**NOTICE TO CUSTOMER:**

**The sale of this product is intended for use of the original purchaser only and for use only on a single computer system. Duplicating, selling, or otherwise distributing this product is a violation of the law ; your license of the product will be terminated at any moment if you are selling or distributing the products.**

Answer **all** questions.

**1** The table shows the length, in centimetres, and maximum diameter, in centimetres, of each of 10 honeydew melons selected at random from those on display at a market stall.

| **Length** | 24 | 25 | 19 | 28 | 27 | 21 | 35 | 23 | 32 | 26 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Maximum diameter** | 18 | 14 | 16 | 11 | 13 | 14 | 12 | 16 | 15 | 14 |

(a) Calculate the value of the product moment correlation coefficient. *(3 marks)*

(b) Interpret your value in the context of this question. *(2 marks)*

**2** The British and Irish Lions 2005 rugby squad contained 50 players. The nationalities and playing positions of these players are shown in the table.

| | | **Nationality** | | | |
|---|---|---|---|---|---|
| | | **English** | **Welsh** | **Scottish** | **Irish** |
| **Playing position** | **Forward** | 14 | 5 | 2 | 6 |
| | **Back** | 8 | 7 | 2 | 6 |

(a) A player was selected at random from the squad for a radio interview. Calculate the probability that the player was:

(i) English; *(2 marks)*

(ii) Irish, given that the player was a back; *(2 marks)*

(iii) a forward, given that the player was not Scottish. *(2 marks)*

(b) Four players were selected at random from the squad to visit a school. Calculate the probability that all four players were English. *(3 marks)*

**3** Payton has a pay-as-you-go internet account. To save money, he reads his e-mail messages just after 6 pm each day. The probability that he has no e-mail messages to read at this time is 0.45, and the number of e-mail messages he receives is independent from day to day.

(a) Calculate the probability that Payton has no e-mail messages to read on exactly 3 days during a 7-day period. *(3 marks)*

(b) Determine the probability that, during June (30 days), Payton has no e-mail messages to read:

(i) on fewer than 15 days; *(2 marks)*

(ii) on more than 10 days; *(2 marks)*

(iii) on at least 12 but at most 18 days. *(3 marks)*

**4** A library allows each member to have up to 15 books on loan at any one time.

The table shows the numbers of books currently on loan to a random sample of 95 members of the library.

| **Number of books on loan** | 0 | 1 | 2 | 3 | 4 | 5–9 | 10–14 | 15 |
|---|---|---|---|---|---|---|---|---|
| **Number of members** | 4 | 13 | 24 | 17 | 15 | 11 | 5 | 6 |

(a) For these data:

(i) state values for the mode and range; *(2 marks)*

(ii) determine values for the median and interquartile range; *(4 marks)*

(iii) calculate estimates of the mean and standard deviation. *(4 marks)*

(b) Making reference to your answers to part (a), give a reason for preferring:

(i) the median and interquartile range to the mean and standard deviation for summarising the given data; *(1 mark)*

(ii) the mean and standard deviation to the mode and range for summarising the given data. *(1 mark)*

**Turn over for the next question**

**5** Bob, a gardener, measures the time taken, $y$ minutes, for 60 grams of weedkiller pellets to dissolve in 10 litres of water at different set temperatures, $x$ °C. His results are shown in the table.

| $x$ | 16 | 20 | 24 | 28 | 32 | 36 | 40 | 44 | 48 | 52 | 56 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 4.7 | 4.3 | 3.8 | 3.5 | 3.0 | 2.7 | 2.4 | 2.0 | 1.8 | 1.6 | 1.1 |

(a) Calculate the equation of the least squares regression line $y = a + bx$. *(4 marks)*

(b) (i) Interpret, in the context of this question, your value for $b$. *(2 marks)*

(ii) Explain why no sensible practical interpretation can be given for your value of $a$. *(2 marks)*

**6** (a) The length, $X$ centimetres, of adult male eels in a river may be assumed to be normally distributed with a mean of 38 and a standard deviation of 5.

Determine:

(i) $P(X < 40)$; *(2 marks)*

(ii) $P(30 < X < 40)$; *(3 marks)*

(iii) the length exceeded by 75% of adult male eels in the river. *(4 marks)*

(b) A sample of 40 adult female eels was taken at random from the river and the length of each eel was measured.

The mean and standard deviation of these lengths were found to be 107 cm and 19.1 cm respectively.

(i) Construct a 98% confidence interval for the mean length of adult female eels in the river. *(4 marks)*

(ii) Hence comment on a claim that, in this river, the average length of adult female eels is more than $2\frac{1}{2}$ times that of adult male eels. *(3 marks)*

**END OF QUESTIONS**

Answer **all** questions.

**1** The table shows the length, in centimetres, and maximum diameter, in centimetres, of each of 10 honeydew melons selected at random from those on display at a market stall.

| **Length** | 24 | 25 | 19 | 28 | 27 | 21 | 35 | 23 | 32 | 26 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Maximum diameter** | 18 | 14 | 16 | 11 | 13 | 14 | 12 | 16 | 15 | 14 |

(a) Calculate the value of the product moment correlation coefficient. *(3 marks)*

(b) Interpret your value in the context of this question. *(2 marks)*

**2** The British and Irish Lions 2005 rugby squad contained 50 players. The nationalities and playing positions of these players are shown in the table.

| | | **Nationality** | | | |
|---|---|---|---|---|---|
| | | **English** | **Welsh** | **Scottish** | **Irish** |
| **Playing position** | **Forward** | 14 | 5 | 2 | 6 |
| | **Back** | 8 | 7 | 2 | 6 |

(a) A player was selected at random from the squad for a radio interview. Calculate the probability that the player was:

(i) a Welsh back; *(1 mark)*

(ii) English; *(2 marks)*

(iii) not English; *(1 mark)*

(iv) Irish, given that the player was a back; *(2 marks)*

(v) a forward, given that the player was not Scottish. *(2 marks)*

(b) Four players were selected at random from the squad to visit a school. Calculate the probability that all four players were English. *(3 marks)*

**3** (a) A sample of 50 washed baking potatoes was selected at random from a large batch. The weights of the 50 potatoes were found to have a mean of 234 grams and a standard deviation of 25.1 grams.

