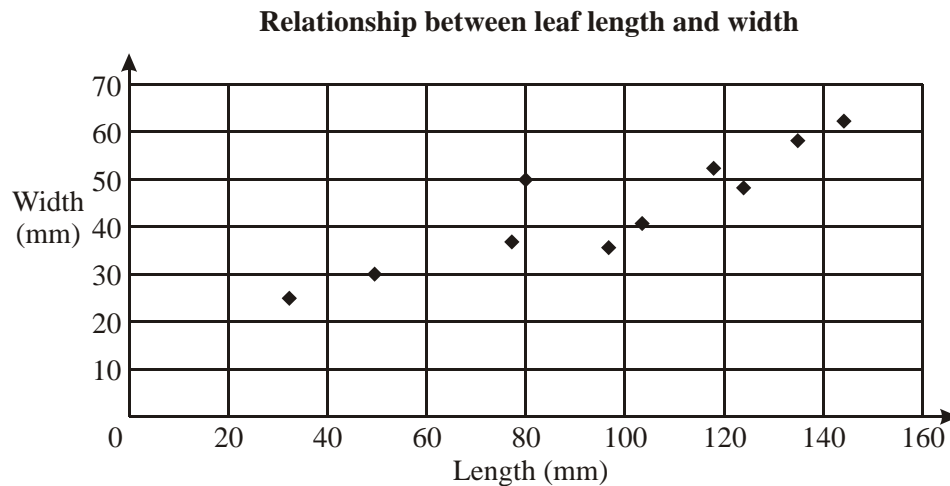


1. The length and width of 10 leaves are shown on the scatter diagram below.



- (a) Plot the point $M(97, 43)$ which represents the mean length and the mean width.
- (b) Draw a suitable line of best fit.
- (c) Write a sentence describing the relationship between leaf length and leaf width for this sample.

Working:

Answer:

- (c)
-

(Total 4 marks)

2. The heights and weights of 10 students selected at random are shown in the table below.

Student	1	2	3	4	5	6	7	8	9	10
Height x cm	155	161	173	150	182	165	170	185	175	145
Weight y kg	50	75	80	46	81	79	64	92	74	108

- (a) Plot this information on a scatter graph. Use a scale of 1 cm to represent 20 cm on the x -axis and 1 cm to represent 10 kg on the y -axis.

(4)

- (b) Calculate the mean height. (1)
- (c) Calculate the mean weight. (1)
- (d) It is given that $S_{xy} = 44.31$.
- (i) By first calculating the standard deviation of the heights, correct to two decimal places, show that the gradient of the line of regression of y on x is 0.276.
- (ii) Calculate the equation of the line of best fit.
- (iii) Draw the line of best fit on your graph. (6)
- (e) Use your line to estimate
- (i) the weight of a student of height 190 cm;
- (ii) the height of a student of weight 72 kg. (2)
- (f) It is decided to remove the data for student number 10 from all calculations. Explain **briefly** what effect this will have on the line of best fit. (1)
- (Total 15 marks)**

3. A survey was conducted in a company to determine whether position in upper management was independent of gender. The results of this survey are tabulated below.

	Managers	Junior executives	Senior executives	Totals
Male	95	130	75	300
Female	65	110	25	200
Totals	160	240	100	500

