210 g.

(a) Write down the median weight of the bags of popcorn.

(1)

(b) Find the standard deviation of the weights of the bags of popcorn.

(5)

A shopkeeper finds that customers will complain if their bag of popcorn weighs less than 180 g.

(c) Find the probability that a customer will complain.

(3)

Question 6 continued

Q6

(Total 9 marks)

7. Tetrahedral dice have four faces. Two fair tetrahedral dice, one red and one blue, have faces numbered 0, 1, 2, and 3 respectively. The dice are rolled and the numbers face down on the two dice are recorded. The random variable R is the score on the red die and the random variable B is the score on the blue die.

(a) Find P(R=3 and B=0).

(2)

The random variable T is R multiplied by B.

(b) Complete the diagram below to represent the sample space that shows all the possible values of T.

3				
2		2		
1	0			
0				
B / R	**0**	**1**	**2**	**3**

Sample space diagram of *T*

(3)

(c) The table below represents the probability distribution of the random variable T.

t	0	1	2	3	4	6	9
P(T=t)	a	b	1/8	1/8	c	1/8	d

Find the values of a, b, c and d.

(3)

Find the values of

(d) E(T),

(2)

(e) Var(T).

(4)

Question 7 continued

Question 7 continued

Q7

(Total 14 marks)

TOTAL FOR PAPER: 75 MARKS

END

mock papers 2

1. A disease is known to be present in 2% of a population. A test is developed to help determine whether or not someone has the disease.

Given that a person has the disease, the test is positive with probability 0.95

Given that a person does not have the disease, the test is positive with probability 0.03

(a) Draw a tree diagram to represent this information.

(3)

A person is selected at random from the population and tested for this disease.

(b) Find the probability that the test is positive.

(3)

A doctor randomly selects a person from the population and tests him for the disease. Given that the test is positive,

(c) find the probability that he does not have the disease.

(2)

(d) Comment on the usefulness of this test.

(1)

2. The age in years of the residents of two hotels are shown in the back to back stem and leaf diagram below.

Abbey Hotel $8|5|0$ means 58 years in Abbey hotel and 50 years in Balmoral hotel Balmoral Hotel

(1)	2	0		
(4)	9751	1		
(4)	9831	2	6	(1)
(11)	99997665332	3	447	(3)
(6)	987750	4	005569	(6)
(1)	8	5	000013667	(9)
		6	233457	(6)
		7	015	(3)

For the Balmoral Hotel,

(a) write down the mode of the age of the residents,

(1)

(b) find the values of the lower quartile, the median and the upper quartile.

(3)

(c) (i) Find the mean, $\bar{x}$, of the age of the residents.

(ii) Given that $\sum x^2 = 81\,213$ find the standard deviation of the age of the residents.

(4)

One measure of skewness is found using

$$\frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

(d) Evaluate this measure for the Balmoral Hotel.

(2)

For the Abbey Hotel, the mode is 39, the mean is 33.2, the standard deviation is 12.7 and the measure of skewness is – 0.454

(e) Compare the two age distributions of the residents of each hotel.

(3)

3. The random variable X has probability distribution given in the table below.

x	–1	0	1	2	3
P($X=x$)	p	q	0.2	0.15	0.15

Given that E(X) = 0.55, find

(a) the value of p and the value of q,

(5)

(b) Var(X),

(4)

(c) E($2X - 4$).

(2)

4. Crickets make a noise. The pitch, v kHz, of the noise made by a cricket was recorded at 15 different temperatures, t °C. These data are summarised below.

$$\sum t^2 = 10\,922.81, \sum v^2 = 42.3356, \sum tv = 677.971, \sum t = 401.3, \sum v = 25.08$$

(a) Find S_{tt}, S_{vv} and S_{tv} for these data.

(4)

(b) Find the product moment correlation coefficient between t and v.

(3)

(c) State, with a reason, which variable is the explanatory variable.

(2)

(d) Give a reason to support fitting a regression model of the form $v = a + bt$ to these data.

(1)

(e) Find the value of a and the value of b. Give your answers to 3 significant figures.

(4)

(f) Using this model, predict the pitch of the noise at 19 °C.

(1)

5. A person's blood group is determined by whether or not it contains any of 3 substances *A*, *B* and *C*.

A doctor surveyed 300 patients' blood and produced the table below.

Blood contains	No. of Patients
only *C*	100
A and *C* but not *B*	100
only A	30
B and *C* but not *A*	25
only *B*	12
A, *B* and *C*	10
A and *B* but not *C*	3

(a) Draw a Venn diagram to represent this information.

(4)

(b) Find the probability that a randomly chosen patient's blood contains substance *C*.

(2)

Harry is one of the patients. Given that his blood contains substance *A*,

(c) find the probability that his blood contains all 3 substances.

(2)

Patients whose blood contains none of these substances are called universal blood donors.

(d) Find the probability that a randomly chosen patient is a universal blood donor.

(2)

6. The discrete random variable X can take only the values 2, 3 or 4. For these values the cumulative distribution function is defined by

$$\mathrm{F}(x) = \frac{(x+k)^2}{25} \quad \text{for} \quad x = 2, 3, 4$$

where k is a positive integer.

(a) Find k.

(2)

(b) Find the probability distribution of X.

(3)

7. A packing plant fills bags with cement. The weight X kg of a bag of cement can be modelled by a normal distribution with mean 50 kg and standard deviation 2 kg.

(a) Find P(X>53).

(3)

(b) Find the weight that is exceeded by 99% of the bags.

(5)

Three bags are selected at random.

(c) Find the probability that two weigh more than 53 kg and one weighs less than 53 kg.

(4)

Question 7 continued

(Total 12 marks)

Q7

TOTAL FOR PAPER: 75 MARKS

END

mock papers 3

1. A teacher is monitoring the progress of students using a computer based revision course. The improvement in performance, y marks, is recorded for each student along with the time, x hours, that the student spent using the revision course. The results for a random sample of 10 students are recorded below.

x hours	1.0	3.5	4.0	1.5	1.3	0.5	1.8	2.5	2.3	3.0
y marks	5	30	27	10	–3	–5	7	15	–10	20

[You may use $\sum x = 21.4$, $\sum y = 96$, $\sum x^2 = 57.22$, $\sum xy = 313.7$]

(a) Calculate S_{xx} and S_{xy}.

