OCR Maths S1

Past Paper Pack

2005-2014

1 **(i)** Calculate the value of Spearman's rank correlation coefficient between the two sets of rankings, A and B, shown in Table 1. [4]

A	1	2	3	4	5
B	4	1	3	2	5

Table 1

(ii) The value of Spearman's rank correlation coefficient between the set of rankings B and a third set of rankings, C, is known to be -1. Copy and complete Table 2 showing the set of rankings C. [2]

B	4	1	3	2	5
C					

Table 2

2 The probability that a certain sample of radioactive material emits an alpha-particle in one unit of time is 0.14. In one unit of time no more than one alpha-particle can be emitted. The number of units of time up to and including the first in which an alpha-particle is emitted is denoted by T.

(i) Find the value of

(a) $\mathrm{P}(T = 5)$, [3]

(b) $\mathrm{P}(T < 8)$. [3]

(ii) State the value of $\mathrm{E}(T)$. [2]

3 In a supermarket the proportion of shoppers who buy washing powder is denoted by p. 16 shoppers are selected at random.

(i) Given that $p = 0.35$, use tables to find the probability that the number of shoppers who buy washing powder is

(a) at least 8, [3]

(b) between 4 and 9 inclusive. [2]

(ii) Given instead that $p = 0.38$, find the probability that the number of shoppers who buy washing powder is exactly 6. [3]

June 2005

4 The table shows the latitude, x (in degrees correct to 3 significant figures), and the average rainfall y (in cm correct to 3 significant figures) of five European cities.

City	x	y
Berlin	52.5	58.2
Bucharest	44.4	58.7
Moscow	55.8	53.3
St Petersburg	60.0	47.8
Warsaw	52.3	56.6

$[n = 5,\ \Sigma x = 265.0,\ \Sigma y = 274.6,\ \Sigma x^2 = 14\,176.54,\ \Sigma y^2 = 15\,162.22,\ \Sigma xy = 14\,464.10.]$

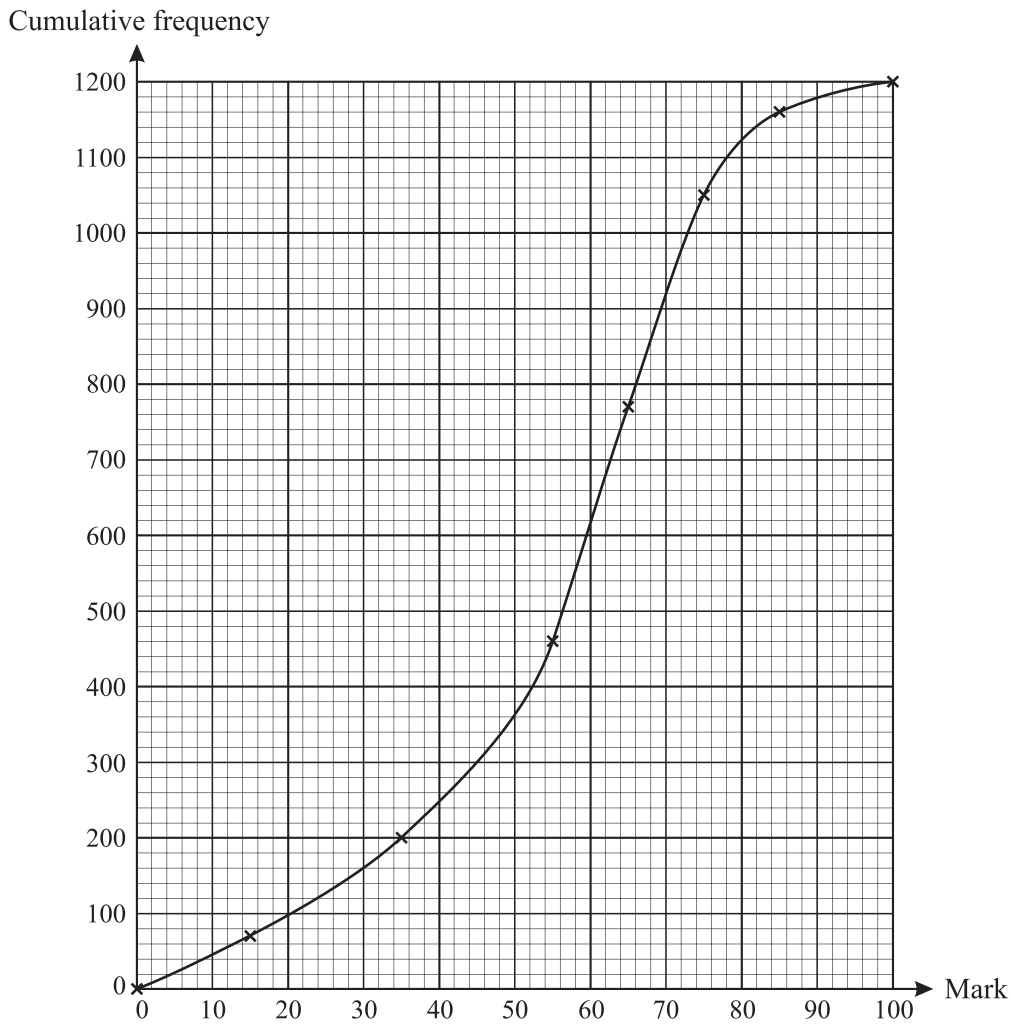
(i) Calculate the product moment correlation coefficient. [3]

(ii) The values of y in the table were in fact obtained from measurements in inches and converted into centimetres by multiplying by 2.54. State what effect it would have had on the value of the product moment correlation coefficient if it had been calculated using inches instead of centimetres. [1]

(iii) It is required to estimate the annual rainfall at Bergen, where $x = 60.4$. Calculate the equation of an appropriate line of regression, giving your answer in simplified form, and use it to find the required estimate. [5]

June 2005

5 The examination marks obtained by 1200 candidates are illustrated on the cumulative frequency graph, where the data points are joined by a smooth curve.



Use the curve to estimate

(i) the interquartile range of the marks, [3]

(ii) x, if 40% of the candidates scored more than x marks, [3]

(iii) the number of candidates who scored more than 68 marks. [2]

Five of the candidates are selected at random, with replacement.

(iv) Estimate the probability that all five scored more than 68 marks. [3]

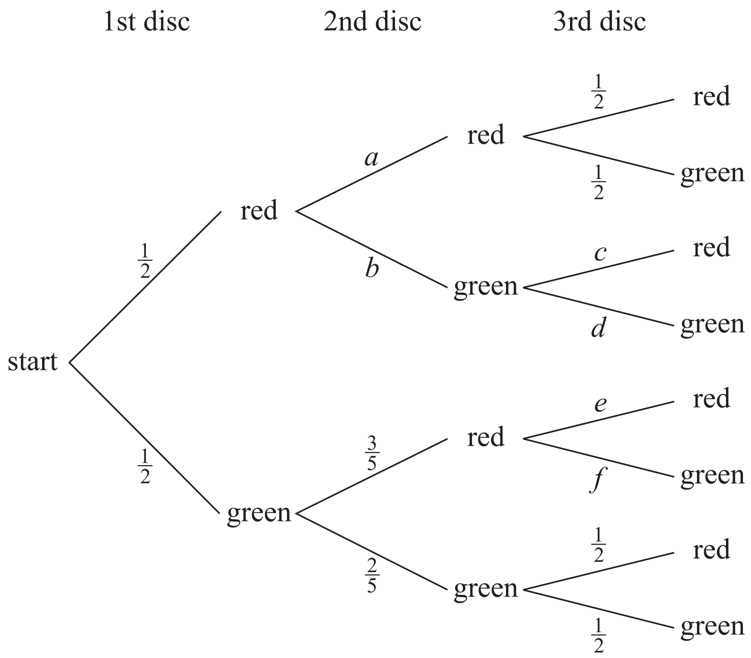
It is subsequently discovered that the candidates' marks in the range 35 to 55 were evenly distributed — that is, roughly equal numbers of candidates scored 35, 36, 37, …, 55.

(v) What does this information suggest about the estimate of the interquartile range found in part **(i)**? [2]

June 2005

6 Two bags contain coloured discs. At first, bag P contains 2 red discs and 2 green discs, and bag Q contains 3 red discs and 1 green disc. A disc is chosen at random from bag P, its colour is noted and it is placed in bag Q. A disc is then chosen at random from bag Q, its colour is noted and it is placed in bag P. A disc is then chosen at random from bag P.

The tree diagram shows the different combinations of three coloured discs chosen.



(i) Write down the values of a, b, c, d, e and f. [4]

The total number of red discs chosen, out of 3, is denoted by R. The table shows the probability distribution of R.

r	0	1	2	3
$P(R = r)$	$\frac{1}{10}$	k	$\frac{9}{20}$	$\frac{1}{5}$

(ii) Show how to obtain the value $P(R = 2) = \frac{9}{20}$. [3]

(iii) Find the value of k. [2]

(iv) Calculate the mean and variance of R. [5]

7 A committee of 7 people is to be chosen at random from 18 volunteers.

(i) In how many different ways can the committee be chosen? [2]

The 18 volunteers consist of 5 people from Gloucester, 6 from Hereford and 7 from Worcester. The committee is to be chosen randomly. Find the probability that the committee will

(ii) consist of 2 people from Gloucester, 2 people from Hereford and 3 people from Worcester, [4]

(iii) include exactly 5 people from Worcester, [4]

(iv) include at least 2 people from each of the three cities. [4]

1 Jenny and John are each allowed two attempts to pass an examination.

(i) Jenny estimates that her chances of success are as follows.

- The probability that she will pass on her first attempt is $\frac{2}{3}$.
- If she fails on her first attempt, the probability that she will pass on her second attempt is $\frac{3}{4}$.

Calculate the probability that Jenny will pass. [3]

(ii) John estimates that his chances of success are as follows.

- The probability that he will pass on his first attempt is $\frac{2}{3}$.
- Overall, the probability that he will pass is $\frac{5}{6}$.

Calculate the probability that if John fails on his first attempt, he will pass on his second attempt. [3]

2 For each of 50 plants, the height, h cm, was measured and the value of $(h - 100)$ was recorded. The mean and standard deviation of $(h - 100)$ were found to be 24.5 and 4.8 respectively.

(i) Write down the mean and standard deviation of h. [2]

The mean and standard deviation of the heights of another 100 plants were found to be 123.0 cm and 5.1 cm respectively.

(ii) Describe briefly how the heights of the second group of plants compare with the first. [2]

(iii) Calculate the mean height of all 150 plants. [2]

3 In Mr Kendall's cupboard there are 3 tins of baked beans and 2 tins of pineapple. Unfortunately his daughter has removed all the labels for a school project and so the tins are identical in appearance. Mr Kendall wishes to use both tins of pineapple for a fruit salad. He opens tins at random until he has opened the two tins of pineapples.

Let X be the number of tins that Mr Kendall opens.

(i) Show that $P(X = 3) = \frac{1}{5}$. [4]

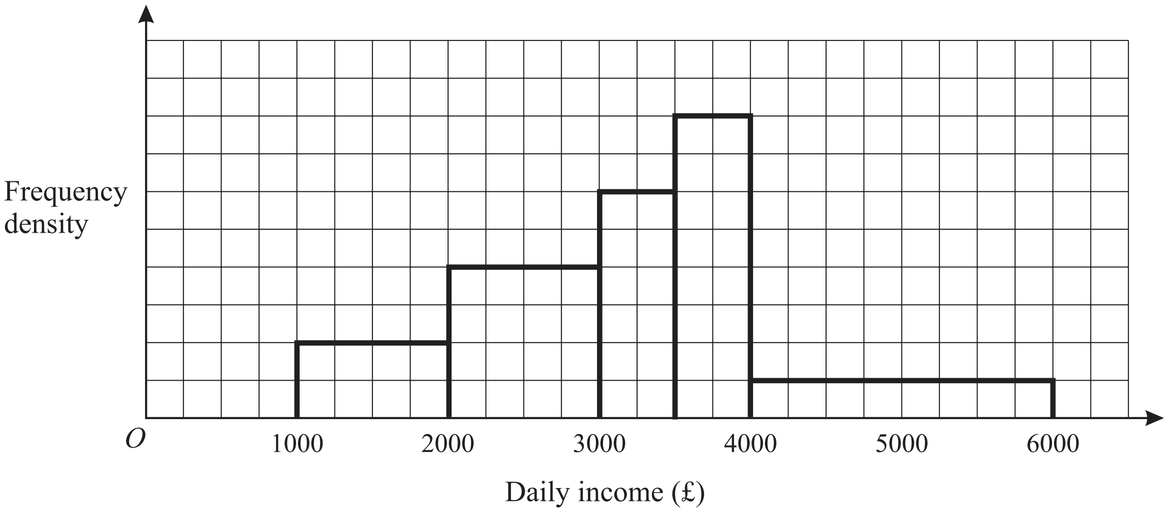
(ii) The probability distribution of X is given in the table below.

x	2	3	4	5
$P(X = x)$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{2}{5}$

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

4 Each day, the Research Department of a retail firm records the firm's daily income, to be used for statistical analysis. The results are summarised by recording the number of days on which the daily income is within certain ranges.

(i)



The histogram shows the results for 300 days. By considering the total area of the histogram,

(a) find the number of days on which the daily income was between £4000 and £6000, [4]

(b) calculate an estimate of the number of days on which the daily income was between £2700 and £3200. [3]

(ii) The Research Department offers to provide any of the following statistical diagrams: histogram, frequency polygon, box-and-whisker plot, cumulative frequency graph, stem-and-leaf diagram and pie chart.

Which one of these statistical diagrams would most easily enable managers to

(a) read off the median and quartile values of the daily income, [1]

(b) find the range of the top 10% of values of the daily income? [1]

5 Andrea practises shots at goal. For each shot the probability of her scoring a goal is $\frac{2}{5}$. Each shot is independent of other shots.

(i) Find the probability that she scores her first goal

(a) on her 5th shot, [2]

(b) before her 5th shot. [3]

(ii) **(a)** Find the probability that she scores exactly 1 goal in her first 5 shots. [3]

(b) Hence find the probability that she scores her **second** goal on her 6th shot. [2]

[Turn over

6 An examination paper consists of two parts. Section A contains questions A1, A2, A3 and A4. Section B contains questions B1, B2, B3, B4, B5, B6 and B7.

Candidates must choose three questions from section A and four questions from section B. The order in which they choose the questions does not matter.

(i) In how many ways can the seven questions be chosen? [3]

(ii) Assuming that all selections are equally likely, find the probability that a particular candidate chooses question A1 but does **not** choose question B1. [3]

(iii) Following a change of syllabus, the form of the examination remains the same except that candidates who choose question A1 are not allowed to choose question B1. In how many ways can the seven questions now be chosen? [3]

7 Past experience has shown that when seeds of a certain type are planted, on average 90% will germinate. A gardener plants 10 of these seeds in a tray and waits to see how many will germinate.

(i) Name an appropriate distribution with which to model the number of seeds that germinate, giving the value(s) of any parameters. State any assumption(s) needed for the model to be valid. [4]

(ii) Use your model to find the probability that fewer than 8 seeds germinate. [2]

Later the gardener plants 20 trays of seeds, with 10 seeds in each tray.