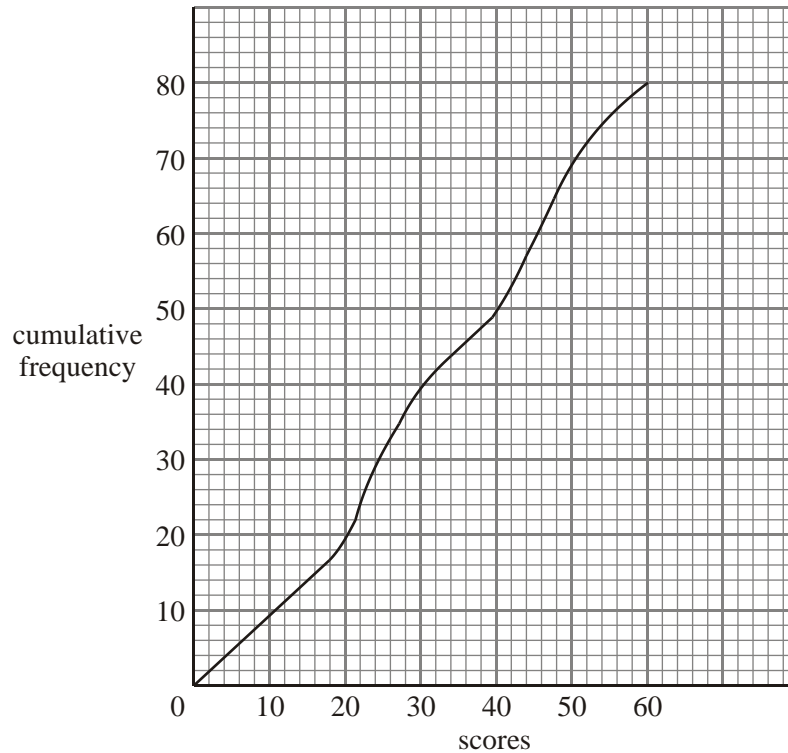
The table below shows the expected number of males and females at each level, if they were represented proportionally to the total numbers of males and females employed.

	Managers	Junior executives	Senior executives	Totals
Male	a	c	60	300
Female	b	d	40	200
Totals	160	240	100	500

- (a) (i) Show that the expected number of **Male Managers** (a) is 96.
- (ii) Hence find the values of b , c and d . (5)

- (b) (i) Write a suitable null hypothesis for this data.
- (ii) Write a suitable alternate hypothesis for this data. (2)
- (c) (i) Perform a chi-squared test of independence for this data to show the value of χ^2 is 12.8 to 3 significant figures.
- (ii) Calculate the number of degrees of freedom, and write down the critical value of χ^2 at the 5% significance level.
- (iii) What conclusion can be drawn regarding gender and position in upper management? (6)
- (Total 13 marks)**

4. The cumulative frequency graph below shows the examination scores of 80 students.



From the graph find

- (a) the median value;
- (b) the interquartile range;
- (c) the 35th percentile;

(d) the percentage of students who scored 50 or above on this examination.

Working:

Answers:

- (a)
- (b)
- (c)
- (d)

(Total 8 marks)

5. For the set of {8, 4, 2, 10, 2, 5, 9, 12, 2, 6}

- (a) calculate the mean;
- (b) find the mode;
- (c) find the median.

Working:

Answers:

- (a)
- (b)
- (c)

(Total 4 marks)

6. For his Mathematical Studies Project a student gave his classmates a questionnaire to fill out. The results for the question on the gender of the student and specific subjects taken by the student are given in the table below, which is a 2×3 contingency table of **observed** values.

	History	Biology	French	
Female	22	20	18	(60)
Male	20	11	9	(40)
	(42)	(31)	(27)	

The following is the table for the **expected** values.

	History	Biology	French
Female	p	18.6	16.2
Male	q	r	10.8

- (a) Calculate the values of p , q and r .

(3)

The chi-squared test is used to determine if the choice of subject is independent of gender, at the 5% level of significance.

- (b) (i) State a suitable null hypothesis H_0 .
- (ii) Show that the number of degrees of freedom is two.
- (iii) Write down the critical value of chi-squared at the 5% level of significance.

(3)

- (c) The calculated value of chi-squared is 1.78. Do you accept H_0 ? Explain your answer.

(2)

(Total 8 marks)

7. Ten students were asked for their average grade at the end of their last year of high school and their average grade at the end of their last year at university. The results were put into a table as follows:

Student	High School grade, x	University grade, y
1	90	3.2
2	75	2.6
3	80	3.0
4	70	1.6
5	95	3.8
6	85	3.1
7	90	3.8
8	70	2.8
9	95	3.0
10	85	3.5
Total	835	30.4

- (a) Given that $s_x = 8.96$, $s_y = 0.610$ and $s_{xy} = 4.16$, find the correlation coefficient r , giving your answer to two decimal places. (2)
- (b) Describe the correlation between the high school grades and the university grades. (2)
- (c) Find the equation of the regression line for y on x in the form $y = ax + b$. (2)
- (Total 6 marks)**

8. Ten students were given two tests, one on Mathematics and one on English. The table shows the results of the tests for each of the ten students.