(3)

(b) Find the equation of the least squares regression line of y on x in the form $y = a + bx$.

(4)

(c) Give an interpretation of the gradient of your regression line.

(1)

Rosemary spends 3.3 hours using the revision course.

(d) Predict her improvement in marks.

(2)

Lee spends 8 hours using the revision course claiming that this should give him an improvement in performance of over 60 marks.

(e) Comment on Lee's claim.

(1)

2. A group of office workers were questioned for a health magazine and $\frac{2}{5}$ were found to take regular exercise. When questioned about their eating habits $\frac{2}{3}$ said they always eat breakfast and, of those who always eat breakfast $\frac{9}{25}$ also took regular exercise.

Find the probability that a randomly selected member of the group

(a) always eats breakfast and takes regular exercise,

(2)

(b) does not always eat breakfast and does not take regular exercise.

(4)

(c) Determine, giving your reason, whether or not always eating breakfast and taking regular exercise are statistically independent.

(2)

3. When Rohit plays a game, the number of points he receives is given by the discrete random variable X with the following probability distribution.

x	0	1	2	3
$P(X = x)$	0.4	0.3	0.2	0.1

(a) Find $E(X)$.

(2)

(b) Find $F(1.5)$.

(2)

(c) Show that $Var(X) = 1$

(4)

(d) Find $Var(5 - 3X)$.

(2)

Rohit can win a prize if the total number of points he has scored after 5 games is at least 10. After 3 games he has a total of 6 points.
You may assume that games are independent.

(e) Find the probability that Rohit wins the prize.

(6)

4. In a study of how students use their mobile telephones, the phone usage of a random sample of 11 students was examined for a particular week.

The total length of calls, y minutes, for the 11 students were

17, 23, 35, 36, 51, 53, 54, 55, 60, 77, 110

(a) Find the median and quartiles for these data.

(3)

A value that is greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$ or smaller than $Q_1 - 1.5 \times (Q_3 - Q_1)$ is defined as an outlier.

(b) Show that 110 is the only outlier.

(2)

(c) Using the graph paper on page 15 draw a box plot for these data indicating clearly the position of the outlier.

(3)

The value of 110 is omitted.

(d) Show that S_{yy} for the remaining 10 students is 2966.9

(3)

These 10 students were each asked how many text messages, x, they sent in the same week.

The values of S_{xx} and S_{xy} for these 10 students are $S_{xx} = 3463.6$ and $S_{xy} = -18.3$.

(e) Calculate the product moment correlation coefficient between the number of text messages sent and the total length of calls for these 10 students.

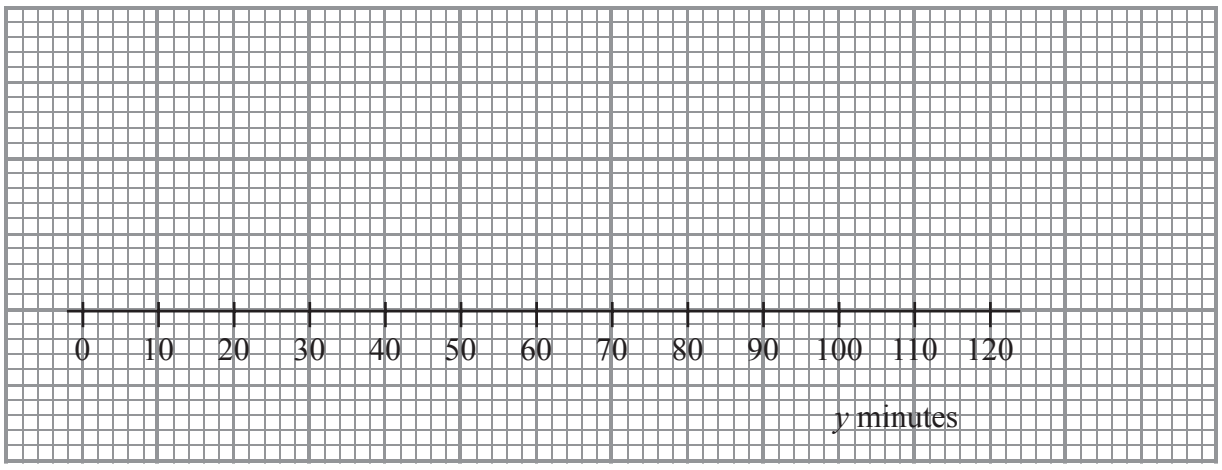
(2)

A parent believes that a student who sends a large number of text messages will spend fewer minutes on calls.

(f) Comment on this belief in the light of your calculation in part (e).

(1)

Question 4 continued



5. In a shopping survey a random sample of 104 teenagers were asked how many hours, to the nearest hour, they spent shopping in the last month. The results are summarised in the table below.

Number of hours	Mid-point	Frequency
0 – 5	2.75	20
6 – 7	6.5	16
8 – 10	9	18
11 – 15	13	25
16 – 25	20.5	15
26 – 50	38	10

A histogram was drawn and the group (8 – 10) hours was represented by a rectangle that was 1.5 cm wide and 3 cm high.

(a) Calculate the width and height of the rectangle representing the group (16 – 25) hours.

(3)

(b) Use linear interpolation to estimate the median and interquartile range.

(5)

(c) Estimate the mean and standard deviation of the number of hours spent shopping.

(4)

(d) State, giving a reason, the skewness of these data.

(2)

(e) State, giving a reason, which average and measure of dispersion you would recommend to use to summarise these data.

(2)

6. The random variable X has a normal distribution with mean 30 and standard deviation 5.

(a) Find P($X < 39$).

(2)

(b) Find the value of d such that P($X < d$) = 0.1151

(4)

(c) Find the value of e such that P($X > e$) = 0.1151

(2)

(d) Find P($d < X < e$).

(2)

Question 6 continued

(Total 10 marks)

Q6

TOTAL FOR PAPER: 75 MARKS

END

mock papers 4

1. The volume of a sample of gas is kept constant. The gas is heated and the pressure, p, is measured at 10 different temperatures, t. The results are summarised below.