Construct a 95% confidence interval for the mean weight of potatoes in the batch. *(4 marks)*

(b) The batch of potatoes is purchased by a market stallholder. He sells them to his customers by allowing them to choose any 5 potatoes for £1.

Give a reason why such chosen potatoes are unlikely to represent a random sample from the batch. *(1 mark)*

**4** A library allows each member to have up to 15 books on loan at any one time.

The table shows the numbers of books currently on loan to a random sample of 95 members of the library.

| **Number of books on loan** | 0 | 1 | 2 | 3 | 4 | 5–9 | 10–14 | 15 |
|---|---|---|---|---|---|---|---|---|
| **Number of members** | 4 | 13 | 24 | 17 | 15 | 11 | 5 | 6 |

(a) For these data:

(i) state values for the mode and range; *(2 marks)*

(ii) determine values for the median and interquartile range; *(4 marks)*

(iii) calculate estimates of the mean and standard deviation. *(4 marks)*

(b) Making reference to your answers to part (a), give a reason for preferring:

(i) the median and interquartile range to the mean and standard deviation for summarising the given data; *(1 mark)*

(ii) the mean and standard deviation to the mode and range for summarising the given data. *(1 mark)*

**5** Bob, a gardener, measures the time taken, $y$ minutes, for 60 grams of weedkiller pellets to dissolve in 10 litres of water at different set temperatures, $x$ °C. His results are shown in the table.

| $x$ | 16 | 20 | 24 | 28 | 32 | 36 | 40 | 44 | 48 | 52 | 56 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 4.7 | 4.3 | 3.8 | 3.5 | 3.0 | 2.7 | 2.4 | 2.0 | 1.8 | 1.6 | 1.1 |

(a) State why the explanatory variable is temperature. *(1 mark)*

(b) Calculate the equation of the least squares regression line $y = a + bx$. *(4 marks)*

(c) (i) Interpret, in the context of this question, your value for $b$. *(2 marks)*

(ii) Explain why no sensible practical interpretation can be given for your value of $a$. *(2 marks)*

(d) (i) Estimate the time taken to dissolve 60 grams of weedkiller pellets in 10 litres of water at 30 °C. *(2 marks)*

(ii) Show why the equation cannot be used to make a valid estimate of the time taken to dissolve 60 grams of weedkiller pellets in 10 litres of water at 75 °C. *(2 marks)*

**6** Each weekday, Monday to Friday, Trina catches a train from her local station. She claims that the probability that the train arrives on time at the station is 0.4, and that the train's arrival time is independent from day to day.

(a) Assuming her claims to be true, determine the probability that the train arrives on time at the station:

(i) on at most 3 days during a 2-week period (10 days); *(2 marks)*

(ii) on more than 10 days but fewer than 20 days during an 8-week period. *(3 marks)*

(b) (i) Assuming Trina's claims to be true, determine the mean and standard deviation for the number of times during a week (5 days) that the train arrives on time at the station. *(3 marks)*

(ii) Each week, for a period of 13 weeks, Trina's travelling colleague, Suzie, records the number of times that the train arrives on time at the station. Suzie's results are

2 2 4 1 2 3 3 2 2 0 3 2 0

Calculate the mean and standard deviation of these values. *(3 marks)*

(iii) Hence comment on the likely validity of Trina's claims. *(2 marks)*

**7** (a) Electra is employed by E & G Ltd to install electricity meters in new houses on an estate. Her time, $X$ minutes, to install a meter may be assumed to be normally distributed with a mean of 48 and a standard deviation of 20.

Determine:

(i) $P(X < 60)$; *(2 marks)*

(ii) $P(30 < X < 60)$; *(3 marks)*

(iii) the time, $k$ minutes, such that $P(X < k) = 0.9$. *(4 marks)*

(b) Gazali is employed by E & G Ltd to install gas meters in the same new houses. His time, $Y$ minutes, to install a meter has a mean of 37 and a standard deviation of 25.

(i) Explain why $Y$ is unlikely to be normally distributed. *(2 marks)*

(ii) State why $\overline{Y}$, the mean of a random sample of 35 gas meter installations, is likely to be approximately normally distributed. *(1 mark)*

(iii) Determine $P(\overline{Y} > 40)$. *(4 marks)*

**END OF QUESTIONS**

**Practice 3**

Leave blank

**1.** A young family were looking for a new 3 bedroom semi-detached house. A local survey recorded the price $x$, in £1000, and the distance $y$, in miles, from the station of such houses. The following summary statistics were provided

$$S_{xx} = 113\,573, \quad S_{yy} = 8.657, \quad S_{xy} = -808.917$$

(a) Use these values to calculate the product moment correlation coefficient.

**(2)**

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

**(1)**

Another family asked for the distances to be measured in km rather than miles.

(c) State the value of the product moment correlation coefficient in this case.

**(1)**

Leave blank

**Question 1 continued**

**(Total 4 marks)**

**Q1**

Leave blank

**2.** The box plot in Figure 1 shows a summary of the weights of the luggage, in kg, for each musician in an orchestra on an overseas tour.

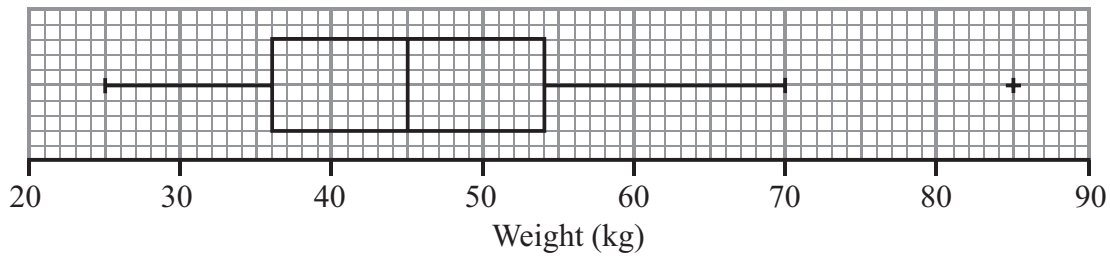



**Figure 1**

The airline's recommended weight limit for each musician's luggage was 45 kg.
Given that none of the musicians' luggage weighed exactly 45 kg,

(a) state the proportion of the musicians whose luggage was below the recommended weight limit.