Student	A	B	C	D	E	F	G	H	I	J
Mathematics (x)	8.6	13.4	12.8	9.3	1.3	9.4	13.1	4.9	13.5	9.6
English (y)	33	51	30	48	12	23	46	18	36	50

- (a) Given s_{xy} (the covariance) is 35.85, calculate, correct to two decimal places, the product moment correlation coefficient (r). (6)
- (b) Use your result from part (a) to comment on the statement:

'Those who do well in Mathematics also do well in English.'

(2)
(Total 8 marks)

9. The table shows the number of children in 50 families.

Number of children	Frequency	Cumulative frequency
1	3	3
2	m	22
3	12	34
4	p	q
5	5	48
6	2	50
	T	

- (a) Write down the value of T .

(b) Find the values of m , p and q .

Working:

Answers:

(a)

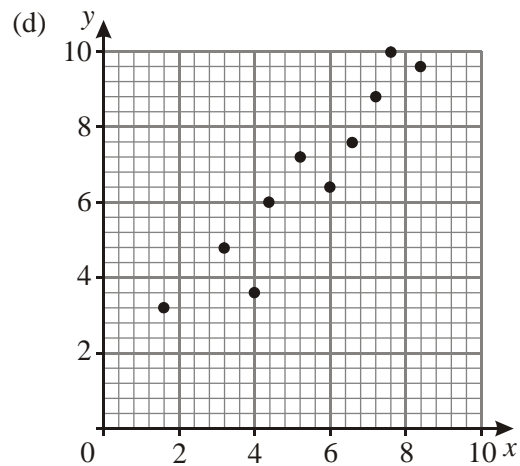
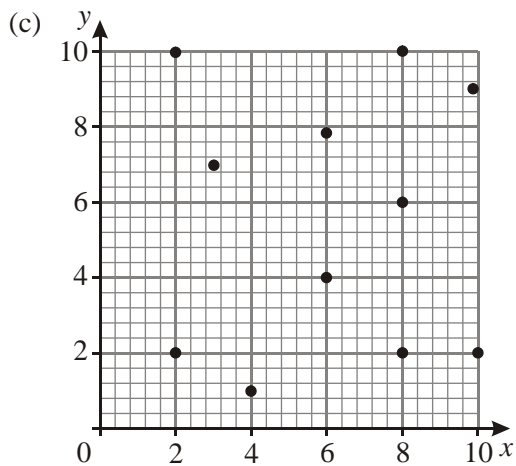
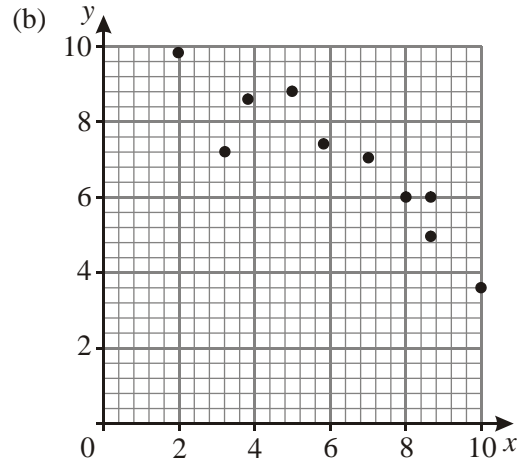
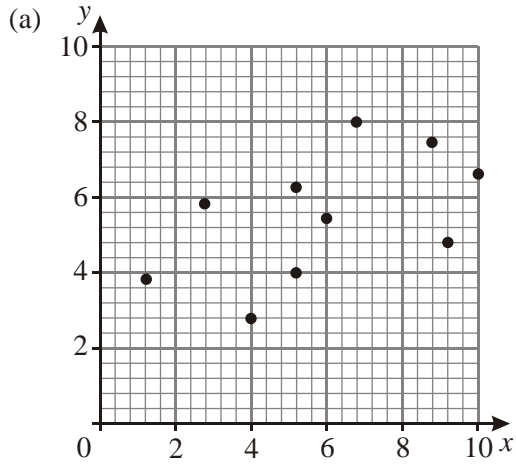
(b)

(Total 4 marks)

10. Statements I, II, III, IV and V represent descriptions of the correlation between two variables.

- I High positive linear correlation
- II Low positive linear correlation
- III No correlation
- IV Low negative linear correlation
- V High negative linear correlation

Which statement **best** represents the relationship between the two variables shown in each of the scatter diagrams below.