$\sum p = 445$ $\sum p^2 = 38\,125$ $\sum t = 240$ $\sum t^2 = 27\,520$ $\sum pt = 26\,830$

(a) Find S_{pp} and S_{pt}.

(3)

Given that $S_{tt} = 21\,760$,

(b) calculate the product moment correlation coefficient.

(2)

(c) Give an interpretation of your answer to part (b).

(1)

2. On a randomly chosen day the probability that Bill travels to school by car, by bicycle or on foot is $\frac{1}{2}$, $\frac{1}{6}$ and $\frac{1}{3}$ respectively. The probability of being late when using these methods of travel is $\frac{1}{5}$, $\frac{2}{5}$ and $\frac{1}{10}$ respectively.

(a) Draw a tree diagram to represent this information.

(3)

(b) Find the probability that on a randomly chosen day

(i) Bill travels by foot and is late,

(ii) Bill is not late.

(4)

(c) Given that Bill is late, find the probability that he did not travel on foot.

(4)

3. The variable x was measured to the nearest whole number. Forty observations are given in the table below.

x	10 – 15	16 – 18	19 –
Frequency	15	9	16

A histogram was drawn and the bar representing the 10 – 15 class has a width of 2 cm and a height of 5 cm. For the 16 – 18 class find

(a) the width,

(1)

(b) the height

(2)

of the bar representing this class.

4. A researcher measured the foot lengths of a random sample of 120 ten-year-old children. The lengths are summarised in the table below.

Foot length, l, (cm)	Number of children
$10 \leqslant l < 12$	5
$12 \leqslant l < 17$	53
$17 \leqslant l < 19$	29
$19 \leqslant l < 21$	15
$21 \leqslant l < 23$	11
$23 \leqslant l < 25$	7

(a) Use interpolation to estimate the median of this distribution.

(2)

(b) Calculate estimates for the mean and the standard deviation of these data.

(6)

One measure of skewness is given by

$$\text{Coefficient of skewness} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

(c) Evaluate this coefficient and comment on the skewness of these data.

(3)

Greg suggests that a normal distribution is a suitable model for the foot lengths of ten-year-old children.

(d) Using the value found in part (c), comment on Greg's suggestion, giving a reason for your answer.

(2)

5. The weight, w grams, and the length, l mm, of 10 randomly selected newborn turtles are given in the table below.

l	49.0	52.0	53.0	54.5	54.1	53.4	50.0	51.6	49.5	51.2
w	29	32	34	39	38	35	30	31	29	30

(You may use $S_{ll} = 33.381$ $S_{wl} = 59.99$ $S_{ww} = 120.1$)

(a) Find the equation of the regression line of w on l in the form $w = a + bl$.

(5)

(b) Use your regression line to estimate the weight of a newborn turtle of length 60 mm.

(2)

(c) Comment on the reliability of your estimate giving a reason for your answer.

(2)

6. The discrete random variable X has probability function

$$P(X = x) = \begin{cases} a(3-x) & x = 0,1,2 \\ b & x = 3 \end{cases}$$

(a) Find $P(X = 2)$ and complete the table below.

x	0	1	2	3
$P(X = x)$	$3a$	$2a$		b

(1)

Given that $E(X) = 1.6$

(b) Find the value of a and the value of b.

(5)

Find

(c) $P(0.5 < X < 3)$,

(2)

(d) $E(3X - 2)$.

(2)

(e) Show that the $Var(X) = 1.64$

(3)

(f) Calculate $Var(3X - 2)$.

(2)

7. (a) Given that P(A) = a and P(B) = b express P($A \cup B$) in terms of a and b when

(i) A and B are mutually exclusive,
(ii) A and B are independent.

(2)

Two events R and Q are such that

P($R \cap Q'$) = 0.15, P(Q) = 0.35 and P($R|Q$) = 0.1

Find the value of

(b) P($R \cup Q$),

(1)

(c) P($R \cap Q$),

(2)

(d) P(R).

(2)

8. The lifetimes of bulbs used in a lamp are normally distributed.
A company *X* sells bulbs with a mean lifetime of 850 hours and a standard deviation of 50 hours.

(a) Find the probability of a bulb, from company *X*, having a lifetime of less than 830 hours.

(3)

(b) In a box of 500 bulbs, from company *X*, find the expected number having a lifetime of less than 830 hours.

(2)

A rival company *Y* sells bulbs with a mean lifetime of 860 hours and 20% of these bulbs have a lifetime of less than 818 hours.

(c) Find the standard deviation of the lifetimes of bulbs from company *Y*.

(4)

Both companies sell the bulbs for the same price.

(d) State which company you would recommend. Give reasons for your answer.

(2)

Question 8 continued

(Total 11 marks)

Q8

TOTAL FOR PAPER: 75 MARKS

END

mock papers 5

1. A jar contains 2 red, 1 blue and 1 green bead. Two beads are drawn at random from the jar without replacement.

(a) In the space below, draw a tree diagram to illustrate all the possible outcomes and associated probabilities. State your probabilities clearly.

(3)

(b) Find the probability that a blue bead and a green bead are drawn from the jar.

(2)

2. The 19 employees of a company take an aptitude test. The scores out of 40 are illustrated in the stem and leaf diagram below.

	2|6 means a score of 26	
0	7	(1)
1	88	(2)
2	4468	(4)
3	2333459	(7)
4	00000	(5)

Find

(a) the median score,

(1)

(b) the interquartile range.

(3)

The company director decides that any employees whose scores are so low that they are outliers will undergo retraining.

An outlier is an observation whose value is less than the lower quartile minus 1.0 times the interquartile range.