**(1)**

A quarter of the musicians had to pay a charge for taking heavy luggage.

(b) State the smallest weight for which the charge was made.

**(1)**

(c) Explain what you understand by the + on the box plot in Figure 1, and suggest an instrument that the owner of this luggage might play.

**(2)**

Leave blank

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

**(2)**

One musician of the orchestra suggests that the weights of luggage, in kg, can be modelled by a normal distribution with quartiles as given in Figure 1.

(e) Find the standard deviation of this normal distribution.

**(4)**

Leave blank

**Question 2 continued**

Leave blank

**Question 2 continued**

**(Total 10 marks)**

**Q2**

Leave blank

**3.** A student is investigating the relationship between the price ( $y$ pence) of 100g of chocolate and the percentage ($x$%) of cocoa solids in the chocolate.
The following data is obtained

| Chocolate brand | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* |
|---|---|---|---|---|---|---|---|---|
| $x$ (% cocoa) | 10 | 20 | 30 | 35 | 40 | 50 | 60 | 70 |
| $y$ (pence) | 35 | 55 | 40 | 100 | 60 | 90 | 110 | 130 |

(You may use: $\sum x = 315, \sum x^2 = 15\,225, \sum y = 620, \sum y^2 = 56\,550, \sum xy = 28\,750$)

(a) On the graph paper on page 9 draw a scatter diagram to represent these data.

**(2)**

(b) Show that $S_{xy} = 4337.5$ and find $S_{xx}$.

**(3)**

The student believes that a linear relationship of the form $y = a + bx$ could be used to describe these data.

(c) Use linear regression to find the value of $a$ and the value of $b$, giving your answers to 1 decimal place.

**(4)**

(d) Draw the regression line on your scatter diagram.

**(2)**

The student believes that one brand of chocolate is overpriced.

(e) Use the scatter diagram to

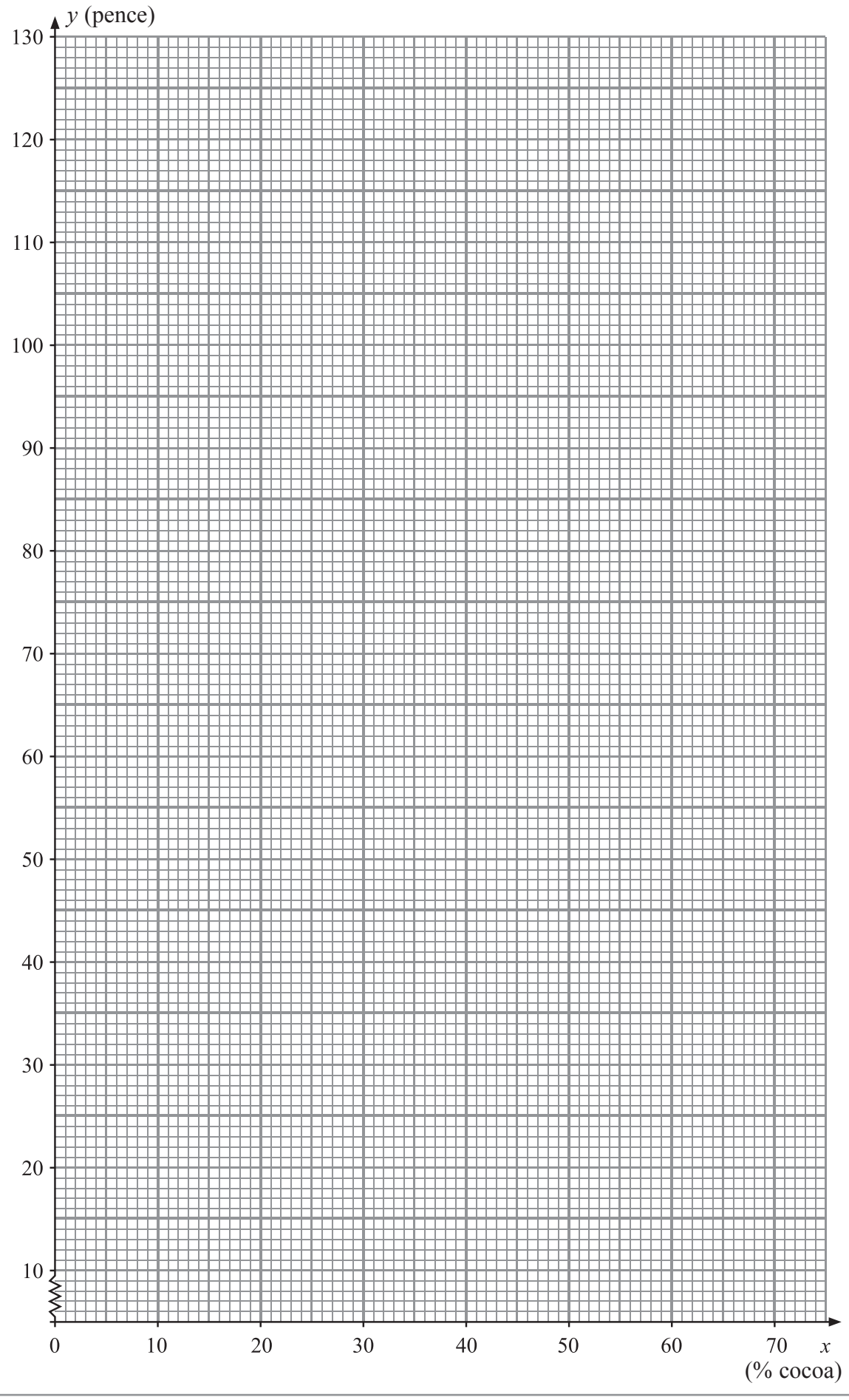
(i) state which brand is overpriced,
(ii) suggest a fair price for this brand.

Give reasons for both your answers.