Answers:

- (a)
- (b)
- (c)
- (d)

(Total 4 marks)

11. The cumulative frequency table below shows the ages of 200 students at a college.

Age	Number of Students	Cumulative Frequency
17	3	3
18	72	75
19	62	137
20	31	m
21	12	180
22	9	189
23–25	5	194
> 25	6	n

- (a) What are the values of m and n ?
- (b) How many students are younger than 20?
- (c) Find the value in years of the lower quartile.

Working:

Answers:

- (a)
- (b)
- (c)

(Total 8 marks)

12. David looked at a passage from a book. He recorded the number of words in each sentence as shown in the following frequency table.

Class interval (number of words)	Frequency f
1–5	16
6–10	28
11–15	26
16–20	14
21–25	10
26–30	3
31–35	1
36–40	0
41–45	2

- (a) Find the class interval in which the median lies.
- (b) Estimate, **correct to the nearest whole number**, the mean number of words in a sentence.

Working:

Answers:

(a)

(b)

(Total 4 marks)

13. Several candy bars were purchased and the following table shows the weight and the cost of each bar.

	Yummy	Chox	Marz	Twin	Chunx	Lite	BigC	Bite
Weight (g)	60	85	80	65	95	50	100	45
Cost (Euros)	1.10	1.50	1.40	1.20	1.80	1.00	1.70	0.90

- (a) Given that $s_x = 19.2$, $s_y = 0.307$ and $s_{xy} = 5.81$, find the correlation coefficient, r , giving your answer correct to 3 decimal places.

(2)

- (b) Describe the correlation between the weight of a candy bar and its cost. (1)
- (c) Calculate the equation of the regression line for y on x . (3)
- (d) Use your equation to estimate the cost of a candy bar weighing 109 g. (2)
- (Total 8 marks)**

- 14.** Members of a certain club are required to register for one of three games, billiards, snooker or darts.

The number of club members of each gender choosing each game in a particular year is shown in the table below.

	Billiards	Snooker	Darts
Male	39	16	8
Female	21	14	17

- (a) Use a χ^2 (Chi-squared) test at the 5% significance level to test whether choice of games is independent of gender. State clearly the null and alternative hypotheses tested, the expected values, and the number of degrees of freedom used. (13)

The following year the choice of games was widened and the figures for that year are as follows:

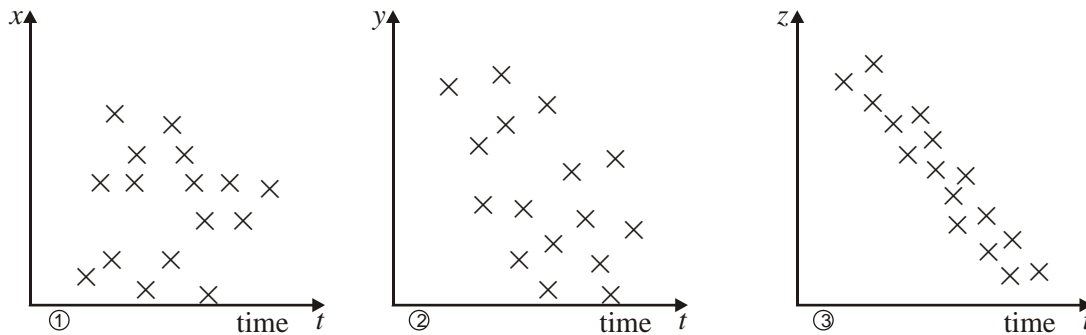
	Billiards	Snooker	Darts	Fencing
Male	4	15	8	10
Female	10	21	17	37

- (b) If the χ^2 test were applied to this new set of data,
- (i) why would it be necessary to combine billiards with another game?
- (ii) which other game would you combine with billiards and why? (2)

A club member is to be selected at random.