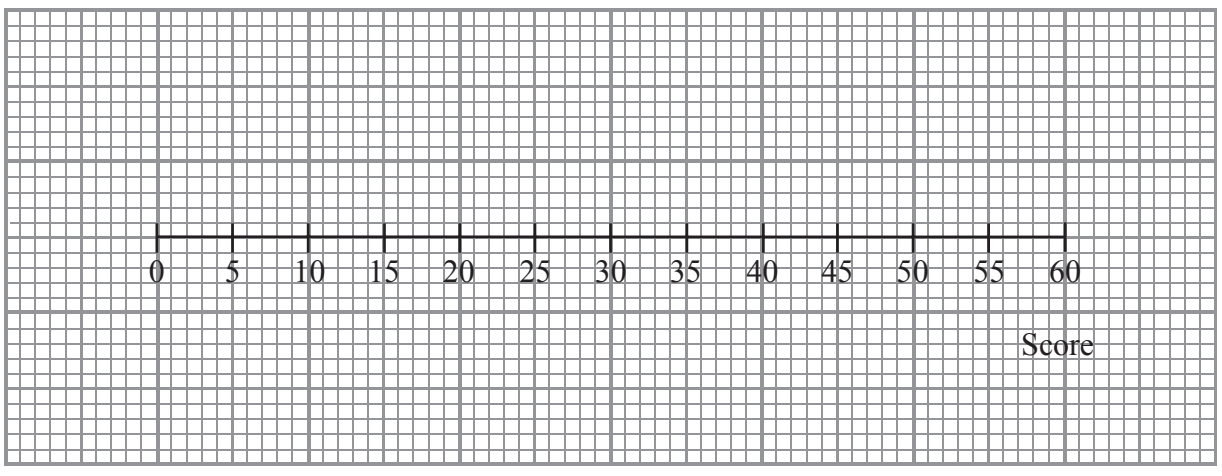
(c) Explain why there is only one employee who will undergo retraining.

(2)

(d) On the graph paper on page 5, draw a box plot to illustrate the employees' scores.

(3)

Question 2 continued



Q2

(Total 9 marks)

Turn over

3. The birth weights, in kg, of 1500 babies are summarised in the table below.

Weight (kg)	Midpoint, $x$kg	Frequency, f
0.0 – 1.0	0.50	1
1.0 – 2.0	1.50	6
2.0 – 2.5	2.25	60
2.5 – 3.0		280
3.0 – 3.5	3.25	820
3.5 – 4.0	3.75	320
4.0 – 5.0	4.50	10
5.0 – 6.0		3

[You may use $\sum fx = 4841$ and $\sum fx^2 = 15\,889.5$]

(a) Write down the missing midpoints in the table above.

(2)

(b) Calculate an estimate of the mean birth weight.

(2)

(c) Calculate an estimate of the standard deviation of the birth weight.

(3)

(d) Use interpolation to estimate the median birth weight.

(2)

(e) Describe the skewness of the distribution. Give a reason for your answer.

(2)

4. There are 180 students at a college following a general course in computing. Students on this course can choose to take up to three extra options.

112 take systems support,
70 take developing software,
81 take networking,
35 take developing software and systems support,
28 take networking and developing software,
40 take systems support and networking,
4 take all three extra options.

(a) In the space below, draw a Venn diagram to represent this information.

(5)

A student from the course is chosen at random.

Find the probability that this student takes

(b) none of the three extra options,

(1)

(c) networking only.

(1)

Students who want to become technicians take systems support and networking. Given that a randomly chosen student wants to become a technician,

(d) find the probability that this student takes all three extra options.

(2)

5. The probability function of a discrete random variable X is given by

$$p(x) = kx^2 \qquad x = 1, 2, 3$$

where k is a positive constant.

(a) Show that $k = \frac{1}{14}$

(2)

Find

(b) $P(X \geqslant 2)$

(2)

(c) $E(X)$

(2)

(d) $\text{Var}(1 - X)$

(4)

6. The blood pressures, p mmHg, and the ages, t years, of 7 hospital patients are shown in the table below.

Patient	A	B	C	D	E	F	G
t	42	74	48	35	56	26	60
p	98	130	120	88	182	80	135

$$[\sum t = 341, \sum p = 833, \sum t^2 = 18\,181, \sum p^2 = 106\,397, \sum tp = 42\,948]$$

(a) Find S_{pp}, S_{tp} and S_{tt} for these data.

(4)

(b) Calculate the product moment correlation coefficient for these data.

(3)

(c) Interpret the correlation coefficient.

(1)

(d) On the graph paper on page 17, draw the scatter diagram of blood pressure against age for these 7 patients.

(2)

(e) Find the equation of the regression line of p on t.

(4)

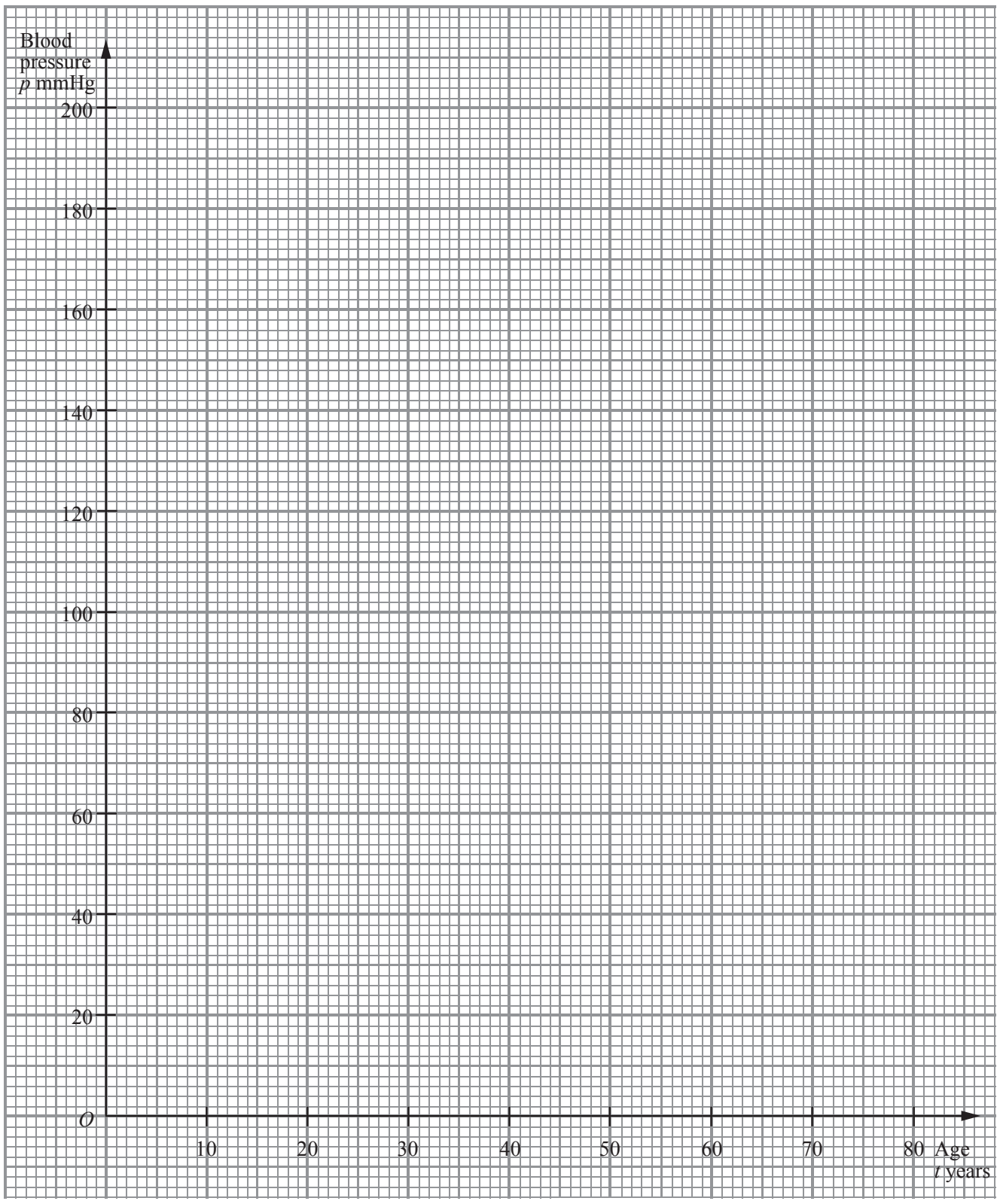
(f) Plot your regression line on your scatter diagram.