**(4)**

Leave blank

**Question 3 continued**



Leave blank

**The data on page 8 has been repeated here to help you**

| Chocolate brand | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* |
|---|---|---|---|---|---|---|---|---|
| $x$ (% cocoa) | 10 | 20 | 30 | 35 | 40 | 50 | 60 | 70 |
| $y$ (pence) | 35 | 55 | 40 | 100 | 60 | 90 | 110 | 130 |

(You may use: $\sum x = 315, \sum x^2 = 15\,225, \sum y = 620, \sum y^2 = 56\,550, \sum xy = 28\,750$)

**Question 3 continued**

Leave blank

**Question 3 continued**

**(Total 15 marks)**

**Q3**

Leave blank

**4.** A survey of the reading habits of some students revealed that, on a regular basis, 25% read quality newspapers, 45% read tabloid newspapers and 40% do not read newspapers at all.

(a) Find the proportion of students who read both quality and tabloid newspapers.

**(3)**

(b) In the space on page 13 draw a Venn diagram to represent this information.

**(3)**

A student is selected at random. Given that this student reads newspapers on a regular basis,

(c) find the probability that this student only reads quality newspapers.

**(3)**

**Question 4 continued**

Leave blank

**Q4**

**(Total 9 marks)**

Leave blank

**5.**

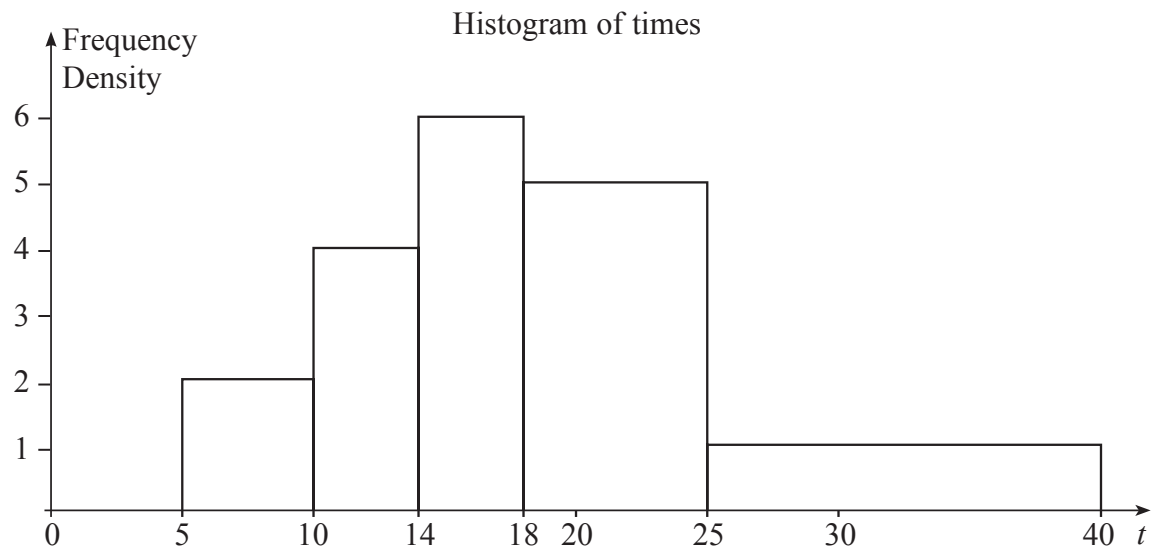



**Figure 2**

Figure 2 shows a histogram for the variable $t$ which represents the time taken, in minutes, by a group of people to swim 500m.

(a) Complete the frequency table for $t$.

| $t$ | 5–10 | 10–14 | 14–18 | 18–25 | 25–40 |
|---|---|---|---|---|---|
| Frequency | 10 | 16 | 24 | | |

**(2)**

(b) Estimate the number of people who took longer than 20 minutes to swim 500m.

**(2)**

(c) Find an estimate of the mean time taken.

**(4)**

(d) Find an estimate for the standard deviation of $t$.

**(3)**

(e) Find the median and quartiles for $t$.

**(4)**

One measure of skewness is found using $\frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$.

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

**(2)**

Leave blank

**Question 5 continued**

**Question 5 continued**

Leave blank

Leave blank

**Question 5 continued**

**(Total 17 marks)**

**Q5**

Leave blank

**6.** The random variable $X$ has a normal distribution with mean 20 and standard deviation 4.

(a) Find P($X > 25$).

**(3)**

(b) Find the value of $d$ such that P($20 < X < d$) = 0.4641

**(4)**

Leave blank

**Question 6 continued**

**(Total 7 marks)**

**Q6**

Leave blank

**7.** The random variable $X$ has probability distribution

| $x$ | 1 | 3 | 5 | 7 | 9 |
|---|---|---|---|---|---|
| $P(X = x)$ | 0.2 | $p$ | 0.2 | $q$ | 0.15 |

(a) Given that $E(X) = 4.5$, write down two equations involving $p$ and $q$.

**(3)**

Find

(b) the value of $p$ and the value of $q$,

**(3)**

(c) $P(4 < X \leqslant 7)$.

**(2)**

Given that $E(X^2) = 27.4$, find

(d) $Var(X)$,

**(2)**

(e) $E(19 - 4X)$,

**(1)**

(f) $Var(19 - 4X)$.