(iii) Calculate the probability that there are at least 19 trays in each of which at least 8 seeds germinate. [4]

8 The table shows the population, x million, of each of nine countries in Western Europe together with the population, y million, of its capital city.

	Germany	United Kingdom	France	Italy	Spain	The Netherlands	Portugal	Austria	Switzerland
x	82.1	59.2	59.1	56.7	39.2	15.9	9.9	8.1	7.3
y	3.5	7.0	9.0	2.7	2.9	0.8	0.7	1.6	0.1

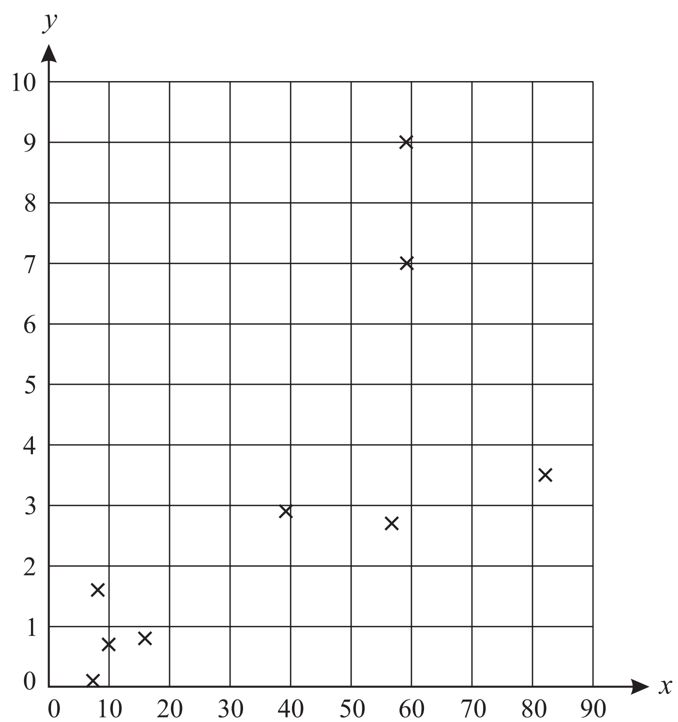
$[n = 9,\ \Sigma x = 337.5,\ \Sigma x^2 = 18\,959.11,\ \Sigma y = 28.3,\ \Sigma y^2 = 161.65,\ \Sigma xy = 1533.76.]$

(i) (a) Calculate Spearman's rank correlation coefficient, r_s. [5]

(b) Explain what your answer indicates about the populations of these countries and their capital cities. [1]

(ii) Calculate the product moment correlation coefficient, r. [2]

The data are illustrated in the scatter diagram.



(iii) By considering the diagram, state the effect on the value of the product moment correlation coefficient, r, if the data for France and the United Kingdom were removed from the calculation. [1]

(iv) In a certain country in Africa, most people live in remote areas and hence the population of the country is unknown. However, the population of the capital city is known to be approximately 1 million. An official suggests that the population of this country could be estimated by using a regression line drawn on the above scatter diagram.

(a) State, with a reason, whether the regression line of y on x or the regression line of x on y would need to be used. [2]

(b) Comment on the reliability of such an estimate in this situation. [2]

1 Some observations of bivariate data were made and the equations of the two regression lines were found to be as follows.

$$y \text{ on } x: \quad y = -0.6x + 13.0$$
$$x \text{ on } y: \quad x = -1.6y + 21.0$$

(i) State, with a reason, whether the correlation between x and y is negative or positive. [1]

(ii) Neither variable is controlled. Calculate an estimate of the value of x when $y = 7.0$. [2]

(iii) Find the values of $\bar{x}$ and $\bar{y}$. [3]

2 A bag contains 5 black discs and 3 red discs. A disc is selected at random from the bag. If it is red it is replaced in the bag. If it is black, it is not replaced. A second disc is now selected at random from the bag.

Find the probability that

(i) the second disc is black, given that the first disc was black, [1]

(ii) the second disc is black, [3]

(iii) the two discs are of different colours. [3]

3 Each of the 7 letters in the word DIVIDED is printed on a separate card. The cards are arranged in a row.

(i) How many different arrangements of the letters are possible? [3]

(ii) In how many of these arrangements are all three Ds together? [2]

The 7 cards are now shuffled and 2 cards are selected at random, without replacement.

(iii) Find the probability that at least one of these 2 cards has D printed on it. [3]

4 **(i)** The random variable X has the distribution $B(25, 0.2)$. Using the tables of cumulative binomial probabilities, or otherwise, find $P(X \geqslant 5)$. [2]

(ii) The random variable Y has the distribution $B(10, 0.27)$. Find $P(Y = 3)$. [2]

(iii) The random variable Z has the distribution $B(n, 0.27)$. Find the smallest value of n such that $P(Z \geqslant 1) > 0.95$. [3]

5 The probability distribution of a discrete random variable, X, is given in the table.

x	0	1	2	3
$P(X = x)$	$\frac{1}{3}$	$\frac{1}{4}$	p	q

It is given that the expectation, $E(X)$, is $1\frac{1}{4}$.

(i) Calculate the values of p and q. [5]

(ii) Calculate the standard deviation of X. [4]

June 2006

6 The table shows the total distance travelled, in thousands of miles, and the amount of commission earned, in thousands of pounds, by each of seven sales agents in 2005.

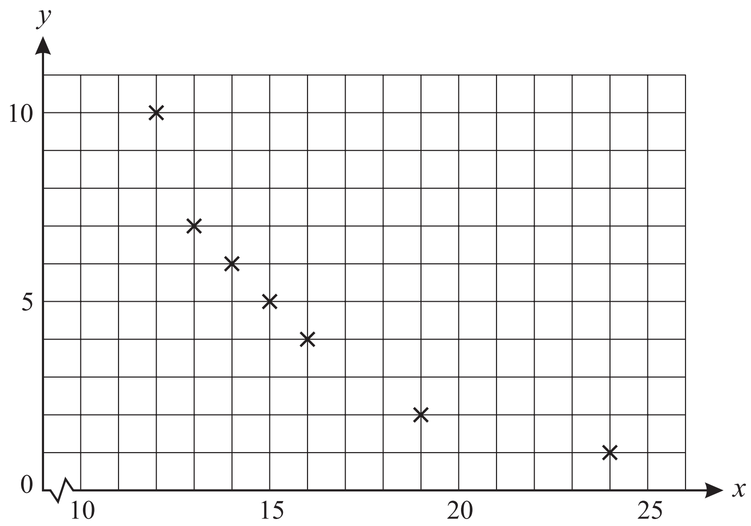
Agent	*A*	*B*	*C*	*D*	*E*	*F*	*G*
Distance travelled	18	15	12	14	16	24	13
Commission earned	18	45	19	24	27	22	23

(i) (a) Calculate Spearman's rank correlation coefficient, r_s, for these data. [5]

(b) Comment briefly on your value of r_s with reference to this context. [1]

(c) After these data were collected, agent *A* found that he had made a mistake. He had actually travelled 19 000 miles in 2005. State, with a reason, but without further calculation, whether the value of Spearman's rank correlation coefficient will increase, decrease or stay the same. [2]

The agents were asked to indicate their level of job satisfaction during 2005. A score of 0 represented no job satisfaction, and a score of 10 represented high job satisfaction. Their scores, *y*, together with the data for distance travelled, *x*, are illustrated in the scatter diagram below.



(ii) For this scatter diagram, what can you say about the value of

(a) Spearman's rank correlation coefficient, [1]

(b) the product moment correlation coefficient? [1]

[Questions 7 and 8 are printed overleaf.]

7 In a UK government survey in 2000, smokers were asked to estimate the time between their waking and their having the first cigarette of the day. For heavy smokers, the results were as follows.

Time between waking and first cigarette	1 to 4 minutes	5 to 14 minutes	15 to 29 minutes	30 to 59 minutes	At least 60 minutes
Percentage of smokers	31	27	19	14	9

Times are given correct to the nearest minute.

(i) Assuming that 'At least 60 minutes' means 'At least 60 minutes but less than 240 minutes', calculate estimates for the mean and standard deviation of the time between waking and first cigarette for these smokers. [6]

(ii) Find an estimate for the interquartile range of the time between waking and first cigarette for these smokers. Give your answer correct to the nearest minute. [4]

(iii) The meaning of 'At least 60 minutes' is now changed to 'At least 60 minutes but less than 480 minutes'. Without further calculation, state whether this would cause an increase, a decrease or no change in the estimated value of

(a) the mean, [1]

(b) the standard deviation, [1]

(c) the interquartile range. [1]

8 Henry makes repeated attempts to light his gas fire. He makes the modelling assumption that the probability that the fire will light on any attempt is $\frac{1}{3}$.

Let X be the number of attempts at lighting the fire, up to and including the successful attempt.

(i) Name the distribution of X, stating a further modelling assumption needed. [2]

In the rest of this question, you should use the distribution named in part **(i)**.

(ii) Calculate

(a) $\mathrm{P}(X = 4)$, [3]

(b) $\mathrm{P}(X < 4)$. [3]

(iii) State the value of $\mathrm{E}(X)$. [1]

(iv) Henry has to light the fire once a day, starting on March 1st. Calculate the probability that the first day on which fewer than 4 attempts are needed to light the fire is March 3rd. [3]

1 Part of the probability distribution of a variable, X, is given in the table.

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

(i) Find $P(X = 0)$. [2]

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

2 The table contains data concerning five households selected at random from a certain town.

Number of people in the household	2	3	3	5	7
Number of cars belonging to people in the household	1	1	3	2	4

(i) Calculate the product moment correlation coefficient, r, for the data in the table. [5]

(ii) Give a reason why it would not be sensible to use your answer to draw a conclusion about all the households in the town. [1]

3 The digits 1, 2, 3, 4 and 5 are arranged in random order, to form a five-digit number.

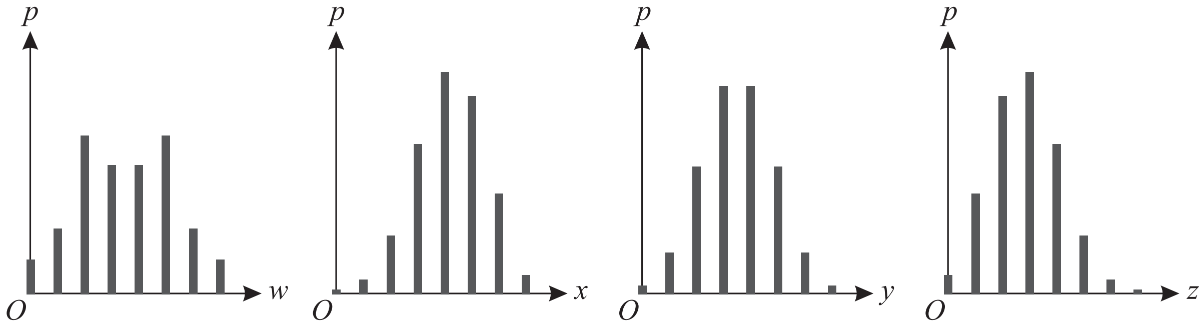
(i) How many different five-digit numbers can be formed? [1]

(ii) Find the probability that the five-digit number is

(a) odd, [2]

(b) less than 23 000. [3]

4 Each of the variables W, X, Y and Z takes eight integer values only. The probability distributions are illustrated in the following diagrams.



(i) For which one or more of these variables is

(a) the mean equal to the median, [1]

(b) the mean greater than the median? [1]

(ii) Give a reason why **none** of these diagrams could represent a geometric distribution. [1]

(iii) Which one of these diagrams could **not** represent a binomial distribution? Explain your answer briefly. [2]

5 A chemical solution was gradually heated. At five-minute intervals the time, x minutes, and the temperature, y °C, were noted.

x	0	5	10	15	20	25	30	35
y	0.8	3.0	6.8	10.9	15.6	19.6	23.4	26.7

$[n = 8,\ \Sigma x = 140,\ \Sigma y = 106.8,\ \Sigma x^2 = 3500,\ \Sigma y^2 = 2062.66,\ \Sigma xy = 2685.0.]$

(i) Calculate the equation of the regression line of y on x. [4]

(ii) Use your equation to estimate the temperature after 12 minutes. [2]

(iii) It is given that the value of the product moment correlation coefficient is close to +1. Comment on the reliability of using your equation to estimate y when

(a) $x = 17$,

(b) $x = 57$.

[2]

6 A coin is biased so that the probability that it will show heads on any throw is $\frac{2}{3}$. The coin is thrown repeatedly.

The number of throws up to and including the first head is denoted by X. Find

(i) $P(X = 4)$, [3]

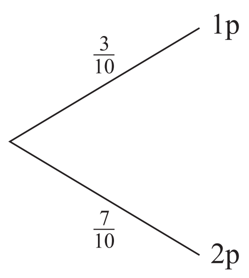
(ii) $P(X < 4)$, [3]

(iii) $E(X)$. [2]

7 A bag contains three 1p coins and seven 2p coins. Coins are removed at random one at a time, **without** replacement, until the total value of the coins removed is **at least** 3p. Then no more coins are removed.

(i) Copy and complete the probability tree diagram. [5]

First coin



Find the probability that

(ii) exactly two coins are removed, [3]

(iii) the total value of the coins removed is 4p. [3]

Jan 2007

8 In the 2001 census, the household size (the number of people living in each household) was recorded. The percentages of households of different sizes were then calculated. The table shows the percentages for two wards, Withington and Old Moat, in Manchester.

	Household size						
	1	2	3	4	5	6	7 or more
Withington	34.1	26.1	12.7	12.8	8.2	4.0	2.1
Old Moat	35.1	27.1	14.7	11.4	7.6	2.8	1.3

(i) Calculate the median and interquartile range of the household size for Withington. [3]

(ii) Making an appropriate assumption for the last class, which should be stated, calculate the mean and standard deviation of the household size for Withington. Give your answers to an appropriate degree of accuracy. [6]

The corresponding results for Old Moat are as follows.