(2)

(g) Use your regression line to estimate the blood pressure of a 40 year old patient.

(2)

Question 6 continued



7. The heights of a population of women are normally distributed with mean μ cm and standard deviation σ cm. It is known that 30% of the women are taller than 172 cm and 5% are shorter than 154 cm.

(a) Sketch a diagram to show the distribution of heights represented by this information.

(3)

(b) Show that $\mu = 154 + 1.6449\sigma$.

(3)

(c) Obtain a second equation and hence find the value of μ and the value of σ.

(4)

A woman is chosen at random from the population.

(d) Find the probability that she is taller than 160 cm.

(3)

Question 7 continued

Q7

(Total 13 marks)

TOTAL FOR PAPER: 75 MARKS

END

mock papers 6

1. Gary compared the total attendance, x, at home matches and the total number of goals, y, scored at home during a season for each of 12 football teams playing in a league. He correctly calculated:

$$S_{xx} = 1022500 \quad S_{yy} = 130.9 \quad S_{xy} = 8825$$

(a) Calculate the product moment correlation coefficient for these data.

(2)

(b) Interpret the value of the correlation coefficient.

(1)

Helen was given the same data to analyse. In view of the large numbers involved she decided to divide the attendance figures by 100. She then calculated the product moment correlation coefficient between $\frac{x}{100}$ and y.

(c) Write down the value Helen should have obtained.

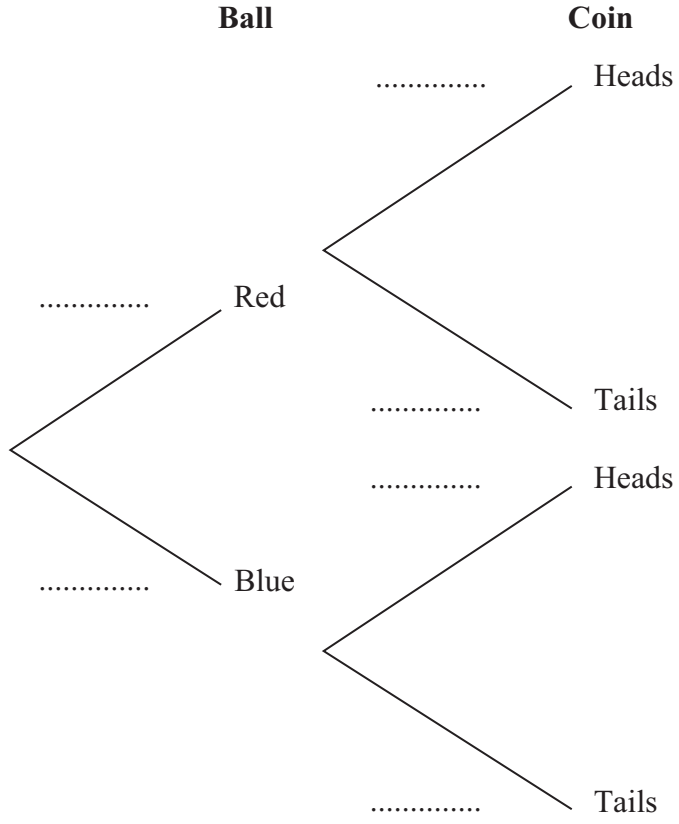
(1)

2. An experiment consists of selecting a ball from a bag and spinning a coin. The bag contains 5 red balls and 7 blue balls. A ball is selected at random from the bag, its colour is noted and then the ball is returned to the bag.

When a red ball is selected, a biased coin with probability $\frac{2}{3}$ of landing heads is spun.

When a blue ball is selected a fair coin is spun.

(a) Complete the tree diagram below to show the possible outcomes and associated probabilities.



(2)

Shivani selects a ball and spins the appropriate coin.

(b) Find the probability that she obtains a head.

(2)

Given that Tom selected a ball at random and obtained a head when he spun the appropriate coin,

(c) find the probability that Tom selected a red ball.

(3)

Shivani and Tom each repeat this experiment.

(d) Find the probability that the colour of the ball Shivani selects is the same as the colour of the ball Tom selects.

(3)

3. The discrete random variable X has probability distribution given by

x	-1	0	1	2	3
$\mathrm{P}(X = x)$	$\frac{1}{5}$	a	$\frac{1}{10}$	a	$\frac{1}{5}$

where a is a constant.

(a) Find the value of a.