**(2)**

Leave blank

**Question 7 continued**

**Question 7 continued**

Leave blank

Leave blank

**Question 7 continued**

**Q7**

**(Total 13 marks)**

**TOTAL FOR PAPER: 75 MARKS**

**END**

**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

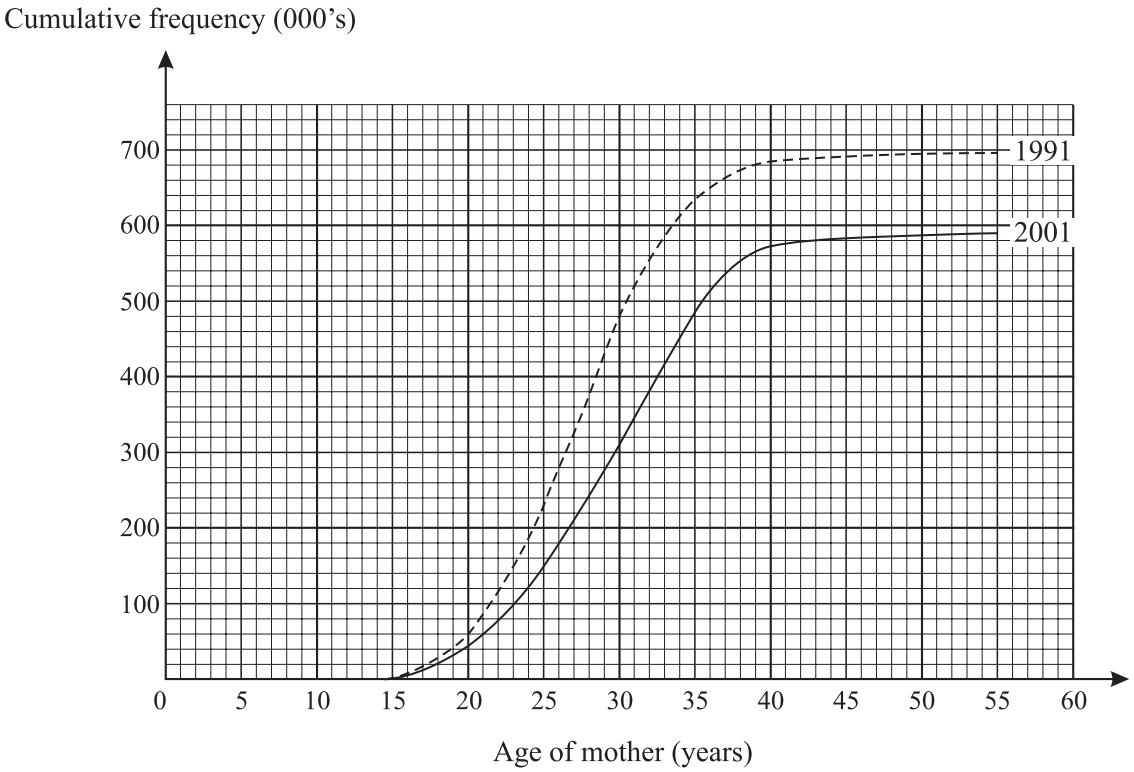
**(a)** the second disc is blue, given that the first disc was blue, [1]

**(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]

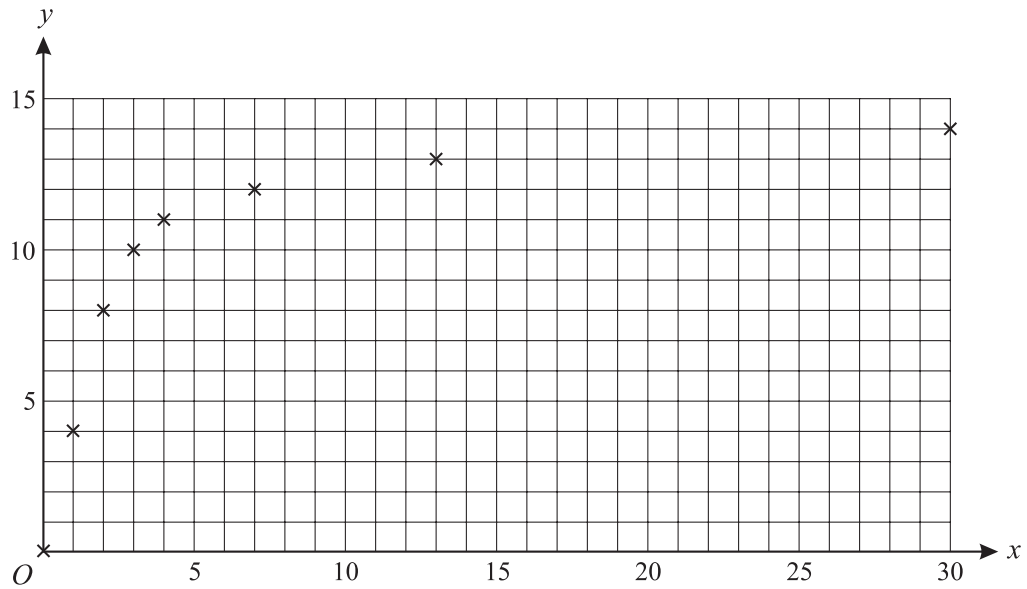
**(b)** Did the women who gave birth in 2001 tend to be younger or older or about the same age as the women who gave birth in 1991? Using the table and your values from part **(a)**, give two reasons for your answer. [3]

**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]

**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]

Answer **all** questions.

**1** The table shows the times taken, $y$ minutes, for a wood glue to dry at different air temperatures, $x$ °C.

| $x$ | 10 | 12 | 15 | 18 | 20 | 22 | 25 | 28 | 30 |
|---|---|---|---|---|---|---|---|---|---|
| $y$ | 42.9 | 40.6 | 38.5 | 35.4 | 33.0 | 30.7 | 28.0 | 25.3 | 22.6 |

(a) Calculate the equation of the least squares regression line $y = a + bx$. *(4 marks)*

(b) Estimate the time taken for the glue to dry when the air temperature is 21 °C. *(2 marks)*

**2** A basket in a stationery store contains a total of 400 marker and highlighter pens. Of the marker pens, some are permanent and the rest are non-permanent. The colours and types of pen are shown in the table.

| | **Colour** | | | |
|---|---|---|---|---|
| **Type** | **Black** | **Blue** | **Red** | **Green** |
| **Permanent marker** | 44 | 66 | 32 | 18 |
| **Non-permanent marker** | 36 | 53 | 21 | 10 |
| **Highlighter** | 0 | 41 | 37 | 42 |

A pen is selected at random from the basket. Calculate the probability that it is:

(a) a blue pen; *(1 mark)*

(b) a marker pen; *(2 marks)*

(c) a blue pen or a marker pen; *(2 marks)*

(d) a green pen, given that it is a highlighter pen. *(2 marks)*

**3** [**Figure 1**, printed on the insert, is provided for use in this question.]

The table shows, for each of a sample of 12 handmade decorative ceramic plaques, the length, $x$ millimetres, and the width, $y$ millimetres.

| **Plaque** | $x$ | $y$ |
|---|---|---|
| A | 232 | 109 |
| B | 235 | 112 |
| C | 236 | 114 |
| D | 234 | 118 |
| E | 230 | 117 |
| F | 230 | 113 |
| G | 246 | 121 |
| H | 240 | 125 |
| I | 244 | 128 |
| J | 241 | 122 |
| K | 246 | 126 |
| L | 245 | 123 |

(a) Calculate the value of the product moment correlation coefficient between $x$ and $y$. *(3 marks)*

(b) Interpret your value in the context of this question. *(2 marks)*

(c) On **Figure 1**, complete the scatter diagram for these data. *(3 marks)*

(d) In fact, the 6 plaques A, B, ..., F are from a different source to the 6 plaques G, H, ..., L.