Median	Interquartile range	Mean	Standard deviation
2	2	2.4	1.5

(iii) State one advantage of using the median rather than the mean as a measure of the average household size. [1]

(iv) By comparing the values for Withington with those for Old Moat, explain briefly why the interquartile range may be less suitable than the standard deviation as a measure of the variation in household size. [1]

(v) For one of the above wards, the value of Spearman's rank correlation coefficient between household size and percentage is -1. Without any calculation, state which ward this is. Explain your answer. [2]

9 A variable X has the distribution $\mathrm{B}(11, p)$.

(i) Given that $p = \frac{3}{4}$, find $\mathrm{P}(X = 5)$. [2]

(ii) Given that $\mathrm{P}(X = 0) = 0.05$, find p. [4]

(iii) Given that $\mathrm{Var}(X) = 1.76$, find the two possible values of p. [5]

1 The table shows the probability distribution for a random variable X.

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

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

2 Two judges each placed skaters from five countries in rank order.

Position	1st	2nd	3rd	4th	5th
Judge 1	UK	France	Russia	Poland	Canada
Judge 2	Russia	Canada	France	UK	Poland

Calculate Spearman's rank correlation coefficient, r_s, for the two judges' rankings. [5]

3 **(i)** How many different teams of 7 people can be chosen, without regard to order, from a squad of 15? [2]

(ii) The squad consists of 6 forwards and 9 defenders. How many different teams containing 3 forwards and 4 defenders can be chosen? [2]

4 A bag contains 6 white discs and 4 blue discs. Discs are removed at random, one at a time, **without** replacement.

(i) Find the probability that

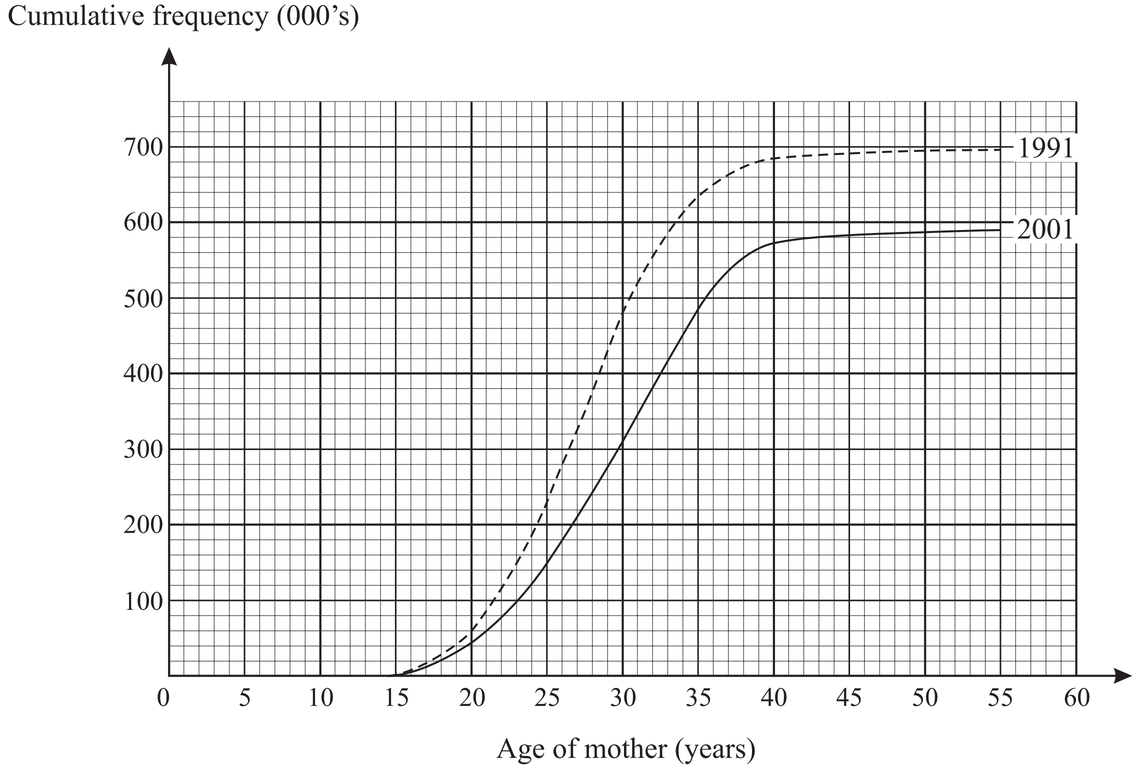
(a) the second disc is blue, given that the first disc was blue, [1]

(b) the second disc is blue, [3]

(c) the third disc is blue, given that the first disc was blue. [3]

(ii) The random variable X is the number of discs which are removed up to and including the first blue disc. State whether the variable X has a geometric distribution. Explain your answer briefly. [1]

5 The numbers of births, in thousands, to mothers of different ages in England and Wales, in 1991 and 2001 are illustrated by the cumulative frequency curves.



(i) In which of these two years were there more births? How many more births were there in this year? [2]

(ii) The following quantities were estimated from the diagram.

Year	Median age (years)	Interquartile range (years)	Proportion of mothers giving birth aged below 25	Proportion of mothers giving birth aged 35 or above
1991	27.5	7.3	33%	9%
2001				18%

(a) Find the values missing from the table. [5]

(b) Did the women who gave birth in 2001 tend to be younger or older or about the same age as the women who gave birth in 1991? Using the table and your values from part **(a)**, give two reasons for your answer. [3]

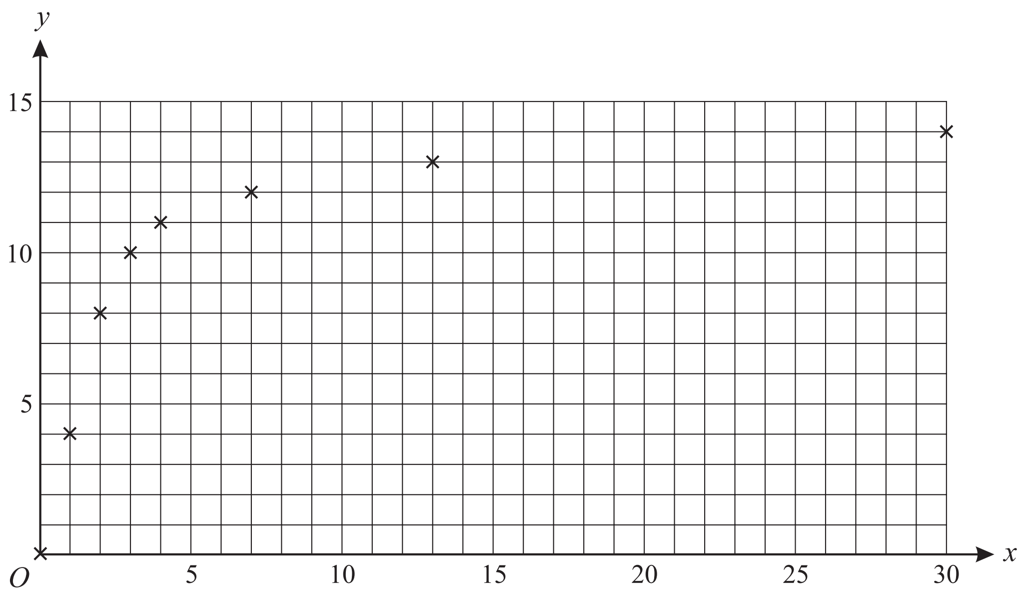
June 2007

6 A machine with artificial intelligence is designed to improve its efficiency rating with practice. The table shows the values of the efficiency rating, y, after the machine has carried out its task various numbers of times, x.

x	0	1	2	3	4	7	13	30
y	0	4	8	10	11	12	13	14

$[n = 8,\ \Sigma x = 60,\ \Sigma y = 72,\ \Sigma x^2 = 1148,\ \Sigma y^2 = 810,\ \Sigma xy = 767.]$

These data are illustrated in the scatter diagram.



(i) (a) Calculate the value of r, the product moment correlation coefficient. [3]

(b) Without calculation, state with a reason the value of r_s, Spearman's rank correlation coefficient. [2]

(ii) A researcher suggests that the data for $x = 0$ and $x = 1$ should be ignored. Without calculation, state with a reason what effect this would have on the value of

(a) r, [2]

(b) r_s. [2]

(iii) Use the diagram to estimate the value of y when $x = 29$. [1]

(iv) Jack finds the equation of the regression line of y on x for all the data, and uses it to estimate the value of y when $x = 29$. Without calculation, state with a reason whether this estimate or the one found in part **(iii)** will be the more reliable. [2]

June 2007

7 On average, 25% of the packets of a certain kind of soup contain a voucher. Kim buys one packet of soup each week for 12 weeks. The number of vouchers she obtains is denoted by X.

(i) State two conditions needed for X to be modelled by the distribution $\mathrm{B}(12, 0.25)$. [2]

In the rest of this question you should assume that these conditions are satisfied.

(ii) Find $\mathrm{P}(X \leqslant 6)$. [2]

In order to claim a free gift, 7 vouchers are needed.

(iii) Find the probability that Kim will be able to claim a free gift at some time during the 12 weeks. [1]

(iv) Find the probability that Kim will be able to claim a free gift in the 12th week but not before. [4]

8 **(i)** A biased coin is thrown twice. The probability that it shows heads both times is 0.04. Find the probability that it shows tails both times. [3]

(ii) Another coin is biased so that the probability that it shows heads on any throw is p. The probability that the coin shows heads exactly once in two throws is 0.42. Find the two possible values of p. [5]

9 **(i)** A random variable X has the distribution $\mathrm{Geo}\left(\frac{1}{5}\right)$. Find

(a) $\mathrm{E}(X)$, [2]

(b) $\mathrm{P}(X = 4)$, [2]

(c) $\mathrm{P}(X > 4)$. [2]

(ii) A random variable Y has the distribution $\mathrm{Geo}(p)$, and $q = 1 - p$.

(a) Show that $\mathrm{P}(Y \text{ is odd}) = p + q^2p + q^4p + \ldots$. [1]

(b) Use the formula for the sum to infinity of a geometric progression to show that

$$\mathrm{P}(Y \text{ is odd}) = \frac{1}{1+q}.$$ [4]

1 **(i)** The letters A, B, C, D and E are arranged in a straight line.

(a) How many different arrangements are possible? [2]

(b) In how many of these arrangements are the letters A and B next to each other? [3]

(ii) From the letters A, B, C, D and E, two different letters are selected at random. Find the probability that these two letters are A and B. [2]

2 A random variable T has the distribution $\text{Geo}\left(\frac{1}{5}\right)$. Find

(i) $P(T = 4)$, [2]

(ii) $P(T > 4)$, [2]

(iii) $E(T)$. [1]

3 A sample of bivariate data was taken and the results were summarised as follows.

$$n = 5 \qquad \Sigma x = 24 \qquad \Sigma x^2 = 130 \qquad \Sigma y = 39 \qquad \Sigma y^2 = 361 \qquad \Sigma xy = 212$$

(i) Show that the value of the product moment correlation coefficient r is 0.855, correct to 3 significant figures. [2]

(ii) The ranks of the data were found. One student calculated Spearman's rank correlation coefficient r_s, and found that $r_s = 0.7$. Another student calculated the product moment coefficient, R, of these ranks. State which one of the following statements is true, and explain your answer briefly.

(A) $R = 0.855$
(B) $R = 0.7$
(C) It is impossible to give the value of R without carrying out a calculation using the original data.

[2]

(iii) All the values of x are now multiplied by a scaling factor of 2. State the new values of r and r_s. [2]

4 A supermarket has a large stock of eggs. 40% of the stock are from a firm called Eggzact. 12% of the stock are brown eggs from Eggzact.

An egg is chosen at random from the stock. Calculate the probability that

(i) this egg is brown, given that it is from Eggzact, [2]

(ii) this egg is from Eggzact and is not brown. [2]

5 **(i)** 20% of people in the large town of Carnley support the Residents' Party. 12 people from Carnley are selected at random. Out of these 12 people, the number who support the Residents' Party is denoted by U.

Find

(a) $P(U \leqslant 5)$, [2]

(b) $P(U \geqslant 3)$. [3]

(ii) 30% of people in Carnley support the Commerce Party. 15 people from Carnley are selected at random. Out of these 15 people, the number who support the Commerce Party is denoted by V.

Find $P(V = 4)$. [3]

6 The probability distribution for a random variable Y is shown in the table.

y	1	2	3
$P(Y = y)$	0.2	0.3	0.5

(i) Calculate $E(Y)$ and $Var(Y)$. [5]

Another random variable, Z, is independent of Y. The probability distribution for Z is shown in the table.

z	1	2	3
$P(Z = z)$	0.1	0.25	0.65

One value of Y and one value of Z are chosen at random. Find the probability that

(ii) $Y + Z = 3$, [3]

(iii) $Y \times Z$ is even. [3]

7 **(i)** Andrew plays 10 tennis matches. In each match he either wins or loses.

(a) State, in this context, two conditions needed for a binomial distribution to arise. [2]

(b) Assuming these conditions are satisfied, define a variable in this context which has a binomial distribution. [1]

(ii) The random variable X has the distribution $B(21, p)$, where $0 < p < 1$.

Given that $P(X = 10) = P(X = 9)$, find the value of p. [5]

8 The stem-and-leaf diagram shows the age in completed years of the members of a sports club.

Male		Female
8 8 7 6	1	6 6 6 7 7 8 8 9
7 6 5 5 3 3 2 1	2	1 3 3 4 5 7 8 8 9 9
9 8 4 4 3	3	2 3 3 4 7
5 2 1	4	0 1 8
9 0	5	0

Key: 1 | 4 | 0 represents a male aged 41 and a female aged 40.

(i) Find the median and interquartile range for the males. [3]

(ii) The median and interquartile range for the females are 27 and 15 respectively. Make two comparisons between the ages of the males and the ages of the females. [2]

(iii) The mean age of the males is 30.7 and the mean age of the females is 27.5, each correct to 1 decimal place. Give one advantage of using the median rather than the mean to compare the ages of the males with the ages of the females. [1]

A record was kept of the number of hours, X, spent by each member at the club in a year. The results were summarised by

$$n = 49, \quad \Sigma(x - 200) = 245, \quad \Sigma(x - 200)^2 = 9849.$$

(iv) Calculate the mean and standard deviation of X. [6]

9 It is thought that the pH value of sand (a measure of the sand's acidity) may affect the extent to which a particular species of plant will grow in that sand. A botanist wished to determine whether there was any correlation between the pH value of the sand on certain sand dunes, and the amount of each of two plant species growing there. She chose random sections of equal area on each of eight sand dunes and measured the pH values. She then measured the area within each section that was covered by each of the two species. The results were as follows.

	Dune	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
	pH value, x	8.5	8.5	9.5	8.5	6.5	7.5	8.5	9.0
Area, $y\,\text{cm}^2$, covered	Species *P*	150	150	575	330	45	15	340	330
	Species *Q*	170	15	80	230	75	25	0	0

The results for species *P* can be summarised by

$$n = 8, \quad \Sigma x = 66.5, \quad \Sigma x^2 = 558.75, \quad \Sigma y = 1935, \quad \Sigma y^2 = 711\,275, \quad \Sigma xy = 17\,082.5.$$

(i) Give a reason why it might be appropriate to calculate the equation of the regression line of y on x rather than x on y in this situation. [1]

(ii) Calculate the equation of the regression line of y on x for species *P*, in the form $y = a + bx$, giving the values of a and b correct to 3 significant figures. [4]

(iii) Estimate the value of y for species *P* on sand where the pH value is 7.0. [2]

The values of the product moment correlation coefficient between x and y for species *P* and *Q* are $r_P = 0.828$ and $r_Q = 0.0302$.