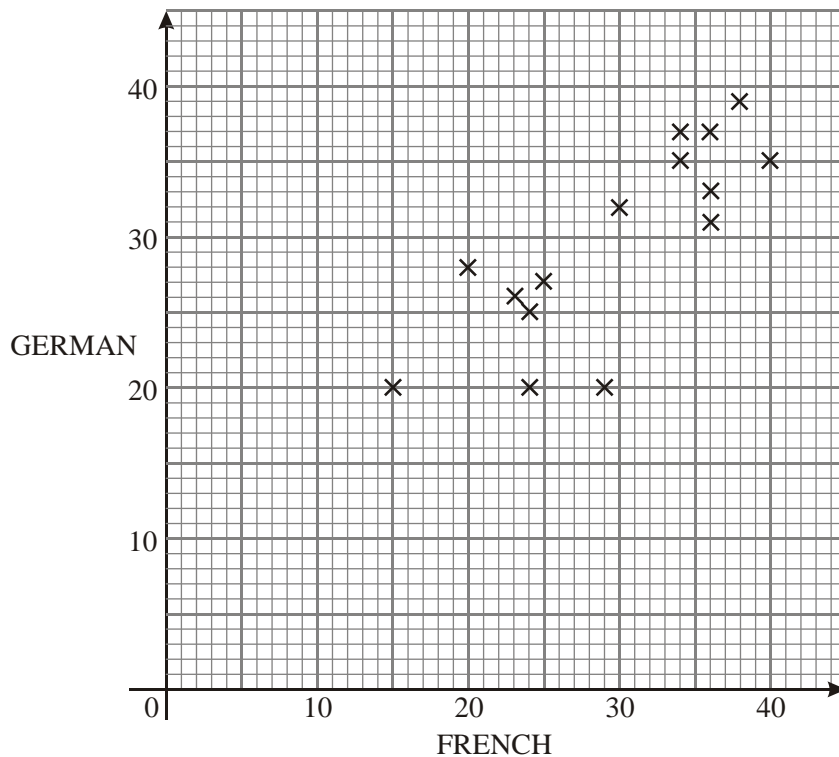
- (c) What is the probability that the club member selected is a
- (i) female who chose billiards or snooker?
- (ii) male or female who chose darts or fencing? (2)
- (Total 17 marks)**

15. The sketches below represent scatter diagrams for the way in which variables x , y and z change over time, t , in a given chemical experiment. They are labelled 1, 2 and 3.



- (a) State which of the diagrams indicate that the pair of variables
- is not correlated. (1)
 - shows strong linear correlation. (1)
- (b) A student is given a piece of paper with five numbers written on it. She is told that three of these numbers are the product moment correlation coefficients for the three pairs of variables shown above. The five numbers are
- $0.9, -0.85, -0.20, 0.04, 1.60$
- For each sketch above state which of these five numbers is the most appropriate value for the correlation coefficient. (3)
 - For the two remaining numbers, state why you reject them for this experiment. (2)
- (c) Another variable, w , over time, t , gave the following information
- $$t = 124 \quad w = 250 \quad s_t = 6.08 \quad s_w = 10.50 \quad s_{tw} = 55.00$$
- for 20 data points.
- Calculate
- the product moment correlation coefficient for this data. (2)
 - the equation of the regression line of w on t in the form $w = at + b$. (5)
- (Total 14 marks)**

16. The diagram below shows the marks scored by pupils in a French test and a German test. The mean score on the French test is 29 marks and on the German test is 31 marks.



- Describe the relationship between the marks scored in the two tests.
- On the graph mark the point M which represents the mean of the distribution.
- Draw a suitable line of best fit.
- Idris scored 32 marks on the French test. Use your graph to estimate the mark Idris scored on the German test.