(2)

(b) Write down $\mathrm{E}(X)$.

(1)

(c) Find $\mathrm{Var}(X)$.

(3)

The random variable $Y = 6 - 2X$

(d) Find $\mathrm{Var}(Y)$.

(2)

(e) Calculate $\mathrm{P}(X \geqslant Y)$.

(3)

4. The Venn diagram in Figure 1 shows the number of students in a class who read any of 3 popular magazines A, B and C.

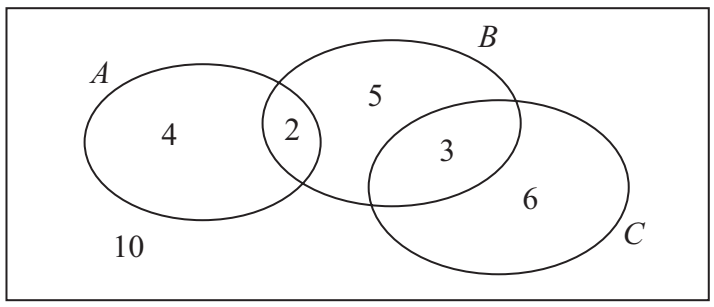


Figure 1

One of these students is selected at random.

(a) Show that the probability that the student reads more than one magazine is $\frac{1}{6}$.

(2)

(b) Find the probability that the student reads A or B (or both).

(2)

(c) Write down the probability that the student reads both A and C.

(1)

Given that the student reads at least one of the magazines,

(d) find the probability that the student reads C.

(2)

(e) Determine whether or not reading magazine B and reading magazine C are statistically independent.

(3)

5. A teacher selects a random sample of 56 students and records, to the nearest hour, the time spent watching television in a particular week.

Hours	1 – 10	11 – 20	21 – 25	26 – 30	31 – 40	41 – 59
Frequency	6	15	11	13	8	3
Mid-point	5.5	15.5		28		50

(a) Find the mid-points of the $21 - 25$ hour and $31 - 40$ hour groups.

(2)

A histogram was drawn to represent these data. The $11 - 20$ group was represented by a bar of width 4 cm and height 6 cm.

(b) Find the width and height of the $26 - 30$ group.

(3)

(c) Estimate the mean and standard deviation of the time spent watching television by these students.

(5)

(d) Use linear interpolation to estimate the median length of time spent watching television by these students.

(2)

The teacher estimated the lower quartile and the upper quartile of the time spent watching television to be 15.8 and 29.3 respectively.

(e) State, giving a reason, the skewness of these data.

(2)

6. A travel agent sells flights to different destinations from *Beerow* airport. The distance d, measured in 100 km, of the destination from the airport and the fare £f are recorded for a random sample of 6 destinations.

Destination	*A*	*B*	*C*	*D*	*E*	*F*
d	2.2	4.0	6.0	2.5	8.0	5.0
f	18	20	25	23	32	28

[You may use $\sum d^2 = 152.09$ $\sum f^2 = 3686$ $\sum fd = 723.1$]

(a) Using the axes below, complete a scatter diagram to illustrate this information.

(2)

(b) Explain why a linear regression model may be appropriate to describe the relationship between f and d.

(1)

(c) Calculate S_{dd} and S_{fd}

(4)

(d) Calculate the equation of the regression line of f on d giving your answer in the form $f = a + bd$.

(4)

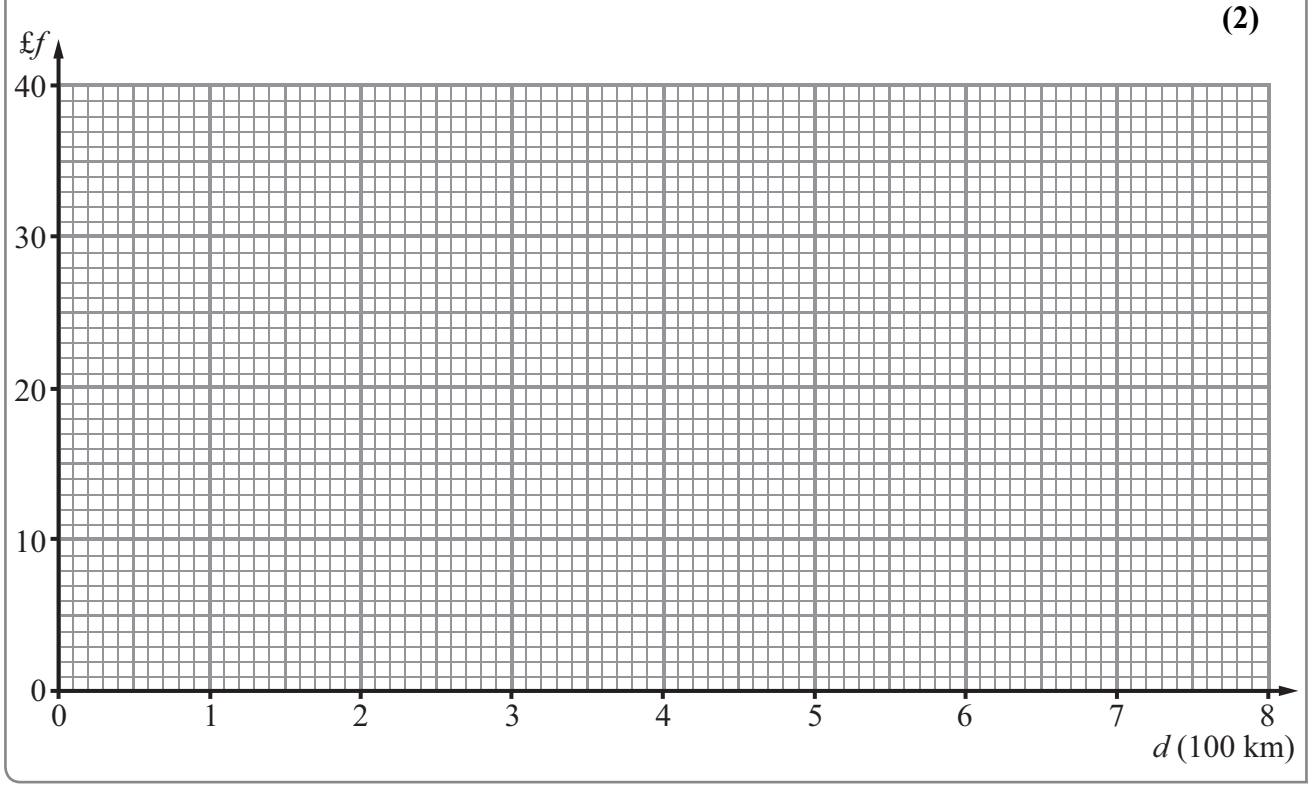
(e) Give an interpretation of the value of b.