(iv) Describe the relationship between the area covered by species *Q* and the pH value. [1]

(v) State, with a reason, whether the regression line of y on x for species *P* will provide a reliable estimate of the value of y when the pH value is

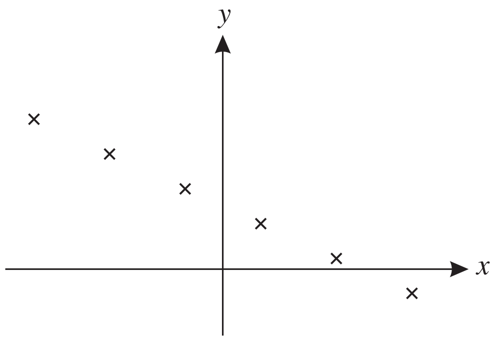
(a) 8, [1]

(b) 4. [1]

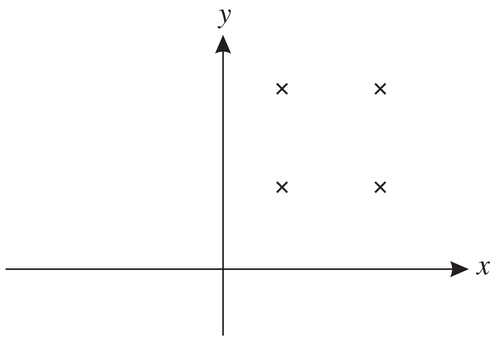
(vi) Assume that the equation of the regression line of y on x for species *Q* is also known. State, with a reason, whether this line will provide a reliable estimate of the value of y when the pH value is 8. [1]

1 **(i)** State the value of the product moment correlation coefficient for each of the following scatter diagrams. [2]

(a)



(b)



(ii) Calculate the value of Spearman's rank correlation coefficient for the following data. [5]

x	3.8	4.1	4.5	5.3
y	1.4	0.8	0.7	1.2

2 A class consists of 7 students from Ashville and 8 from Bewton. A committee of 5 students is chosen at random from the class.

(i) Find the probability that 2 students from Ashville and 3 from Bewton are chosen. [3]

(ii) In fact 2 students from Ashville and 3 from Bewton are chosen. In order to watch a video, all 5 committee members sit in a row. In how many different orders can they sit if no two students from Bewton sit next to each other? [2]

3 **(i)** A random variable X has the distribution $B(8, 0.55)$. Find

(a) $P(X < 7)$, [1]

(b) $P(X = 5)$, [2]

(c) $P(3 \leqslant X < 6)$. [3]

(ii) A random variable Y has the distribution $B\left(10, \frac{5}{12}\right)$. Find

(a) $P(Y = 2)$, [2]

(b) $\text{Var}(Y)$. [1]

4 At a fairground stall, on each turn a player receives prize money with the following probabilities.

Prize money	£0.00	£0.50	£5.00
Probability	$\frac{17}{20}$	$\frac{1}{10}$	$\frac{1}{20}$

(i) Find the probability that a player who has two turns will receive a total of £5.50 in prize money. [3]

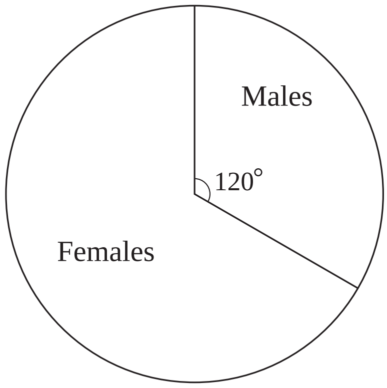
(ii) The stall-holder wishes to make a profit of 20p per turn on average. Calculate the amount the stall-holder should charge for each turn. [4]

5 **(i)** A bag contains 12 red discs and 10 black discs. Two discs are removed at random, without replacement. Find the probability that both discs are red. [2]

(ii) Another bag contains 7 green discs and 8 blue discs. Three discs are removed at random, without replacement. Find the probability that exactly two of these discs are green. [3]

(iii) A third bag contains 45 discs, each of which is either yellow or brown. Two discs are removed at random, without replacement. The probability that both discs are yellow is $\frac{1}{15}$. Find the number of yellow discs which were in the bag at first. [4]

6 **(i)** The numbers of males and females in Year 12 at a school are illustrated in the pie chart. The number of males in Year 12 is 128.

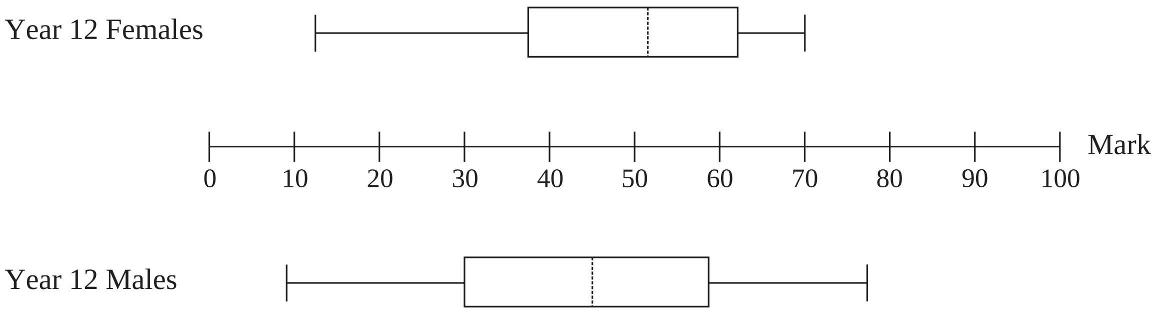


Year 12

(a) Find the number of females in Year 12. [1]

(b) On a corresponding pie chart for Year 13, the angle of the sector representing males is 150°. Explain why this does not necessarily mean that the number of males in Year 13 is more than 128. [1]

(ii) All the Year 12 students took a General Studies examination. The results are illustrated in the box-and-whisker plots.



(a) One student said "The Year 12 pie chart shows that there are more females than males, but the box-and-whisker plots show that there are more males than females."

Comment on this statement. [1]

(b) Give two comparisons between the overall performance of the females and the males in the General Studies examination. [2]

(c) Give one advantage and one disadvantage of using box-and-whisker plots rather than histograms to display the results. [2]

(iii) The mean mark for 102 of the male students was 51. The mean mark for the remaining 26 male students was 59. Calculate the mean mark for all 128 male students. [3]

7 Once each year, Paula enters a lottery for a place in an annual marathon. Each time she enters the lottery, the probability of her obtaining a place is 0.3. Find the probability that

(i) the first time she obtains a place is on her 4th attempt, [3]

(ii) she does not obtain a place on any of her first 6 attempts, [2]

(iii) she needs fewer than 10 attempts to obtain a place, [3]

(iv) she obtains a place exactly twice in her first 5 attempts. [3]

8 A city council attempted to reduce traffic congestion by introducing a congestion charge. The charge was set at £4.00 for the first year and was then increased by £2.00 each year. For each of the first eight years, the council recorded the average number of vehicles entering the city centre per day. The results are shown in the table.

Charge, £x	4	6	8	10	12	14	16	18
Average number of vehicles per day, y million	2.4	2.5	2.2	2.3	2.0	1.8	1.7	1.5

[$n = 8$, $\Sigma x = 88$, $\Sigma y = 16.4$, $\Sigma x^2 = 1136$, $\Sigma y^2 = 34.52$, $\Sigma xy = 168.6$.]

(i) Calculate the product moment correlation coefficient for these data. [3]

(ii) Explain why x is the independent variable. [1]

(iii) Calculate the equation of the regression line of y on x. [4]

(iv) **(a)** Use your equation to estimate the average number of vehicles which will enter the city centre per day when the congestion charge is raised to £20.00. [2]

(b) Comment on the reliability of your estimate. [2]

(v) The council wishes to estimate the congestion charge required to reduce the average number of vehicles entering the city per day to 1.0 million. Assuming that a reliable estimate can be made by extrapolation, state whether they should use the regression line of y on x or the regression line of x on y. Give a reason for your answer. [2]

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]**

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]**

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]**

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]**

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	1 2 3 5 6 8 9
7	0 0 2 4 5 6 7 8
8	0
9	7

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]**

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]**

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]**

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]**

1 20% of packets of a certain kind of cereal contain a free gift. Jane buys one packet a week for 8 weeks. The number of free gifts that Jane receives is denoted by X. Assuming that Jane's 8 packets can be regarded as a random sample, find

(i) $\mathrm{P}(X = 3)$, **[3]**

(ii) $\mathrm{P}(X \geqslant 3)$, **[2]**

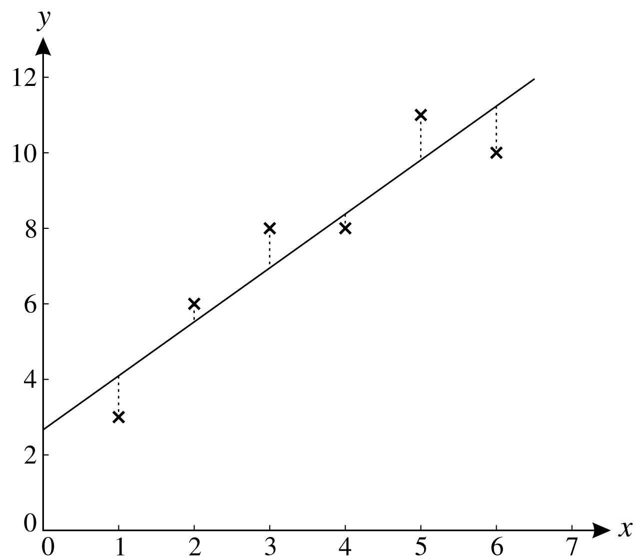
(iii) $\mathrm{E}(X)$. **[2]**

2 Two judges placed 7 dancers in rank order. Both judges placed dancers A and B in the first two places, but in opposite orders. The judges agreed about the ranks for all the other 5 dancers. Calculate the value of Spearman's rank correlation coefficient. **[4]**

3 In an agricultural experiment, the relationship between the amount of water supplied, x units, and the yield, y units, was investigated. Six values of x were chosen and for each value of x the corresponding value of y was measured. The results are shown in the table.

x	1	2	3	4	5	6
y	3	6	8	8	11	10

These results, together with the regression line of y on x, are plotted on the graph.



(i) Give a reason why the regression line of x on y is not suitable in this context. **[1]**

(ii) Explain the significance, for the regression line of y on x, of the distances shown by the vertical dotted lines in the diagram. **[2]**

(iii) Calculate the value of the product moment correlation coefficient, r. **[3]**

(iv) Comment on your value of r in relation to the diagram. **[2]**

June 2009

4 30% of people own a Talk-2 phone. People are selected at random, one at a time, and asked whether they own a Talk-2 phone. The number of people questioned, up to and including the first person who owns a Talk-2 phone, is denoted by X. Find

(i) $\mathrm{P}(X = 4)$, **[3]**

(ii) $\mathrm{P}(X > 4)$, **[2]**

(iii) $\mathrm{P}(X < 6)$. **[3]**

5 The diameters of 100 pebbles were measured. The measurements rounded to the nearest millimetre, x, are summarised in the table.

x	$10 \leqslant x \leqslant 19$	$20 \leqslant x \leqslant 24$	$25 \leqslant x \leqslant 29$	$30 \leqslant x \leqslant 49$
Number of stones	25	22	29	24

These data are to be presented on a statistical diagram.

(i) For a histogram, find the frequency density of the $10 \leqslant x \leqslant 19$ class. **[2]**

(ii) For a cumulative frequency graph, state the coordinates of the first two points that should be plotted. **[2]**

(iii) Why is it not possible to draw an exact box-and-whisker plot to illustrate the data? **[1]**

6 Last year Eleanor played 11 rounds of golf. Her scores were as follows:

79, 71, 80, 67, 67, 74, 66, 65, 71, 66, 64.

(i) Calculate the mean of these scores and show that the standard deviation is 5.31, correct to 3 significant figures. **[4]**

(ii) Find the median and interquartile range of the scores. **[4]**

This year, Eleanor also played 11 rounds of golf. The standard deviation of her scores was 4.23, correct to 3 significant figures, and the interquartile range was the same as last year.

(iii) Give a possible reason why the standard deviation of her scores was lower than last year although her interquartile range was unchanged. **[1]**

In golf, smaller scores mean a better standard of play than larger scores. Ken suggests that since the standard deviation was smaller this year, Eleanor's overall standard has improved.

(iv) Explain why Ken is wrong. **[1]**

(v) State what the smaller standard deviation does show about Eleanor's play. **[1]**

[Questions 7, 8 and 9 are printed overleaf.]

7 Three letters are selected at random from the 8 letters of the word COMPUTER, without regard to order.

(i) Find the number of possible selections of 3 letters. **[2]**

(ii) Find the probability that the letter P is included in the selection. **[3]**

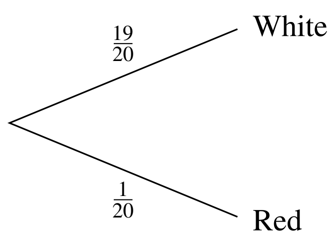
Three letters are now selected at random, one at a time, from the 8 letters of the word COMPUTER, and are placed in order in a line.

(iii) Find the probability that the 3 letters form the word TOP. **[3]**

8 A game at a charity event uses a bag containing 19 white counters and 1 red counter. To play the game once a player takes counters at random from the bag, one at a time, without replacement. If the red counter is taken, the player wins a prize and the game ends. If not, the game ends when 3 white counters have been taken. Niko plays the game once.

(i) (a) Copy and complete the tree diagram showing the probabilities for Niko. **[4]**

First counter



(b) Find the probability that Niko will win a prize. **[3]**

(ii) The number of counters that Niko takes is denoted by X.

(a) Find $P(X = 3)$. **[2]**

(b) Find $E(X)$. **[4]**

9 Repeated independent trials of a certain experiment are carried out. On each trial the probability of success is 0.12.

(i) Find the smallest value of n such that the probability of at least one success in n trials is more than 0.95. **[3]**

(ii) Find the probability that the 3rd success occurs on the 7th trial. **[5]**

Jan 2010

1 Andy makes repeated attempts to thread a needle. The number of attempts up to and including his first success is denoted by X.