Working:

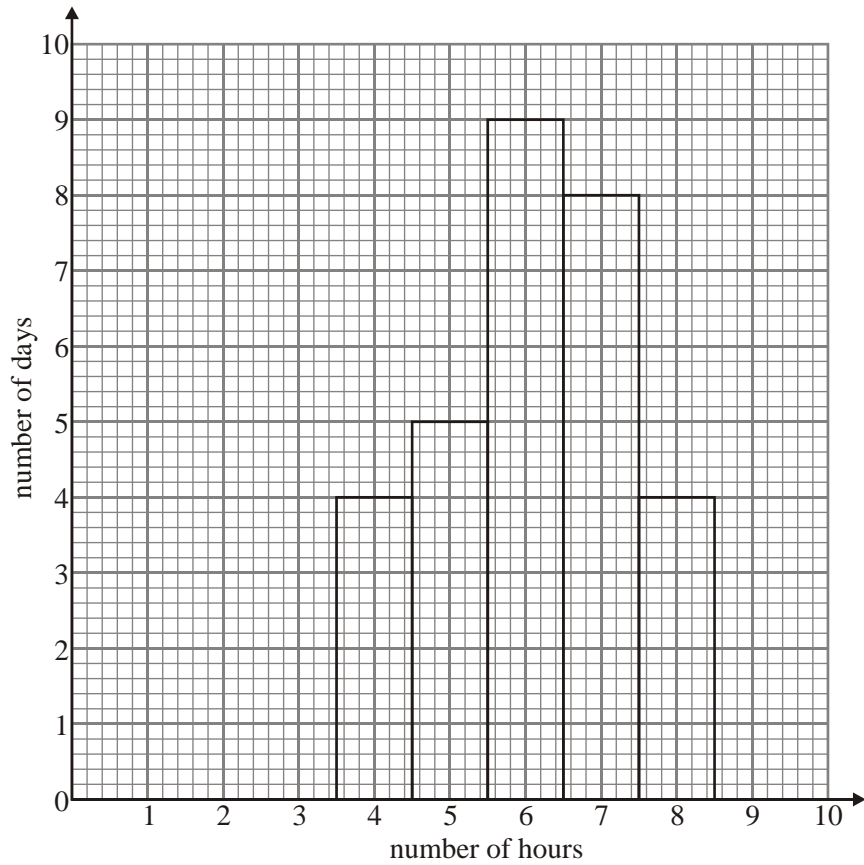
Answers:

(a)

(b)

(Total 4 marks)

17. The number of hours that a professional footballer trains each day in the month of June is represented in the following histogram.



- (a) Write down the modal number of hours trained each day.
 (b) Calculate the mean number of hours he trains each day.

Working:

Answers:

- (a)
 (b)

(Total 8 marks)

18. The following table gives the heights and weights of five sixteen-year-old boys.

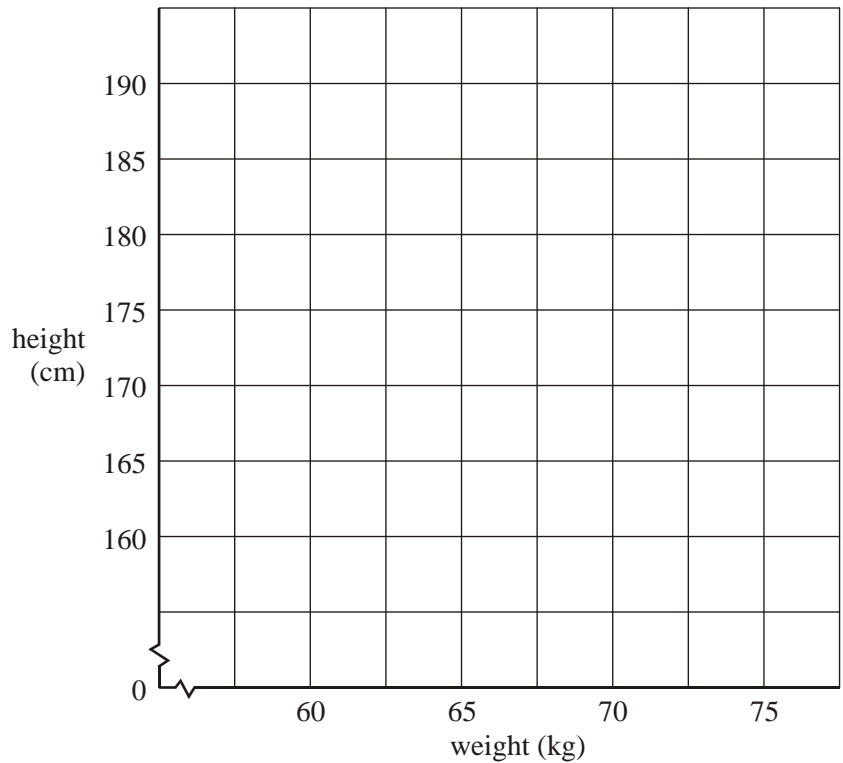
Name	Height	Weight
Blake	182 cm	73 kg
Jorge	173 cm	68 kg
Chin	162 cm	60 kg
Ravi	178 cm	66 kg
Derek	190 cm	75 kg

(a) Find

(i) the mean height;

(ii) the mean weight.