With reference to your scatter diagram, **but without further calculations**, estimate the value of the product moment correlation coefficient between $x$ and $y$ for **each** source of plaque. *(2 marks)*

**4** For the adult population of the UK, 35 per cent of men and 29 per cent of women do not wear glasses or contact lenses.

(a) Determine the probability that, in a random sample of 40 men, at most 15 do not wear glasses or contact lenses. *(3 marks)*

(b) Calculate the probability that, in a random sample of 10 women, exactly 3 do not wear glasses or contact lenses. *(3 marks)*

(c) (i) Calculate the mean and the variance for the number who **do** wear glasses or contact lenses in a random sample of 20 women. *(3 marks)*

(ii) The numbers wearing glasses or contact lenses in 10 groups, each of 20 women, had a mean of 16.5 and a variance of 2.50 .

Comment on the claim that these 10 groups were **not** random samples. *(3 marks)*

**5** Vernon, a service engineer, is expected to carry out a boiler service in one hour.

One hour is subtracted from each of his actual times, and the resulting differences, $x$ minutes, for a random sample of 100 boiler services have a mean, $\bar{x}$, of 1.90 and a standard deviation, $s$, of 3.32 .

(a) Deduce, in minutes, the mean and the standard deviation of Vernon's actual service times for this sample. *(3 marks)*

(b) Construct a 98% confidence interval for the mean time taken by Vernon to carry out a boiler service. *(4 marks)*

(c) Vernon claims that, on average, a boiler service takes much longer than an hour.

Comment, with a justification, on this claim. *(1 mark)*

**6** The length, $L$ centimetres, of *Slimline* bin liners may be modelled by a normal distribution with a mean of 69.5 and a standard deviation of 0.55 .

(a) Determine:

(i) $\mathrm{P}(L < 70)$; *(3 marks)*

(ii) $\mathrm{P}(69 < L < 70)$; *(3 marks)*

(iii) $\mathrm{P}(L = 70)$. *(1 mark)*

(b) Determine the maximum length exceeded by 90% of bin liners. *(4 marks)*

(c) The bin liners are sold in packets of 20, and those in each packet may be considered to be a random sample.

Determine the probability that:

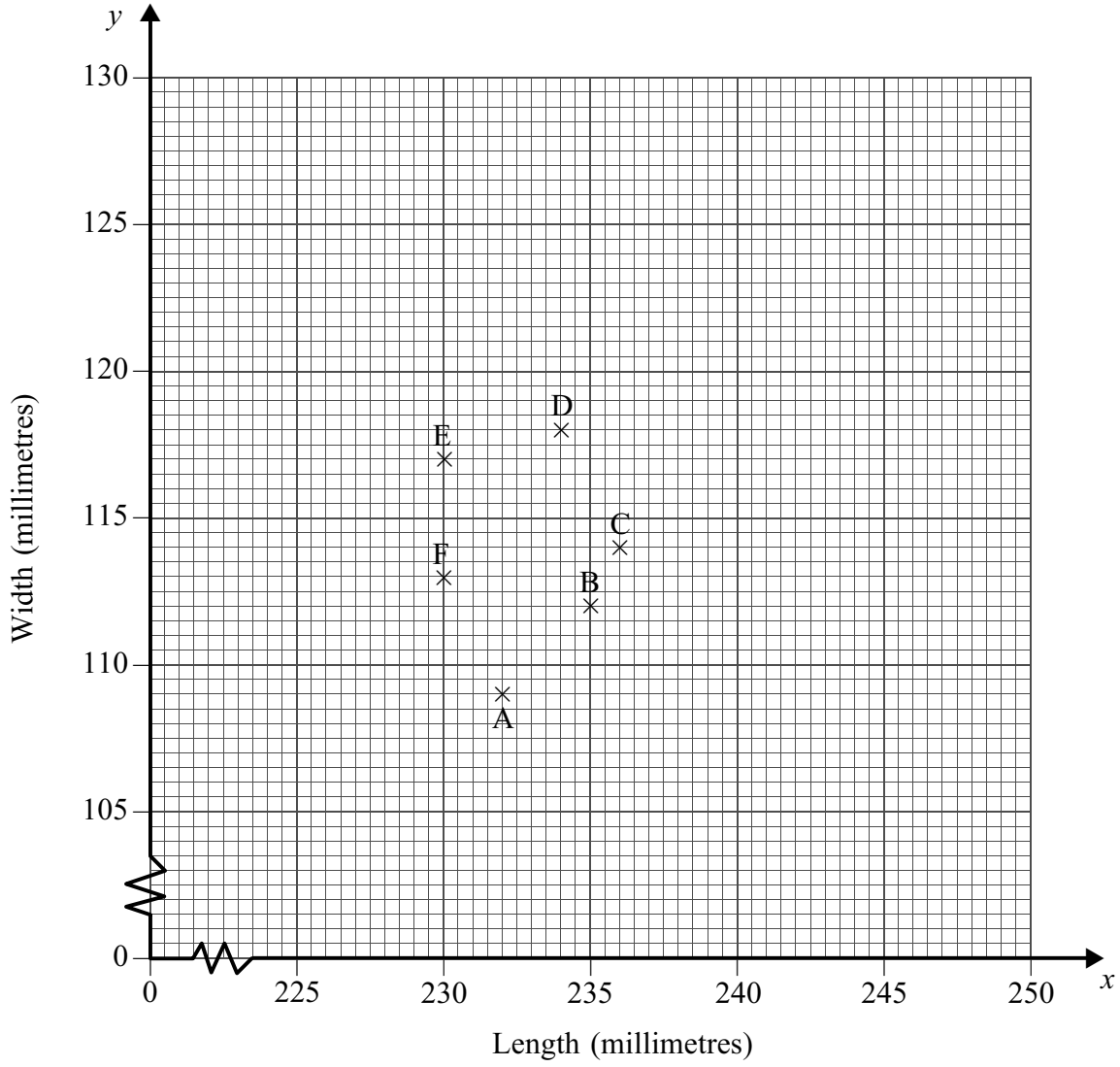
(i) all the bin liners in a packet have lengths less than 70 cm; *(2 marks)*

(ii) the mean length of the bin liners in a packet is greater than 69.25 cm. *(4 marks)*

**END OF QUESTIONS**

**Figure 1 (for use in Question 3)**

**Decorative Plaques**



Answer **all** questions.