(i) State two conditions necessary for X to have a geometric distribution. **[2]**

(ii) Assuming that X has the distribution Geo(0.3), find

(a) $P(X = 5)$, **[2]**

(b) $P(X > 5)$. **[3]**

(iii) Suggest a reason why one of the conditions you have given in part **(i)** might not be satisfied in this context. **[2]**

2 40 people were asked to guess the length of a certain road. Each person gave their guess, l km, correct to the nearest kilometre. The results are summarised below.

l	10–12	13–15	16–20	21–30
Frequency	1	13	20	6

(i) **(a)** Use appropriate formulae to calculate estimates of the mean and standard deviation of l. **[6]**

(b) Explain why your answers are only estimates. **[1]**

(ii) A histogram is to be drawn to illustrate the data. Calculate the frequency density of the block for the 16–20 class. **[2]**

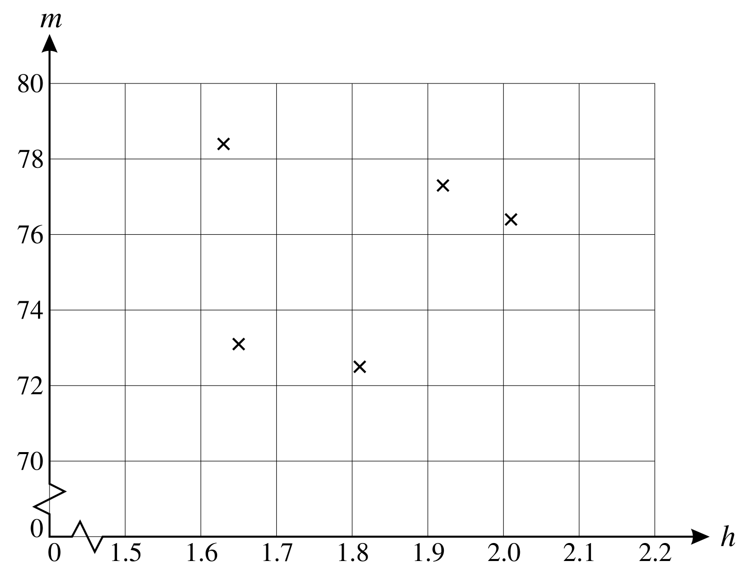
(iii) Explain which class contains the median value of l. **[2]**

(iv) Later, the person whose guess was between 10 km and 12 km changed his guess to between 13 km and 15 km. Without calculation state whether the following will increase, decrease or remain the same:

(a) the mean of l, **[1]**

(b) the standard deviation of l. **[1]**

3 The heights, h m, and weights, m kg, of five men were measured. The results are plotted on the diagram.



The results are summarised as follows.

$$n = 5 \quad \Sigma h = 9.02 \quad \Sigma m = 377.7 \quad \Sigma h^2 = 16.382 \quad \Sigma m^2 = 28\,558.67 \quad \Sigma hm = 681.612$$

(i) Use the summarised data to calculate the value of the product moment correlation coefficient, r. **[3]**

(ii) Comment on your value of r in relation to the diagram. **[2]**

(iii) It was decided to re-calculate the value of r after converting the heights to feet and the masses to pounds. State what effect, if any, this will have on the value of r. **[1]**

(iv) One of the men had height 1.63 m and mass 78.4 kg. The data for this man were removed and the value of r was re-calculated using the original data for the remaining four men. State in general terms what effect, if any, this will have on the value of r. **[1]**

4 A certain four-sided die is biased. The score, X, on each throw is a random variable with probability distribution as shown in the table. Throws of the die are independent.

x	0	1	2	3
$\mathrm{P}(X = x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

(i) Calculate $\mathrm{E}(X)$ and $\mathrm{Var}(X)$. **[5]**

The die is thrown 10 times.

(ii) Find the probability that there are not more than 4 throws on which the score is 1. **[2]**

(iii) Find the probability that there are exactly 4 throws on which the score is 2. **[3]**

5 A washing-up bowl contains 6 spoons, 5 forks and 3 knives. Three of these 14 items are removed at random, without replacement. Find the probability that

(i) all three items are of different kinds, **[3]**

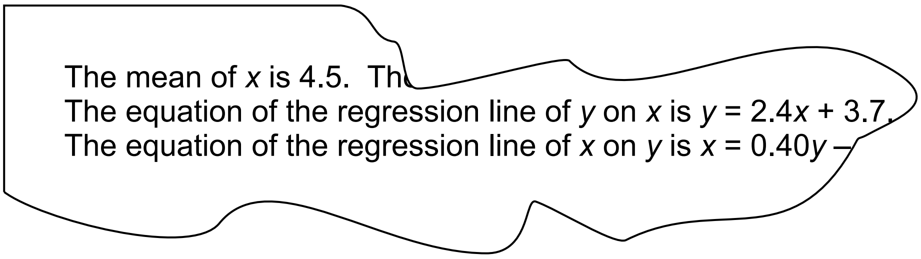
(ii) all three items are of the same kind. **[3]**

6 **(a)** A student calculated the values of the product moment correlation coefficient, r, and Spearman's rank correlation coefficient, r_s, for two sets of bivariate data, A and B. His results are given below.

$$A: \quad r = 0.9 \text{ and } r_s = 1$$
$$B: \quad r = 1 \text{ and } r_s = 0.9$$

With the aid of a diagram where appropriate, explain why the student's results for A could both be correct but his results for B cannot both be correct. **[3]**

(b) An old research paper has been partially destroyed. The surviving part of the paper contains the following incomplete information about some bivariate data from an experiment.



Calculate the missing constant at the end of the equation of the second regression line. **[4]**

7 The table shows the numbers of male and female members of a vintage car club who own either a Jaguar or a Bentley. No member owns both makes of car.

	Male	Female
Jaguar	25	15
Bentley	12	8

One member is chosen at random from these 60 members.

(i) Given that this member is male, find the probability that he owns a Jaguar. **[2]**

Now two members are chosen at random from the 60 members. They are chosen one at a time, without replacement.

(ii) Given that the first one of these members is female, find the probability that both own Jaguars. **[4]**

8 The five letters of the word NEVER are arranged in random order in a straight line.

(i) How many different orders of the letters are possible? **[2]**

(ii) In how many of the possible orders are the two Es next to each other? **[2]**

(iii) Find the probability that the first two letters in the order include exactly one letter E. **[3]**

9 R and S are independent random variables each having the distribution $\text{Geo}(p)$.

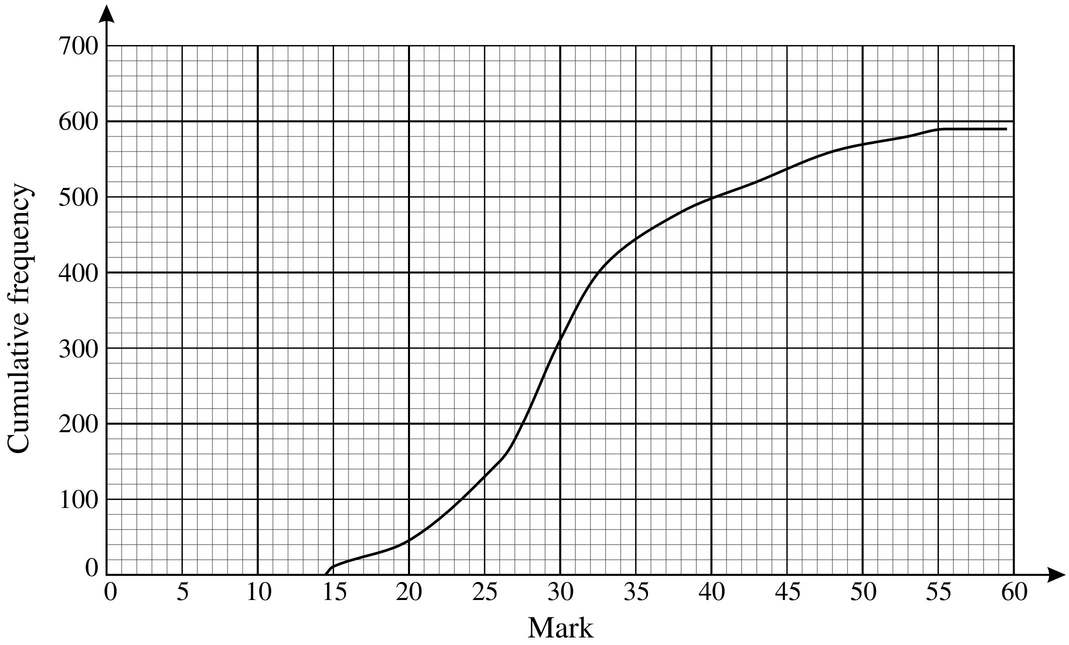
(i) Find $\text{P}(R = 1 \text{ and } S = 1)$ in terms of p. **[1]**

(ii) Show that $\text{P}(R = 3 \text{ and } S = 3) = p^2q^4$, where $q = 1 - p$. **[1]**

(iii) Use the formula for the sum to infinity of a geometric series to show that

$$\text{P}(R = S) = \frac{p}{2-p}.$$ **[5]**

1 The marks of some students in a French examination were summarised in a grouped frequency distribution and a cumulative frequency diagram was drawn, as shown below.



(i) Estimate how many students took the examination. **[1]**

(ii) How can you tell that no student scored more than 55 marks? **[1]**

(iii) Find the greatest possible range of the marks. **[1]**

(iv) The minimum mark for Grade C was 27. The number of students who gained exactly Grade C was the same as the number of students who gained a grade lower than C. Estimate the maximum mark for Grade C. **[3]**

(v) In a German examination the marks of the same students had an interquartile range of 16 marks. What does this result indicate about the performance of the students in the German examination as compared with the French examination? **[3]**

2 Three skaters, A, B and C, are placed in rank order by four judges. Judge P ranks skater A in 1st place, skater B in 2nd place and skater C in 3rd place.

(i) Without carrying out any calculation, state the value of Spearman's rank correlation coefficient for the following ranks. Give a reason for your answer. **[1]**

Skater	A	B	C
Judge P	1	2	3
Judge Q	3	2	1

(ii) Calculate the value of Spearman's rank correlation coefficient for the following ranks. **[3]**

Skater	A	B	C
Judge P	1	2	3
Judge R	3	1	2

(iii) Judge S ranks the skaters at random. Find the probability that the value of Spearman's rank correlation coefficient between the ranks of judge P and judge S is 1. **[3]**

June 2010

3 **(i)** Some values, (x, y), of a bivariate distribution are plotted on a scatter diagram and a regression line is to be drawn. Explain how to decide whether the regression line of y on x or the regression line of x on y is appropriate. **[2]**

(ii) In an experiment the temperature, $x\,^\circ\text{C}$, of a rod was gradually increased from $0\,^\circ\text{C}$, and the extension, y mm, was measured nine times at $50\,^\circ\text{C}$ intervals. The results are summarised below.

$$n = 9 \qquad \Sigma x = 1800 \qquad \Sigma y = 14.4 \qquad \Sigma x^2 = 510\,000 \qquad \Sigma y^2 = 32.6416 \qquad \Sigma xy = 4080$$

(a) Show that the gradient of the regression line of y on x is 0.008 and find the equation of this line. **[4]**

(b) Use your equation to estimate the temperature when the extension is 2.5 mm. **[1]**

(c) Use your equation to estimate the extension for a temperature of $-50\,^\circ\text{C}$. **[1]**

(d) Comment on the meaning and the reliability of your estimate in part **(c)**. **[2]**

4 **(i)** The random variable W has the distribution $\text{B}\left(10, \frac{1}{3}\right)$. Find

(a) $\text{P}(W \leqslant 2)$, **[1]**

(b) $\text{P}(W = 2)$. **[2]**

(ii) The random variable X has the distribution $\text{B}(15, 0.22)$.

(a) Find $\text{P}(X = 4)$. **[2]**

(b) Find $\text{E}(X)$ and $\text{Var}(X)$. **[3]**

5 Each of four cards has a number printed on it as shown.

1	**2**	**3**	**3**

Two of the cards are chosen at random, without replacement. The random variable X denotes the sum of the numbers on these two cards.

(i) Show that $\text{P}(X = 6) = \frac{1}{6}$ and $\text{P}(X = 4) = \frac{1}{3}$. **[3]**

(ii) Write down all the possible values of X and find the probability distribution of X. **[4]**

(iii) Find $\text{E}(X)$ and $\text{Var}(X)$. **[5]**

6 There are 10 numbers in a list. The first 9 numbers have mean 6 and variance 2. The 10th number is 3. Find the mean and variance of all 10 numbers. **[6]**

[Questions 7 and 8 are printed overleaf.]

June 2010

7 The menu below shows all the dishes available at a certain restaurant.

Rice dishes	Main dishes	Vegetable dishes
Boiled rice	Chicken	Mushrooms
Fried rice	Beef	Cauliflower
Pilau rice	Lamb	Spinach
Keema rice	Mixed grill	Lentils
	Prawn	Potatoes
	Vegetarian	

A group of friends decide that they will share a total of 2 different rice dishes, 3 different main dishes and 4 different vegetable dishes from this menu. Given these restrictions,

(i) find the number of possible combinations of dishes that they can choose to share, **[3]**

(ii) assuming that all choices are equally likely, find the probability that they choose boiled rice. **[2]**

The friends decide to add a further restriction as follows. If they choose boiled rice, they will not choose potatoes.

(iii) Find the number of possible combinations of dishes that they can now choose. **[3]**

8 The proportion of people who watch *West Street* on television is 30%. A market researcher interviews people at random in order to contact viewers of *West Street*. Each day she has to contact a certain number of viewers of *West Street*.

(i) Near the end of one day she finds that she needs to contact just one more viewer of *West Street*. Find the probability that the number of further interviews required is

(a) 4, **[3]**

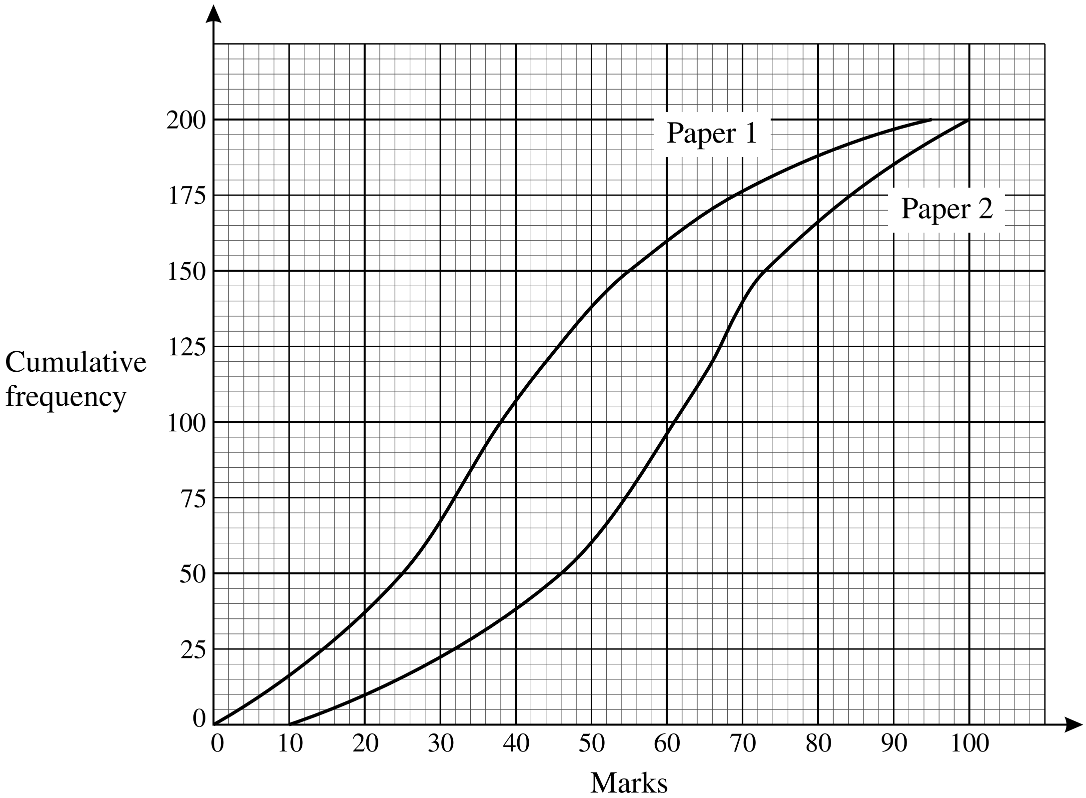
(b) less than 4. **[3]**

(ii) Near the end of another day she finds that she needs to contact just two more viewers of *West Street*. Find the probability that the number of further interviews required is

(a) 5, **[4]**

(b) more than 5. **[2]**

1 200 candidates took each of two examination papers. The diagram shows the cumulative frequency graphs for their marks.