(b) Plot the above data on the grid below and draw the line of best fit.



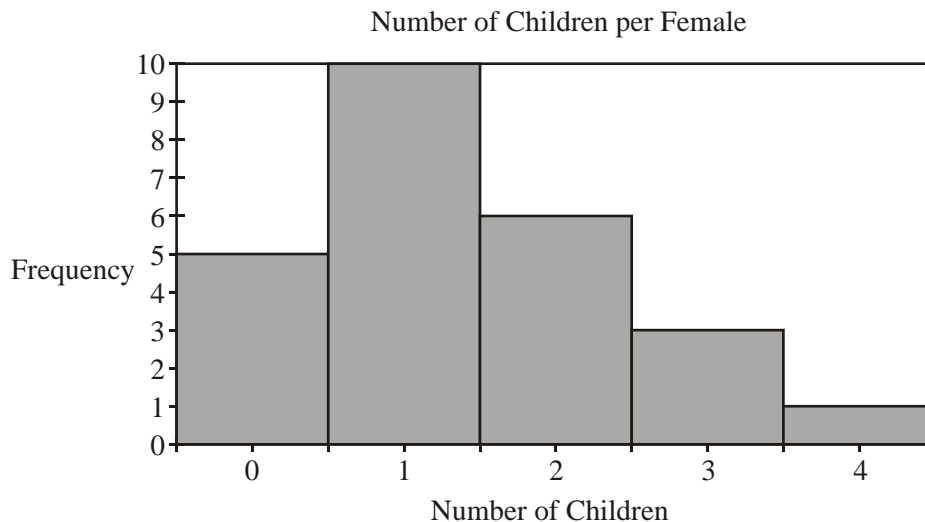
Working:

Answers:

- (a) (i)
- (ii)

(Total 4 marks)

19. A group of 25 females were asked how many children they each had. The results are shown in the histogram below.



- (a) Show that the mean number of children per female is 1.4. (2)
- (b) Show clearly that the standard deviation for this data is approximately 1.06. (3)
- (c) Another group of 25 females was surveyed and it was found that the mean number of children per female was 2.4 and the standard deviation was 2. Use the results from parts (a) and (b) to describe the differences between the number of children the two groups of females have. (2)
- (d) A female is selected at random from the first group. What is the probability that she has more than two children? (2)
- (e) Two females are selected at random from the first group. What is the probability that
- (i) both females have more than two children? (2)
 - (ii) only one of the females has more than two children? (2)
 - (iii) the second female selected has two children given that the first female selected had no children? (1)

(Total 15 marks)

20. The *Type Fast* secretarial training agency has a new computer software spreadsheet package. The agency investigates the number of hours it takes people of varying ages to reach a level of proficiency using this package. Fifteen individuals are tested and the results are summarised in the table below.

Age (x)	32	40	21	45	24	19	17	21	27	54	33	37	23	45	18
Time (in hours) (y)	10	12	8	15	7	8	6	9	11	16	<i>t</i>	13	9	17	5

- (a) (i) Given that $S_y = 3.5$ and $S_{xy} = 36.7$, calculate the product-moment correlation coefficient r for this data. (4)
- (ii) What does the value of the correlation coefficient suggest about the relationship between the two variables? (1)
- (b) Given that the mean time taken was 10.6 hours, write the equation of the regression line for y on x in the form $y = ax + b$. (3)
- (c) Use your equation for the regression line to predict
- (i) the time that it would take a 33 year old person to reach proficiency, giving your answer correct to the nearest hour; (2)
- (ii) the age of a person who would take 8 hours to reach proficiency, giving your answer correct to the nearest year. (2)

(Total 12 marks)