**1** The table shows the times taken, $y$ minutes, for a wood glue to dry at different air temperatures, $x$ °C.

| $x$ | 10 | 12 | 15 | 18 | 20 | 22 | 25 | 28 | 30 |
|---|---|---|---|---|---|---|---|---|---|
| $y$ | 42.9 | 40.6 | 38.5 | 35.4 | 33.0 | 30.7 | 28.0 | 25.3 | 22.6 |

(a) Calculate the equation of the least squares regression line $y = a + bx$. *(4 marks)*

(b) Estimate the time taken for the glue to dry when the air temperature is 21 °C. *(2 marks)*

**2** A basket in a stationery store contains a total of 400 marker and highlighter pens. Of the marker pens, some are permanent and the rest are non-permanent. The colours and types of pen are shown in the table.

| | **Colour** | | | |
|---|---|---|---|---|
| **Type** | **Black** | **Blue** | **Red** | **Green** |
| **Permanent marker** | 44 | 66 | 32 | 18 |
| **Non-permanent marker** | 36 | 53 | 21 | 10 |
| **Highlighter** | 0 | 41 | 37 | 42 |

A pen is selected at random from the basket. Calculate the probability that it is:

(a) a blue pen; *(1 mark)*

(b) a marker pen; *(2 marks)*

(c) a blue pen or a marker pen; *(2 marks)*

(d) a green pen, given that it is a highlighter pen; *(2 marks)*

(e) a non-permanent marker pen, given that it is a red pen. *(2 marks)*

**3** [**Figure 1**, printed on the insert, is provided for use in this question.]

The table shows, for each of a sample of 12 handmade decorative ceramic plaques, the length, $x$ millimetres, and the width, $y$ millimetres.

| **Plaque** | $x$ | $y$ |
|---|---|---|
| A | 232 | 109 |
| B | 235 | 112 |
| C | 236 | 114 |
| D | 234 | 118 |
| E | 230 | 117 |
| F | 230 | 113 |
| G | 246 | 121 |
| H | 240 | 125 |
| I | 244 | 128 |
| J | 241 | 122 |
| K | 246 | 126 |
| L | 245 | 123 |

(a) Calculate the value of the product moment correlation coefficient between $x$ and $y$. *(3 marks)*

(b) Interpret your value in the context of this question. *(2 marks)*

(c) On **Figure 1**, complete the scatter diagram for these data. *(3 marks)*

(d) In fact, the 6 plaques A, B, ..., F are from a different source to the 6 plaques G, H, ..., L.

With reference to your scatter diagram, **but without further calculations**, estimate the value of the product moment correlation coefficient between $x$ and $y$ for **each** source of plaque. *(2 marks)*

**4** The runs scored by a cricketer in 11 innings during the 2006 season were as follows.

47 63 0 28 40 51 $a$ 77 0 13 35

The exact value of $a$ was unknown but it was greater than 100.

(a) Calculate the median and the interquartile range of these 11 values. *(4 marks)*

(b) Give a reason why, for these 11 values:

(i) the mode is **not** an appropriate measure of average;

(ii) the range is **not** an appropriate measure of spread. *(2 marks)*

**5** When a particular make of tennis ball is dropped from a vertical distance of 250 cm on to concrete, the height, $X$ centimetres, to which it first bounces may be assumed to be normally distributed with a mean of 140 and a standard deviation of 2.5 .

(a) Determine:

(i) $P(X < 145)$; *(3 marks)*

(ii) $P(138 < X < 142)$. *(4 marks)*

(b) Determine, to one decimal place, the maximum height exceeded by 85% of first bounces. *(4 marks)*

(c) Determine the probability that, for a random sample of 4 first bounces, the mean height is greater than 139 cm. *(4 marks)*

**6** For the adult population of the UK, 35 per cent of men and 29 per cent of women do not wear glasses or contact lenses.

(a) Determine the probability that, in a random sample of 40 men:

(i) at most 15 do not wear glasses or contact lenses; *(3 marks)*

(ii) more than 10 but fewer than 20 do not wear glasses or contact lenses. *(3 marks)*

(b) Calculate the probability that, in a random sample of 10 women, exactly 3 do not wear glasses or contact lenses. *(3 marks)*

(c) (i) Calculate the mean and the variance for the number who **do** wear glasses or contact lenses in a random sample of 20 women. *(3 marks)*

(ii) The numbers wearing glasses or contact lenses in 10 groups, each of 20 women, had a mean of 16.5 and a variance of 2.50 .

Comment on the claim that these 10 groups were **not** random samples. *(3 marks)*

**7** Vernon, a service engineer, is expected to carry out a boiler service in one hour.

One hour is subtracted from each of his actual times, and the resulting differences, $x$ minutes, for a random sample of 100 boiler services are summarised in the table.

| Difference | Frequency |
|---|---|
| $-6 \leqslant x < -4$ | 4 |
| $-4 \leqslant x < -2$ | 9 |
| $-2 \leqslant x < 0$ | 13 |
| $0 \leqslant x < 2$ | 27 |
| $2 \leqslant x < 4$ | 21 |
| $4 \leqslant x < 6$ | 15 |
| $6 \leqslant x < 8$ | 7 |
| $8 \leqslant x \leqslant 10$ | 4 |
| **Total** | **100** |

(a) (i) Calculate estimates of the mean and the standard deviation of these differences. *(4 marks)*

(ii) Hence deduce, in minutes, estimates of the mean and the standard deviation of Vernon's actual service times for this sample. *(3 marks)*

(b) (i) Construct an approximate 98% confidence interval for the mean time taken by Vernon to carry out a boiler service. *(4 marks)*

(ii) Give a reason why this confidence interval is approximate rather than exact. *(1 mark)*