(i) Estimate the median mark for each of the papers. **[2]**

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

(iii) It is suggested that the marks on Paper 2 were less varied than those on Paper 1. Use interquartile ranges to comment on this suggestion. **[4]**

(iv) The minimum mark for grade A, the top grade, on Paper 1 was 10 marks lower than the minimum mark for grade A on Paper 2. Given that 25 candidates gained grade A in Paper 1, find the number of candidates who gained grade A in Paper 2. **[2]**

(v) The mean and standard deviation of the marks on Paper 1 were 36.5 and 28.2 respectively. Later, a marking error was discovered and it was decided to add 1 mark to each of the 200 marks on Paper 1. State the mean and standard deviation of the new marks on Paper 1. **[2]**

2 The random variable X has the distribution Geo(0.2). Find

(i) $\mathrm{P}(X = 3)$, **[2]**

(ii) $\mathrm{P}(3 \leqslant X \leqslant 5)$, **[3]**

(iii) $\mathrm{P}(X > 4)$. **[3]**

Two independent values of X are found.

(iv) Find the probability that the total of these two values is 3. **[3]**

Jan 2011

3 A firm wishes to assess whether there is a linear relationship between the annual amount spent on advertising, £x thousand, and the annual profit, £y thousand. A summary of the figures for 12 years is as follows.

$$n = 12 \qquad \Sigma x = 86.6 \qquad \Sigma y = 943.8 \qquad \Sigma x^2 = 658.76 \qquad \Sigma y^2 = 83\,663.00 \qquad \Sigma xy = 7351.12$$

(i) Calculate the product moment correlation coefficient, showing that it is greater than 0.9. **[3]**

(ii) Comment briefly on this value in this context. **[1]**

(iii) A manager claims that this result shows that spending more money on advertising in the future will result in greater profits. Make two criticisms of this claim. **[2]**

(iv) Calculate the equation of the regression line of y on x. **[4]**

(v) Estimate the annual profit during a year when £7400 was spent on advertising. **[2]**

4 Jenny and Omar are each allowed two attempts at a high jump.

(i) The probability that Jenny will succeed on her first attempt is 0.6. If she fails on her first attempt, the probability that she will succeed on her second attempt is 0.7. Calculate the probability that Jenny will succeed. **[3]**

(ii) The probability that Omar will succeed on his first attempt is p. If he fails on his first attempt, the probability that he will succeed on his second attempt is also p. The probability that he succeeds is 0.51. Find p. **[4]**

5 30% of packets of Natural Crunch Crisps contain a free gift. Jan buys 5 packets each week.

(i) The number of free gifts that Jan receives in a week is denoted by X. Name a suitable probability distribution with which to model X, giving the value(s) of any parameter(s). State any assumption(s) necessary for the distribution to be a valid model. **[4]**

Assume now that your model is valid.

(ii) Find

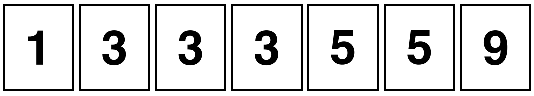
(a) $P(X \leqslant 2)$, **[1]**

(b) $P(X = 2)$. **[2]**

(iii) Find the probability that, in the next 7 weeks, there are exactly 3 weeks in which Jan receives exactly 2 free gifts. **[3]**

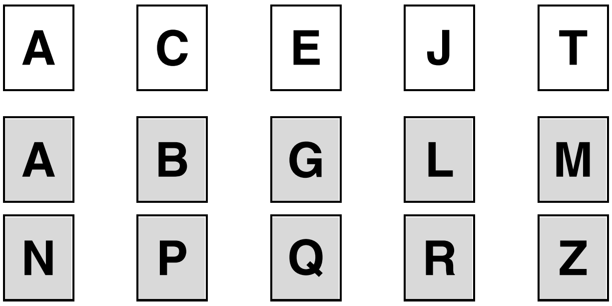
[Questions 6, 7 and 8 are printed overleaf.]

6 **(i)** The diagram shows 7 cards, each with a digit printed on it. The digits form a 7-digit number.



How many different 7-digit numbers can be formed using these cards? **[3]**

(ii) The diagram below shows 5 white cards and 10 grey cards, each with a letter printed on it.



From these cards, 3 white cards and 4 grey cards are selected at random **without** regard to order.

(a) How many selections of seven cards are possible? **[3]**

(b) Find the probability that the seven cards include exactly one card showing the letter A. **[4]**

7 The probability distribution of a discrete random variable, X, is shown below.

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

(i) Find $\mathrm{E}(X)$ in terms of a. **[2]**

(ii) Show that $\mathrm{Var}(X) = 4a(1 - a)$. **[3]**

8 Five dogs, A, B, C, D and E, took part in three races. The order in which they finished the first race was $ABCDE$.

(i) Spearman's rank correlation coefficient between the orders for the 5 dogs in the first two races was found to be -1. Write down the order in which the dogs finished the second race. **[1]**

(ii) Spearman's rank correlation coefficient between the orders for the 5 dogs in the first race and the third race was found to be 0.9.

(a) Show that, in the usual notation (as in the List of Formulae), $\Sigma d^2 = 2$. **[2]**

(b) Hence or otherwise find a possible order in which the dogs could have finished the third race. **[2]**

1 Five salesmen from a certain firm were selected at random for a survey. For each salesman, the annual income, x thousand pounds, and the distance driven last year, y thousand miles, were recorded. The results were summarised as follows.

$$n = 5 \quad \Sigma x = 251 \quad \Sigma x^2 = 14\,323 \quad \Sigma y = 65 \quad \Sigma y^2 = 855 \quad \Sigma xy = 3247$$

(i) (a) Show that the product moment correlation coefficient, r, between x and y is -0.122, correct to 3 significant figures. **[3]**

(b) State what this value of r shows about the relationship between annual income and distance driven last year for these five salesmen. **[1]**

(c) It was decided to recalculate r with the distances measured in kilometres instead of miles. State what effect, if any, this would have on the value of r. **[1]**

(ii) Another salesman from the firm is selected at random. His annual income is known to be £52 000, but the distance that he drove last year is unknown. In order to estimate this distance, a regression line based on the above data is used. Comment on the reliability of such an estimate. **[2]**

2 The orders in which 4 contestants, P, Q, R and S, were placed in two competitions are shown in the table.

Position	1st	2nd	3rd	4th
Competition 1	Q	R	S	P
Competition 2	Q	P	R	S

Calculate Spearman's rank correlation coefficient between these two orders. **[5]**

3 **(i)** A random variable, X, has the distribution $B(12, 0.85)$. Find

(a) $P(X > 10)$, **[2]**

(b) $P(X = 10)$, **[2]**

(c) $Var(X)$. **[2]**

(ii) A random variable, Y, has the distribution $B(2, \frac{1}{4})$. Two independent values of Y are found. Find the probability that the sum of these two values is 1. **[4]**

4 The table shows information about the time, t minutes correct to the nearest minute, taken by 50 people to complete a race.

Time (minutes)	$t \leqslant 27$	$28 \leqslant t \leqslant 30$	$31 \leqslant t \leqslant 35$	$36 \leqslant t \leqslant 45$	$46 \leqslant t \leqslant 60$	$t \geqslant 61$
Number of people	0	4	28	14	4	0

(i) In a histogram illustrating the data, the height of the block for the $31 \leqslant t \leqslant 35$ class is 5.6 cm. Find the height of the block for the $28 \leqslant t \leqslant 30$ class. (There is no need to draw the histogram.) **[3]**

(ii) The data in the table are used to estimate the median time. State, with a reason, whether the estimated median time is more than 33 minutes, less than 33 minutes or equal to 33 minutes. **[3]**

(iii) Calculate estimates of the mean and standard deviation of the data. **[6]**

(iv) It was found that the winner's time had been incorrectly recorded and that it was actually less than 27 minutes 30 seconds. State whether each of the following will increase, decrease or remain the same:

(a) the mean, **[1]**

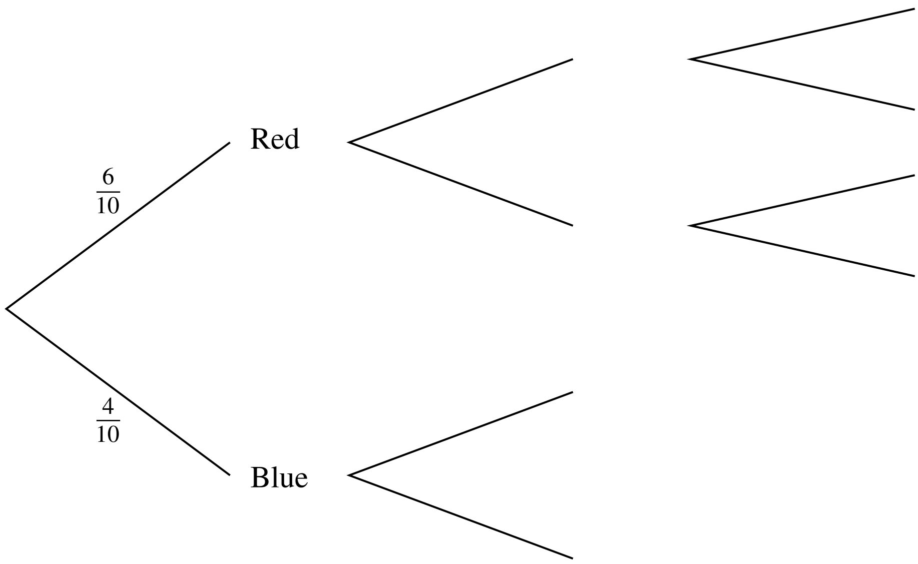
(b) the standard deviation, **[1]**

(c) the median, **[1]**

(d) the interquartile range. **[1]**

5 A bag contains 4 blue discs and 6 red discs. Chloe takes a disc from the bag. If this disc is red, she takes 2 more discs. If not, she takes 1 more disc. Each disc is taken at random and no discs are replaced.

(i) Complete the probability tree diagram in your Answer Book, showing all the probabilities. **[2]**



The total number of blue discs that Chloe takes is denoted by X.

(ii) Show that $\mathrm{P}(X = 1) = \frac{3}{5}$. **[2]**

The complete probability distribution of X is given below.

x	0	1	2
$\mathrm{P}(X = x)$	$\frac{1}{6}$	$\frac{3}{5}$	$\frac{7}{30}$

(iii) Calculate $\mathrm{E}(X)$ and $\mathrm{Var}(X)$. **[5]**

6 A group of 7 students sit in random order on a bench.

(i) **(a)** Find the number of orders in which they can sit. **[1]**

(b) The 7 students include Tom and Jerry. Find the probability that Tom and Jerry sit next to each other. **[3]**

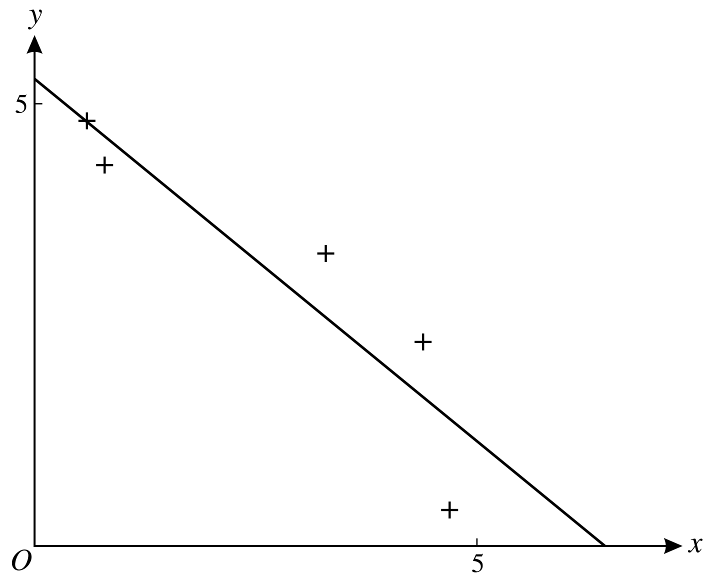
(ii) The students consist of 3 girls and 4 boys. Find the probability that

(a) no two boys sit next to each other, **[2]**

(b) all three girls sit next to each other. **[3]**

June 2011

7 The diagram shows the results of an experiment involving some bivariate data. The least squares regression line of y on x for these results is also shown.



(i) Given that the least squares regression line of y on x is used for an estimation, state which of x or y is treated as the independent variable. **[1]**

(ii) Use the diagram to explain what is meant by 'least squares'. **[2]**

(iii) State, with a reason, the value of Spearman's rank correlation coefficient for these data. **[2]**

(iv) What can be said about the value of the product moment correlation coefficient for these data? **[1]**

8 Ann, Bill, Chris and Dipak play a game with a fair cubical die. Starting with Ann they take turns, in alphabetical order, to throw the die. This process is repeated as many times as necessary until a player throws a 6. When this happens, the game stops and this player is the winner.

Find the probability that

(i) Chris wins on his first throw, **[1]**

(ii) Dipak wins on his second throw, **[3]**

(iii) Ann gets a third throw, **[2]**

(iv) Bill throws the die exactly three times. **[4]**

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.