(1)

Jane is planning her holiday and wishes to fly from *Beerow* airport to a destination t km away. A rival travel agent charges 5p per km.

(f) Find the range of values of t for which the first travel agent is cheaper than the rival.

(2)



7. The distances travelled to work, D km, by the employees at a large company are normally distributed with $D \sim \mathrm{N}(30, 8^2)$.

(a) Find the probability that a randomly selected employee has a journey to work of more than 20 km.

(3)

(b) Find the upper quartile, Q_3, of D.

(3)

(c) Write down the lower quartile, Q_1, of D.

(1)

An outlier is defined as any value of D such that $D < h$ or $D > k$ where

$$h = Q_1 - 1.5 \times (Q_3 - Q_1) \quad \text{and} \quad k = Q_3 + 1.5 \times (Q_3 - Q_1)$$

(d) Find the value of h and the value of k.

(2)

An employee is selected at random.

(e) Find the probability that the distance travelled to work by this employee is an outlier.

(3)

Question 7 continued

Q7

(Total 12 marks)

TOTAL FOR PAPER: 75 MARKS

END

1. A random sample of 50 salmon was caught by a scientist. He recorded the length l cm and weight w kg of each salmon.

The following summary statistics were calculated from these data.

$\sum l = 4027$ $\sum l^2 = 327\,754.5$ $\sum w = 357.1$ $\sum lw = 29\,330.5$ $S_{ww} = 289.6$

(a) Find S_{ll} and S_{lw}

(3)

(b) Calculate, to 3 significant figures, the product moment correlation coefficient between l and w.

(2)

(c) Give an interpretation of your coefficient.

(1)

2. Keith records the amount of rainfall, in mm, at his school, each day for a week. The results are given below.

2.8 5.6 2.3 9.4 0.0 0.5 1.8

Jenny then records the amount of rainfall, x mm, at the school each day for the following 21 days. The results for the 21 days are summarised below.

$$\sum x = 84.6$$

(a) Calculate the mean amount of rainfall during the whole 28 days.

(2)

Keith realises that he has transposed two of his figures. The number 9.4 should have been 4.9 and the number 0.5 should have been 5.0
Keith corrects these figures.

(b) State, giving your reason, the effect this will have on the mean.

(2)

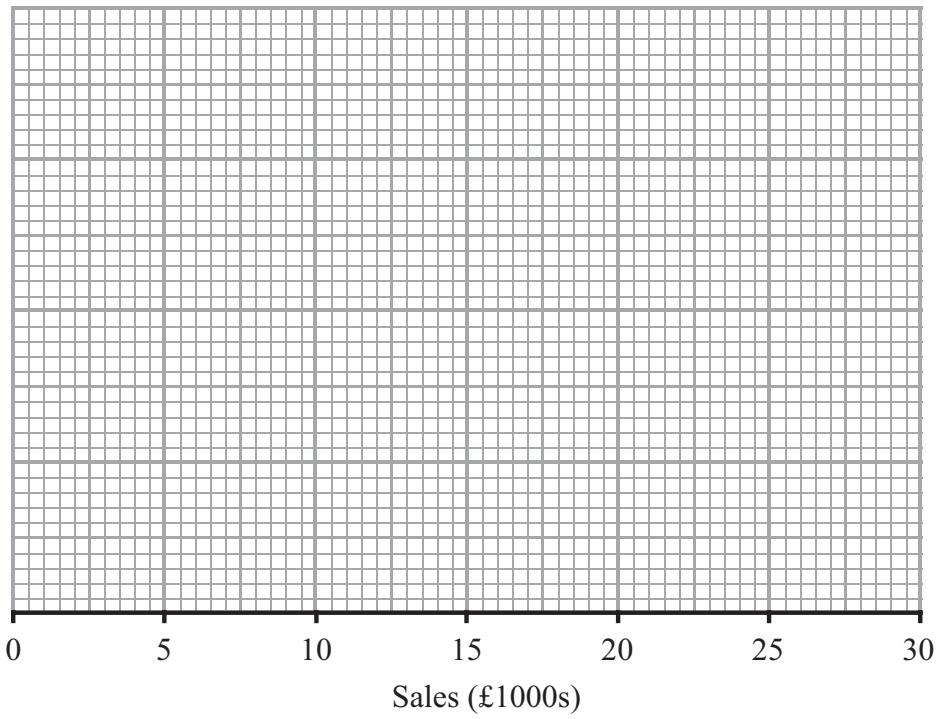
3. Over a long period of time a small company recorded the amount it received in sales per month. The results are summarised below.

	Amount received in sales (£1000s)
Two lowest values	3, 4
Lower quartile	7
Median	12
Upper quartile	14
Two highest values	20, 25

An outlier is an observation that falls
either 1.5 × interquartile range above the upper quartile
or 1.5 × interquartile range below the lower quartile.

(a) On the graph paper below, draw a box plot to represent these data, indicating clearly any outliers.

(5)



(b) State the skewness of the distribution of the amount of sales received. Justify your answer.

(2)

(c) The company claims that for 75 % of the months, the amount received per month is greater than £10 000. Comment on this claim, giving a reason for your answer.

(2)

4. A farmer collected data on the annual rainfall, x cm, and the annual yield of peas, p tonnes per acre.

The data for annual rainfall was coded using $v = \frac{x-5}{10}$ and the following statistics were found.

$$S_{vv} = 5.753 \qquad S_{pv} = 1.688 \qquad S_{pp} = 1.168 \qquad \overline{p} = 3.22 \qquad \overline{v} = 4.42$$

(a) Find the equation of the regression line of p on v in the form $p = a + bv$.