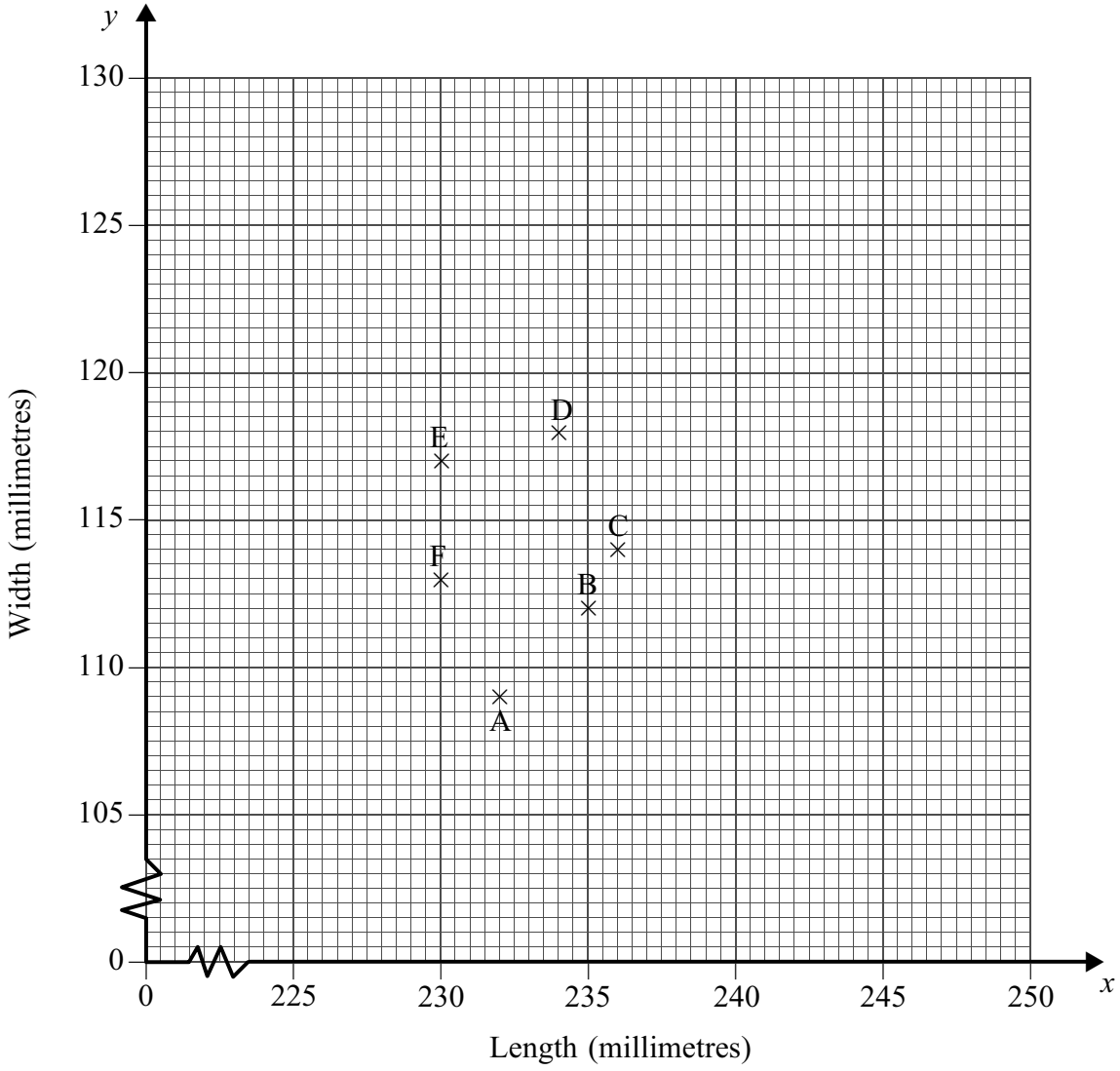
(c) Vernon claims that, more often than not, a boiler service takes more than an hour and that, on average, a boiler service takes much longer than an hour.

Comment, with a justification, on **each** of these claims. *(2 marks)*

**END OF QUESTIONS**

**Figure 1 (for use in Question 3)**

**Decorative Plaques**



**Practice 7**

Leave blank

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

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

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

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

**(3)**

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

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

**(3)**

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

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

**(2)**

(d) Comment on the usefulness of this test.

**(1)**

Leave blank

**Question 1 continued**

**Question 1 continued**

Leave blank

**Question 1 continued**

Leave blank

**Q1**

**(Total 9 marks)**

Leave blank

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

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

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

For the Balmoral Hotel,

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

**(1)**

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

**(3)**

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

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

**(4)**

One measure of skewness is found using

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

(d) Evaluate this measure for the Balmoral Hotel.

**(2)**

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

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

**(3)**

Leave blank

**Question 2 continued**

Leave blank

**Question 2 continued**

Leave blank

**Question 2 continued**

**(Total 13 marks)**

**Q2**

Leave blank

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

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

Given that E($X$) = 0.55, find

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

**(5)**

(b) Var($X$),

**(4)**

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

**(2)**

Leave blank

**Question 3 continued**

**Question 3 continued**

Leave blank

Leave blank

**Question 3 continued**

**(Total 11 marks)**

**Q3**

Leave blank

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

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

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

**(4)**

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

**(3)**

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

**(2)**

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

**(1)**

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

**(4)**

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

**(1)**

Leave blank

**Question 4 continued**

**Question 4 continued**

Leave blank

Leave blank

**Question 4 continued**

**(Total 15 marks)**

**Q4**

Leave blank

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

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

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

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

**(4)**

Leave blank

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

**(2)**

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

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

**(2)**

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

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

**(2)**

**Question 5 continued**

Leave blank

Leave blank

**Question 5 continued**

**(Total 10 marks)**

**Q5**

Leave blank

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

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

where $k$ is a positive integer.

(a) Find $k$.

**(2)**

(b) Find the probability distribution of $X$.

**(3)**

Leave blank

**Question 6 continued**

**Q6**

**(Total 5 marks)**

Leave blank

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

(a) Find P($X$>53).

**(3)**

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

**(5)**

Three bags are selected at random.

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

**(4)**

**Question 7 continued**

Leave blank

Leave blank

**Question 7 continued**

Leave blank

**Question 7 continued**

**(Total 12 marks)**

Q7

**TOTAL FOR PAPER: 75 MARKS**

**END**