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

(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

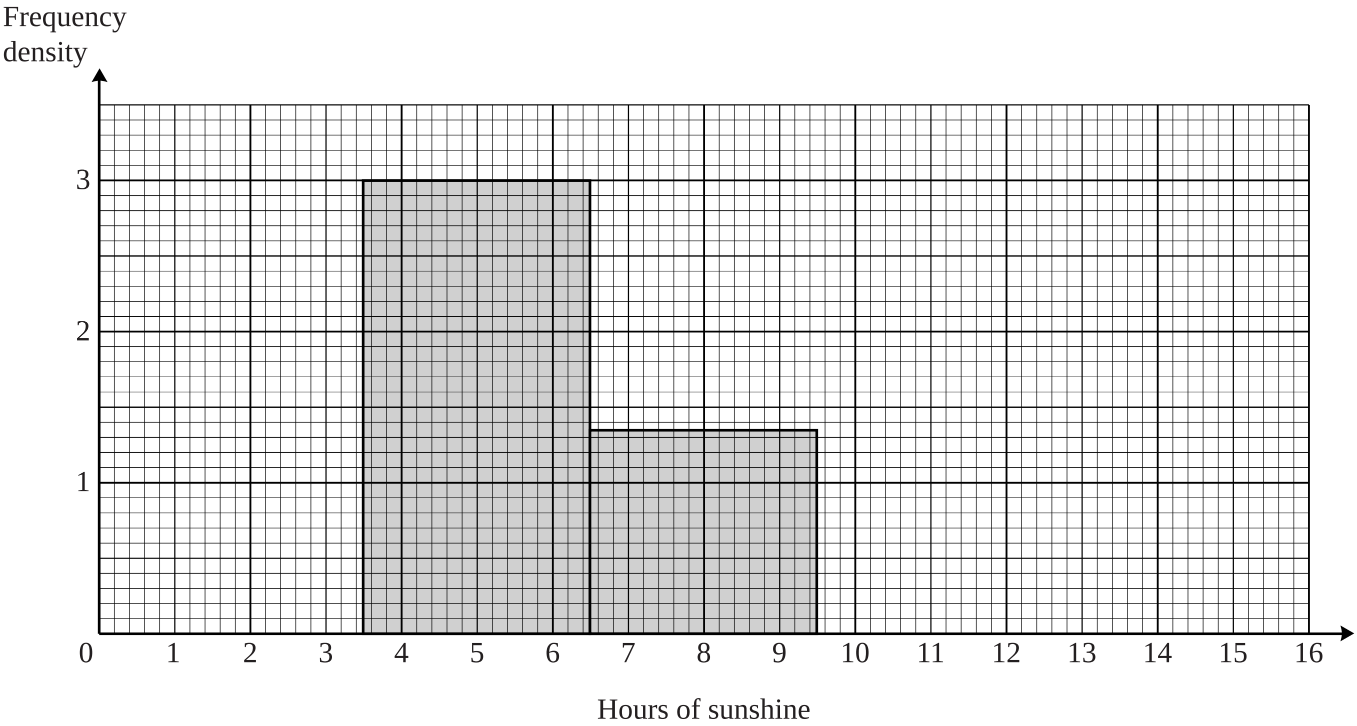
Calculate Spearman's rank correlation coefficient. **[5]**

(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]**

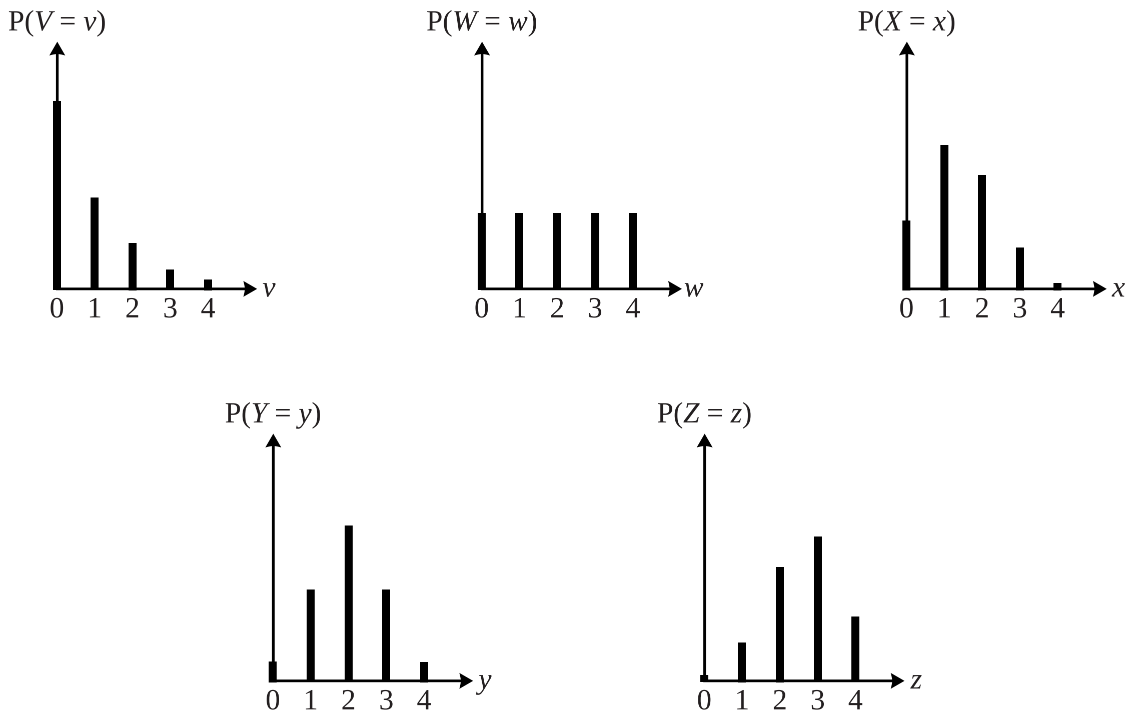
(b) Fred wishes to draw the block for the 10 – 15 class on the same diagram. Calculate the height, in centimetres, of this block. **[2]**

(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 Geo($\frac{1}{2}$). State, with a reason, which variable this is. **[2]**

(ii) One of these variables has the distribution 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) $P(X \geqslant 4)$. **[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]**

1 For each of the last five years the number of tourists, x thousands, visiting Sackton, and the average weekly sales, £ y thousands, in Sackton Stores were noted. The table shows the results.

Year	2007	2008	2009	2010	2011
x	250	270	264	290	292
y	4.2	3.7	3.2	3.5	3.0

(i) Calculate the product moment correlation coefficient r between x and y. **[4]**

(ii) It is required to estimate the average weekly sales at Sackton Stores in a year when the number of tourists is 280 000. Calculate the equation of an appropriate regression line, and use it to find this estimate. **[4]**

(iii) Over a longer period the value of r is –0.8. The mayor says, "This shows that having more tourists causes sales at Sackton Stores to decrease." Give a reason why this statement is not correct. **[1]**

2 The masses, x kg, of 50 bags of flour were measured and the results were summarised as follows.

$$n = 50 \qquad \Sigma(x - 1.5) = 1.4 \qquad \Sigma(x - 1.5)^2 = 0.05$$

Calculate the mean and standard deviation of the masses of these bags of flour. **[6]**

3 The test marks of 14 students are displayed in a stem-and-leaf diagram, as shown below.

0	
1	2 6
2	1 3 5
3	w x 4 8 y z
4	6 7 7

Key: 1 | 6 means 16 marks

(i) Find the lower quartile. **[1]**

(ii) Given that the median is 32, find the values of w and x. **[2]**

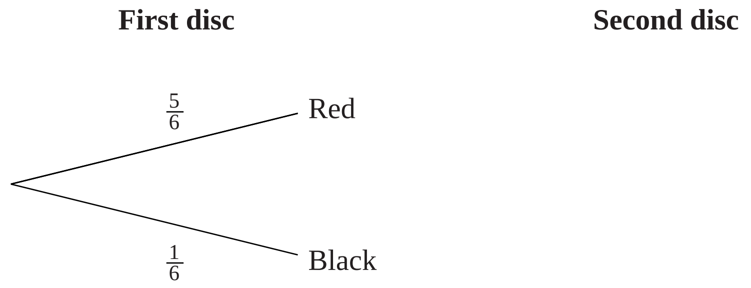
(iii) Find the possible values of the upper quartile. **[2]**

(iv) State one advantage of a stem-and-leaf diagram over a box-and-whisker plot. **[1]**

(v) State one advantage of a box-and-whisker plot over a stem-and-leaf diagram. **[1]**

4 A bag contains 5 red discs and 1 black disc. Tina takes two discs from the bag at random without replacement.

(i) The diagram shows part of a tree diagram to illustrate this situation.



Complete the tree diagram in your Answer Book showing all the probabilities. **[2]**

(ii) Find the probability that exactly one of the two discs is red. **[3]**

All the discs are replaced in the bag. Tony now takes three discs from the bag at random without replacement.

(iii) Given that the first disc Tony takes is red, find the probability that the third disc Tony takes is also red. **[2]**

5 **(i)** Write down the value of Spearman's rank correlation coefficent, r_s, for the following sets of ranks.

(a)

Judge *A* ranks	1	2	3	4
Judge *B* ranks	1	2	3	4

[1]

(b)

Judge *A* ranks	1	2	3	4
Judge *C* ranks	4	3	2	1

[1]

(ii) Calculate the value of r_s for the following ranks.

Judge *A* ranks	1	2	3	4
Judge *D* ranks	2	4	1	3

[3]

(iii) For each of parts **(i)(a)**, **(i)(b)** and **(ii)**, describe in everyday terms the relationship between the two judges' opinions. **[3]**

6 A six-sided die is biased so that the probability of scoring 6 is 0.1 and the probabilities of scoring 1, 2, 3, 4, and 5 are all equal. In a game at a fête, contestants pay £3 to roll this die. If the score is 6 they receive £10 back. If the score is 5 they receive £5 back. Otherwise they receive no money back. Find the organiser's expected profit for 100 rolls of the die. **[5]**

7 **(i)** 5 of the 7 letters A, B, C, D, E, F, G are arranged in a random order in a straight line.

(a) How many different arrangements of 5 letters are possible? **[2]**

(b) How many of these arrangements end with a vowel (A or E)? **[3]**

(ii) A group of 5 people is to be chosen from a list of 7 people.

(a) How many different groups of 5 people can be chosen? **[1]**

(b) The list of 7 people includes Jill and Jo. A group of 5 people is chosen at random from the list. Given that either Jill and Jo are both chosen or neither of them is chosen, find the probability that both of them are chosen. **[3]**

8 **(i)** The random variable X has the distribution B(30, 0.6). Find $P(X \geqslant 16)$. **[2]**

(ii) The random variable Y has the distribution B(4, 0.7).

(a) Find $P(Y = 2)$. **[2]**

(b) Three values of Y are chosen at random. Find the probability that their total is 10. **[6]**

9 **(i)** A clock is designed to chime once each hour, on the hour. The clock has a fault so that each time it is supposed to chime there is a constant probability of $\frac{1}{10}$ that it will not chime. It may be assumed that the clock never stops and that faults occur independently. The clock is started at 5 minutes past midnight on a certain day. Find the probability that the first time it does not chime is

(a) at 0600 on that day, **[3]**

(b) before 0600 on that day. **[3]**

(ii) Another clock is designed to chime twice each hour: on the hour and at 30 minutes past the hour. This clock has a fault so that each time it is supposed to chime there is a constant probability of $\frac{1}{20}$ that it will not chime. It may be assumed that the clock never stops and that faults occur independently. The clock is started at 5 minutes past midnight on a certain day.

(a) Find the probability that the first time it does not chime is at either 0030 or 0130 on that day. **[2]**

(b) Use the formula for the sum to infinity of a geometric progression to find the probability that the first time it does not chime is at 30 minutes past some hour. **[3]**

1 When a four-sided spinner is spun, the number on which it lands is denoted by X, where X is a random variable taking values 2, 4, 6 and 8. The spinner is biased so that $P(X = x) = kx$, where k is a constant.

(i) Show that $P(X = 6) = \frac{3}{10}$. **[2]**

(ii) Find $E(X)$ and $Var(X)$. **[5]**

2 **(i)** Kathryn is allowed three attempts at a high jump. If she succeeds on any attempt, she does not jump again. The probability that she succeeds on her first attempt is $\frac{3}{4}$. If she fails on her first attempt, the probability that she succeeds on her second attempt is $\frac{3}{8}$. If she fails on her first two attempts, the probability that she succeeds on her third attempt is $\frac{3}{16}$. Find the probability that she succeeds. **[3]**

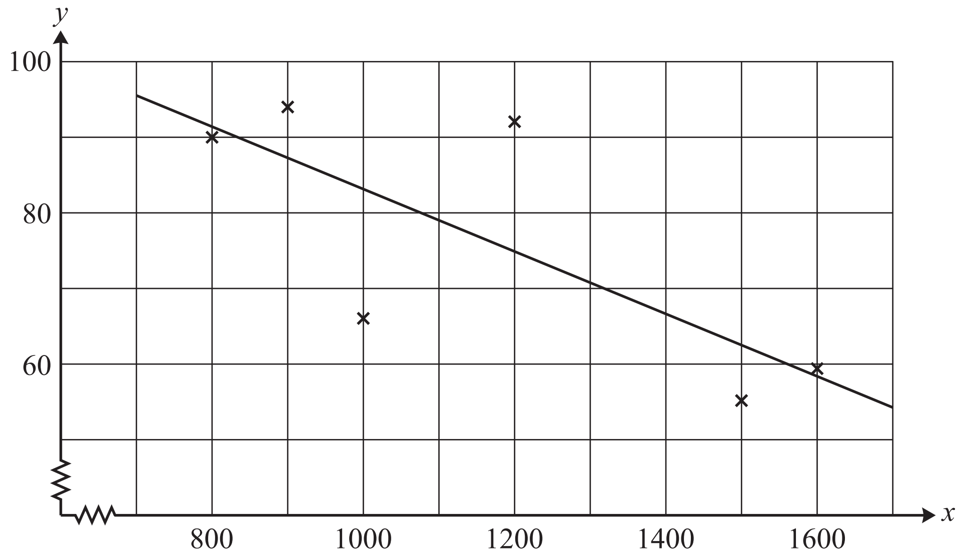
(ii) Khaled is allowed two attempts to pass an examination. If he succeeds on his first attempt, he does not make a second attempt. The probability that he passes at the first attempt is 0.4 and the probability that he passes on either the first or second attempt is 0.58. Find the probability that he passes on the second attempt, given that he failed on the first attempt. **[3]**

3 The Gross Domestic Product per Capita (GDP), x dollars, and the Infant Mortality Rate per thousand (IMR), y, of 6 African countries were recorded and summarised as follows.

$$n = 6 \qquad \sum x = 7000 \qquad \sum x^2 = 8\,700\,000 \qquad \sum y = 456 \qquad \sum y^2 = 36\,262 \qquad \sum xy = 509\,900$$

(i) Calculate the equation of the regression line of y on x for these 6 countries. **[4]**

The original data were plotted on a scatter diagram and the regression line of y on x was drawn, as shown below.