(4)

(b) Using your regression line estimate the annual yield of peas per acre when the annual rainfall is 85 cm.

(2)

5. On a randomly chosen day, each of the 32 students in a class recorded the time, t minutes to the nearest minute, they spent on their homework. The data for the class is summarised in the following table.

Time, t	Number of students
10 – 19	2
20 – 29	4
30 – 39	8
40 – 49	11
50 – 69	5
70 – 79	2

(a) Use interpolation to estimate the value of the median.

(2)

Given that

$$\sum t = 1414 \qquad \text{and} \qquad \sum t^2 = 69\,378$$

(b) find the mean and the standard deviation of the times spent by the students on their homework.

(3)

(c) Comment on the skewness of the distribution of the times spent by the students on their homework. Give a reason for your answer.

(2)

6. The discrete random variable X has the probability distribution

x	1	2	3	4
$P(X = x)$	k	$2k$	$3k$	$4k$

(a) Show that $k = 0.1$

(1)

Find

(b) $E(X)$

(2)

(c) $E(X^2)$

(2)

(d) $\text{Var}(2-5X)$

(3)

Two independent observations X_1 and X_2 are made of X.

(e) Show that $P(X_1 + X_2 = 4) = 0.1$

(2)

(f) Complete the probability distribution table for $X_1 + X_2$

(2)

y	2	3	4	5	6	7	8
$P(X_1 + X_2 = y)$	0.01	0.04	0.10		0.25	0.24	

(g) Find $P(1.5 < X_1 + X_2 \leqslant 3.5)$

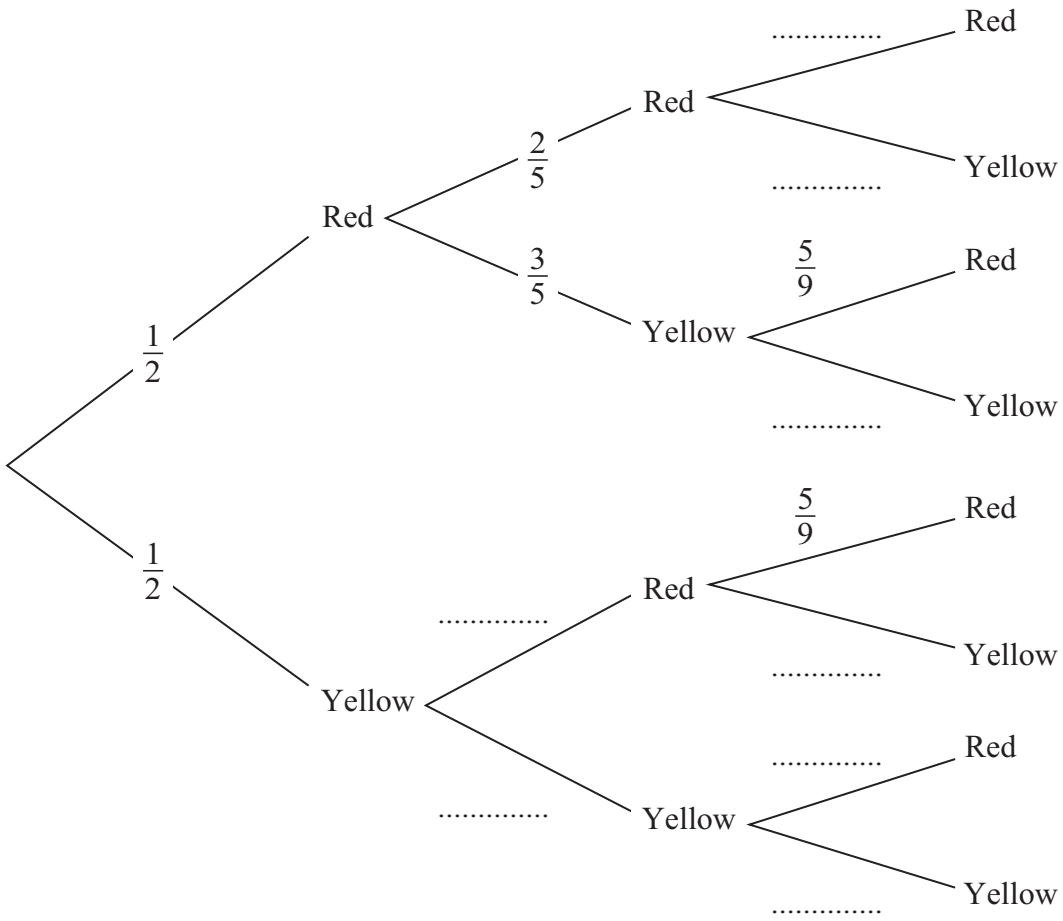
(2)

7. The bag P contains 6 balls of which 3 are red and 3 are yellow.
The bag Q contains 7 balls of which 4 are red and 3 are yellow.
A ball is drawn at random from bag P and placed in bag Q. A second ball is drawn at random from bag P and placed in bag Q.
A third ball is then drawn at random from the 9 balls in bag Q.

The event A occurs when the 2 balls drawn from bag P are of the same colour.
The event B occurs when the ball drawn from bag Q is red.

(a) Complete the tree diagram shown below.

(4)



(b) Find P(A)

(3)

(c) Show that $P(B) = \frac{5}{9}$

(3)

(d) Show that $P(A \cap B) = \frac{2}{9}$

(2)

(e) Hence find $P(A \cup B)$

(2)

(f) Given that all three balls drawn are the same colour, find the probability that they are all red.

(3)

8. The weight, X grams, of soup put in a tin by machine A is normally distributed with a mean of 160 g and a standard deviation of 5 g.
A tin is selected at random.

(a) Find the probability that this tin contains more than 168 g.

(3)

The weight stated on the tin is w grams.

(b) Find w such that $P(X < w) = 0.01$

(3)

The weight, Y grams, of soup put into a carton by machine B is normally distributed with mean μ grams and standard deviation σ grams.

(c) Given that $P(Y < 160) = 0.99$ and $P(Y > 152) = 0.90$ find the value of μ and the value of σ.

(6)

Question 8 continued

Q8

(Total 12 marks)

TOTAL FOR PAPER: 75 MARKS

END