(ii) The GDP for another country, Tanzania, is 1300 dollars. Use the regression line in the diagram to estimate the IMR of Tanzania. **[1]**

(iii) The GDP for Nigeria is 2400 dollars. Give two reasons why the regression line is unlikely to give a reliable estimate for the IMR for Nigeria. **[2]**

(iv) The actual value of the IMR for Tanzania is 96. The data for Tanzania ($x = 1300$, $y = 96$) is now included with the original 6 countries. Calculate the value of the product moment correlation coefficient, r, for all 7 countries. **[4]**

(v) The IMR is now redefined as the infant mortality rate per hundred instead of per thousand, and the value of r is recalculated for all 7 countries. Without calculation state what effect, if any, this would have on the value of r found in part **(iv)**. **[1]**

4 **(i)** How many different 3-digit numbers can be formed using the digits 1, 2 and 3 when

(a) no repetitions are allowed, **[1]**

(b) any repetitions are allowed, **[2]**

(c) each digit may be included at most twice? **[2]**

(ii) How many different **4-digit** numbers can be formed using the digits 1, 2 and 3 when each digit may be included at most twice? **[5]**

Jan 2013

5 A random variable X has the distribution $B(5, \frac{1}{4})$.

(i) Find

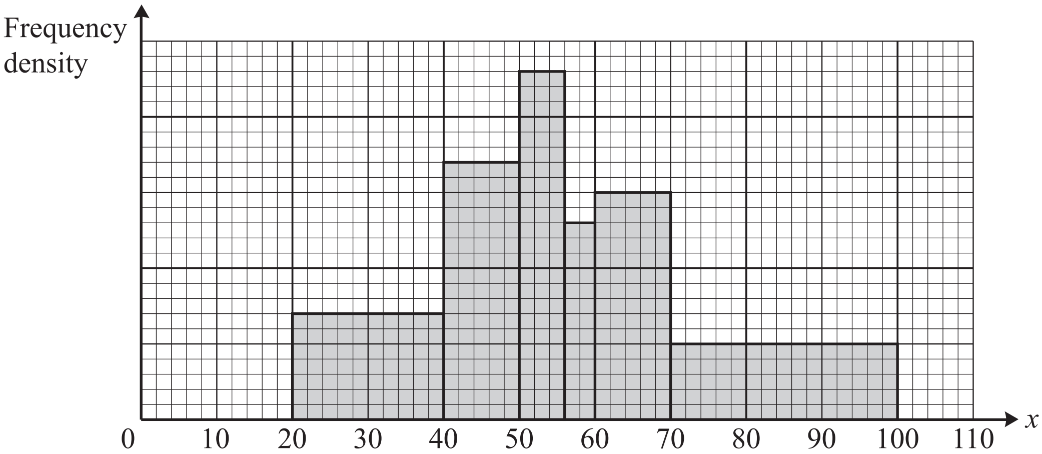
(a) $E(X)$, **[1]**

(b) $P(X = 2)$. **[2]**

(ii) Two values of X are chosen at random. Find the probability that their sum is less than 2. **[4]**

(iii) 10 values of X are chosen at random. Use an appropriate formula to find the probability that exactly 3 of these values are 2s. **[3]**

6 The masses, x grams, of 800 apples are summarised in the histogram.



(i) On the frequency density axis, 1 cm represents a units. Find the value of a. **[3]**

(ii) Find an estimate of the median mass of the apples. **[4]**

7 **(i)** Two judges rank n competitors, where n is an even number. Judge 2 reverses each consecutive pair of ranks given by Judge 1, as shown.

Competitor	C_1	C_2	C_3	C_4	C_5	C_6	....	C_{n-1}	C_n
Judge 1 rank	1	2	3	4	5	6	....	$n-1$	n
Judge 2 rank	2	1	4	3	6	5	....	n	$n-1$

Given that the value of Spearman's coefficient of rank correlation is $\frac{63}{65}$, find n. **[4]**

(ii) An experiment produced some data from a bivariate distribution. The product moment correlation coefficient is denoted by r, and Spearman's rank correlation coefficient is denoted by r_s.

(a) Explain whether the statement

$$r = 1 \Rightarrow r_s = 1$$

is true or false. **[1]**

(b) Use a diagram to explain whether the statement

$$r \neq 1 \Rightarrow r_s \neq 1$$

is true or false. **[2]**

8 Sandra makes repeated, independent attempts to hit a target. On each attempt, the probability that she succeeds is 0.1.

(i) Find the probability that

(a) the first time she succeeds is on her 5th attempt, **[2]**

(b) the first time she succeeds is after her 5th attempt, **[2]**

(c) the second time she succeeds is before her 4th attempt. **[4]**

Jill also makes repeated attempts to hit the target. Each attempt of either Jill or Sandra is independent. Each time that Jill attempts to hit the target, the probability that she succeeds is 0.2. Sandra and Jill take turns attempting to hit the target, with Sandra going first.

(ii) Find the probability that the first person to hit the target is Sandra, on her

(a) 2nd attempt, **[2]**

(b) 10th attempt. **[3]**

June 2013

1 The lengths, in centimetres, of 18 snakes are given below.

24 62 20 65 27 67 69 32 40 53 55 47 33 45 55 56 49 58

(i) Draw an ordered stem-and-leaf diagram for the data. **[3]**

(ii) Find the mean and median of the lengths of the snakes. **[2]**

(iii) It was found that one of the lengths had been measured incorrectly. After this length was corrected, the median increased by 1 cm. Give two possibilities for the incorrect length and give a corrected value in each case. **[2]**

2 **(i)** The table shows the times, in minutes, spent by five students revising for a test, and the grades that they achieved in the test.

Student	Ann	Bill	Caz	Den	Ed
Time revising	0	60	35	100	45
Grade	C	D	E	B	A

Calculate Spearman's rank correlation coefficient. **[5]**

(ii) The table below shows the ranks given by two judges to four competitors.

Competitor	P	Q	R	S
Judge 1 rank	1	2	3	4
Judge 2 rank	3	2	1	4

Spearman's rank correlation coefficient for these ranks is denoted by r_s. With the same set of ranks for Judge 1, write down a different set of ranks for Judge 2 which gives the same value of r_s. There is no need to find the value of r_s. **[2]**

3 The probability distribution of a random variable X is shown.

x	1	3	5	7
$\mathrm{P}(X=x)$	0.4	0.3	0.2	0.1

(i) Find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$. **[5]**

(ii) Three independent values of X, denoted by X_1, X_2 and X_3, are chosen. Given that $X_1 + X_2 + X_3 = 19$, write down all the possible sets of values for X_1, X_2 and X_3 and hence find $\mathrm{P}(X_1 = 7)$. **[2]**

(iii) 11 independent values of X are chosen. Use an appropriate formula to find the probability that exactly 4 of these values are 5s. **[3]**

4 At a stall in a fair, contestants have to estimate the mass of a cake. A group of 10 people made estimates, m kg, and for each person the value of $(m - 5)$ was recorded. The mean and standard deviation of $(m - 5)$ were found to be 0.74 and 0.13 respectively.

(i) Write down the mean and standard deviation of m. **[2]**

The mean and standard deviation of the estimates made by another group of 15 people were found to be 5.6 kg and 0.19 kg respectively.

(ii) Calculate the mean of all 25 estimates. **[2]**

(iii) Fiona claims that if a group's estimates are more consistent, they are likely to be more accurate. Given that the true mass of the cake is 5.65 kg, comment on this claim. **[2]**

5 The table shows some of the values of the seasonally adjusted Unemployment Rate (UR), x%, and the Consumer Price Index (CPI), y%, in the United Kingdom from April 2008 to July 2010.

Date	April 2008	July 2008	October 2008	January 2009	April 2009	July 2009	October 2009	January 2010	April 2010	July 2010
UR, x%	5.2	5.7	6.1	6.8	7.5	7.8	7.8	7.9	7.8	7.7
CPI, y%	3.0	4.4	4.5	3.0	2.3	1.8	1.5	3.5	3.7	3.1

These data are summarised below.

$$n = 10 \quad \Sigma x = 70.3 \quad \Sigma x^2 = 503.45 \quad \Sigma y = 30.8 \quad \Sigma y^2 = 103.94 \quad \Sigma xy = 211.9$$

(i) Calculate the product moment correlation coefficient, r, for the data, showing that $-0.6 < r < -0.5$. **[3]**

(ii) Karen says "The negative value of r shows that when the Unemployment Rate increases, it causes the Consumer Price Index to decrease." Give a criticism of this statement. **[1]**

(iii) (a) Calculate the equation of the regression line of x on y. **[3]**

(b) Use your equation to estimate the value of the Unemployment Rate in a month when the Consumer Price Index is 4.0%. **[2]**

6 The diagram shows five cards, each with a letter on it.

A	B	C	D	E

The letters A and E are vowels; the letters B, C and D are consonants.

(i) Two of the five cards are chosen at random, without replacement. Find the probability that they both have vowels on them. **[2]**

(ii) The two cards are replaced. Now three of the five cards are chosen at random, without replacement. Find the probability that they include exactly one card with a vowel on it. **[3]**

(iii) The three cards are replaced. Now four of the five cards are chosen at random without replacement. Find the probability that they include the card with the letter B on it. **[2]**

7 In a factory, an inspector checks a random sample of 30 mugs from a large batch and notes the number, X, which are defective. He then deals with the batch as follows.

- If $X < 2$, the batch is accepted.
- If $X > 2$, the batch is rejected.
- If $X = 2$, the inspector selects another random sample of only 15 mugs from the batch. If this second sample contains 1 or more defective mugs, the batch is rejected. Otherwise the batch is accepted.

It is given that 5% of mugs are defective.

(i) (a) Find the probability that the batch is rejected after just the first sample is checked. **[3]**

(b) Show that the probability that the batch is rejected is 0.327, correct to 3 significant figures. **[5]**

(ii) Batches are checked one after another. Find the probability that the first batch to be rejected is either the 4th or the 5th batch that is checked. **[3]**

8 **(i)** A bag contains 12 black discs, 10 white discs and 5 green discs. Three discs are drawn at random from the bag, without replacement. Find the probability that all three discs are of different colours. **[3]**

(ii) A bag contains 30 red discs and 20 blue discs. A second bag contains 50 discs, each of which is either red or blue. A disc is drawn at random from each bag. The probability that these two discs are of different colours is 0.54. Find the number of red discs that were in the second bag at the start. **[4]**

9 A game is played with a token on a board with a grid printed on it. The token starts at the point (0, 0) and moves in steps. Each step is either 1 unit in the positive x-direction with probability 0.8, or 1 unit in the positive y-direction with probability 0.2. The token stops when it reaches a point with a y-coordinate of 1. It is given that the token stops at $(X, 1)$.

(i) **(a)** Find the probability that $X = 10$. **[2]**

(b) Find the probability that $X < 10$. **[3]**

(ii) Find the expected number of steps taken by the token. **[2]**

(iii) Hence, write down the value of $\mathrm{E}(X)$. **[1]**

1 The stem-and-leaf diagram shows the heights, in metres to the nearest 0.1 m, of a random sample of trees of species A.

5	
5	9
6	1 4
6	5 5 9
7	2 3 3 4
7	5 6 6 6 7 8
8	0 3 4
8	5

Key: 6 | 4 means 6.4 m

(i) Find the median and interquartile range of the heights. **[3]**

(ii) The heights, in metres to the nearest 0.1 m, of a random sample of trees of species B are given below.

7.6 5.2 8.5 5.2 6.3 6.3 6.8 7.2 6.7 7.3 5.4 7.5 7.4 6.0 6.7

In the answer book, complete the back-to-back stem-and-leaf diagram. **[2]**

(iii) Make two comparisons between the heights of the two species of tree. **[2]**

2 **(a)** The probability distribution of a random variable W is shown in the table.

w	0	2	4
$P(W = w)$	0.3	0.4	0.3

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

(b) The random variable X has probability distribution given by

$$P(X = x) = k(x + 1) \quad \text{for } x = 1, 2, 3, 4.$$

(i) Show that $k = \frac{1}{14}$. **[1]**

(ii) Calculate $E(X)$. **[3]**

3 The table shows information about the numbers of people per household in 280 900 households in the north-west of England in 2001.

Number of people	1	2	3	4	5 or more
Number of households	86 900	92 500	45 000	37 100	19 400

(i) Taking '5 or more' to mean '5 or 6', calculate estimates of the mean and standard deviation of the number of people per household. **[5]**

(ii) State the values of the median and upper quartile of the number of people per household. **[2]**

4 Each time Ben attempts to complete a crossword in his daily newspaper, the probability that he succeeds is $\frac{2}{3}$. The random variable X denotes the number of times that Ben succeeds in 9 attempts.

(i) Find

(a) $P(X = 6)$, **[3]**

(b) $P(X < 6)$, **[1]**

(c) $E(X)$ and $Var(X)$. **[2]**

Ben notes three values, X_1, X_2 and X_3, of X.

(ii) State the total number of attempts to complete a crossword that are needed to obtain three values of X. Hence find $P(X_1 + X_2 + X_3 = 18)$. **[4]**

5 Tariq collected information about typical prices, £y million, of four-bedroomed houses at varying distances, x miles, from a large city. He chose houses at 10-mile intervals from the city. His results are shown below.

x	10	20	30	40	50	60	70	80
y	1.2	1.4	1.2	0.9	0.8	0.5	0.5	0.3

$n = 8$ $\Sigma x = 360$ $\Sigma x^2 = 20400$ $\Sigma y = 6.8$ $\Sigma y^2 = 6.88$ $\Sigma xy = 241$

(i) Use an appropriate formula to calculate the product moment correlation coefficient, r, showing that $-1.0 < r < -0.9$. **[3]**

(ii) State what this value of r shows in this context. **[1]**

(iii) Tariq decides to recalculate the value of r with the house prices measured in hundreds of thousands of pounds, instead of millions of pounds. State what effect, if any, this will have on the value of r. **[1]**

(iv) Calculate the equation of the regression line of y on x. **[3]**

(v) Explain why the regression line of y on x, rather than x on y, should be used for estimating a value of x from a given value of y. **[1]**

June 2014

6 Fiona and James collected the results for six hockey teams at the end of the season. They then carried out various calculations using Spearman's rank correlation coefficient, r_s.

(i) Fiona calculated the value of r_s between the number of goals scored FOR each team and the number of goals scored AGAINST each team. She found that $r_s = -1$. Complete the table in the answer book showing the ranks.

Team	A	B	C	D	E	F
Number of goals FOR (rank)	1	2	3	4	5	6
Number of goals AGAINST (rank)						

[1]

(ii) James calculated the value of r_s between the number of goals scored and the number of points gained by the 6 teams. He found the value of r_s to be 1. He then decided to include the results of another two teams in the calculation of r_s. The table shows the ranks for these two teams.

Team	G	H
Number of goals scored (rank)	7	8
Number of points gained (rank)	8	7

Calculate the value of r_s for all 8 teams. **[4]**

7 The table shows the numbers of members of a swimming club in certain categories.

	Male	Female
Adults	78	45
Children	52	n

It is given that $\frac{5}{8}$ of the female members are children.

(i) Find the value of n. **[2]**

(ii) Find the probability that a member chosen at random is either female or a child (or both). **[2]**

The table below shows the corresponding numbers for an athletics club.

	Male	Female
Adults	6	4
Children	5	10

(iii) Two members of the athletics club are chosen at random for a photograph.

(a) Find the probability that one of these members is a female child and the other is an adult male. **[2]**

(b) Find the probability that exactly one of these members is female and exactly one is a child. **[2]**

June 2014

8 A group of 8 people, including Kathy, David and Harpreet, are planning a theatre trip.

(i) Four of the group are chosen at random, without regard to order, to carry the refreshments. Find the probability that these 4 people include Kathy and David but not Harpreet. **[3]**

(ii) The 8 people sit in a row. Kathy and David sit next to each other and Harpreet sits at the left-hand end of the row. How many different arrangements of the 8 people are possible? **[3]**

(iii) The 8 people stand in a line to queue for the exit. Kathy and David stand next to each other and Harpreet stands next to them. How many different arrangements of the 8 people are possible? **[3]**

9 Each day Harry makes repeated attempts to light his gas fire. If the fire lights he makes no more attempts. On each attempt, the probability that the fire will light is 0.3 independent of all other attempts. Find the probability that

(i) the fire lights on the 5th attempt, **[2]**

(ii) Harry needs more than 1 attempt but fewer than 5 attempts to light the fire. **[3]**

If the fire does not light on the 6th attempt, Harry stops and the fire remains unlit.

(iii) Find the probability that, on a particular day, the fire lights. **[3]**

(iv) Harry's week starts on Monday. Find the probability that, during a certain week, the first day on which the fire lights is Wednesday. **[2]**

END OF QUESTION PAPER