Bivariate Statistics Review Packet

1. The same 100 students are also asked how many meals on average they have per day.
    The data collected is organized in the following table.

	3 or fewer meals per day	4 or 5 meals per day	More than 5 meals per day	Total
Male	15	25	15	55
Female	12	20	13	45
Total	27	45	28	100

A χ^2 test is carried out at the 5 % level of significance.

(a) Write down the null hypothesis, H_0, for this test.

(1)

(b) Write down the number of degrees of freedom for this test.

(1)

(c) Write down the critical value for this test.

(1)

(d) Show that the expected number of females that have more than 5 meals per day is 13, correct to the nearest integer.

(2)

(e) Use your graphic display calculator to find the χ^2_{calc} for this data.

(2)

(f) Decide whether H_0 must be accepted. Justify your answer.

(2)
(Total 9 marks)

2. In a mountain region there appears to be a relationship between the number of trees growing in the region and the depth of snow in winter. A set of 10 areas was chosen, and in each area the number of trees was counted and the depth of snow measured.
The results are given in the table below.

Number of trees (x)	Depth of snow in cm (y)
45	30
75	50
66	40
27	25
44	30
28	5
60	35
35	20
73	45
47	25

(a) Use your graphic display calculator to find

    (i) the mean number of trees;

    (ii) the standard deviation of the number of trees;

    (iii) the mean depth of snow;

    (iv) the standard deviation of the depth of snow.

(4)

The covariance, $S_{xy} = 188.5$.

(b) Write down the product-moment correlation coefficient, r.

(2)

(c) Write down the equation of the regression line of y on x.

(2)

(d) If the number of trees in an area is 55, estimate the depth of snow.

(2)

(e) (i) Use the equation of the regression line to estimate the depth of snow in an area with 100 trees.

    (ii) Decide whether the answer in (e)(i) is a valid estimate of the depth of snow in the area. Give a reason for your answer.

(3)
(Total 13 marks)

3. In a study on 100 students there seemed to be a difference between males and females in their choice of favourite car colour. The results are given in the table below.

A χ^2 test was conducted.

	Blue	Red	Green
Males	14	6	8
Females	31	24	17

(a) Write down the total number of male students.

(1)

(b) Show that the expected frequency for males, whose favourite car colour is blue, is 12.6.

(2)

The calculated value of χ^2 is 1.367.

(c) (i) Write down the null hypothesis for this test.

(ii) Write down the number of degrees of freedom.

(iii) Write down the critical value of χ^2 at the 5 % significance level.

(iv) Determine whether the null hypothesis should be accepted. Give a reason for your answer.

(5)
(Total 8 marks)

4. Jorge conducted a survey of 200 drivers. He asked two questions:

How long have you had your driving licence?
Do you wear a seat belt when driving?

The replies are summarized in the table below.

	Wear a seat belt	**Do not wear a seat belt**
Licence less than 2 years	38	42
Licence between 2 and 15 years	30	45
Licence more than 15 years	30	15

(a) Jorge applies a χ^2 test at the 5 % level to investigate whether wearing a seat belt is associated with the time a driver has had their licence.

   (i) Write down the null hypothesis, H_0.

   (ii) Write down the number of degrees of freedom.

   (iii) Show that the expected number of drivers that wear a seat belt and have had their driving licence for more than 15 years is 22, correct to the nearest whole number.

   (iv) Write down the χ^2 test statistic for this data.

   (v) Does Jorge accept H_0? Give a reason for your answer.

(8)

(b) Consider the 200 drivers surveyed. One driver is chosen at random.
Calculate the probability that

   (i) this driver wears a seat belt;

   (ii) the driver does not wear a seat belt, **given that** the driver has held a licence for more than 15 years.

(4)

(c) Two drivers are chosen at random. Calculate the probability that

   (i) both wear a seat belt.

   (ii) at least one wears a seat belt.

(6)
(Total 18 marks)

5. The marks obtained by 8 candidates in Physics and Chemistry tests are given below.

Physics (x)	6	8	10	11	10	5	4	12
Chemistry (y)	8	11	14	13	11	7	5	15

(a) Write down the product moment correlation coefficient, r.

(1)

(b) Write down, in the form $y = mx + c$, the equation of the regression line y on x for the 8 candidates.

(2)

A ninth candidate obtained a score of 7 in the Physics test but was absent for the Chemistry test.

(c) Use your answer to (b) to estimate the score he would have obtained on the Chemistry test.

(2)

(d) Give a reason why it is valid to use this regression line to estimate the score on the Chemistry test.

(1)
(Total 6 marks)

6. Manuel conducts a survey on a random sample of 751 people to see which television programme type they watch most from the following: Drama, Comedy, Film, News. The results are as follows.

	Drama	Comedy	Film	News
Males under 25	22	65	90	35
Males 25 and over	36	54	67	17
Females under 25	22	59	82	15
Females 25 and over	64	39	38	46

Manuel decides to ignore the ages and to test at the 5% level of significance whether the most watched programme type is independent of **gender**.

(a) Draw a table with 2 rows and 4 columns of data so that Manuel can perform a chi-squared test.

(3)

(b) State Manuel's null hypothesis and alternative hypothesis.

(1)

(c) Find the expected frequency for the number of females who had "Comedy" as their most-watched programme type. Give your answer to the nearest whole number.

(2)

(d) Using your graphic display calculator, or otherwise, find the chi-squared statistic for Manuel's data.

(3)

(e) (i) State the number of degrees of freedom available for this calculation.

(ii) State the critical value for Manuel's test.

(iii) State his conclusion.

(3)
(Total 12 marks)

7. Tania wishes to see whether there is any correlation between a person's age and the number of objects on a tray which could be remembered after looking at them for a certain time.

She obtains the following table of results.

Age (x years)	15	21	36	40	44	55
Number of objects remembered (y)	17	20	15	16	17	12

(a)   Use your graphic display calculator to find the equation of the regression line of y on x.

(2)

(b)   Use your equation to estimate the number of objects remembered by a person aged 28 years.

(1)

(c)   Use your graphic display calculator to find the correlation coefficient r.

(1)

(d)   Comment on your value for r.

(2)
(Total 6 marks)

8. In an experiment a vertical spring was fixed at its upper end. It was stretched by hanging different weights on its lower end. The length of the spring was then measured. The following readings were obtained.

Load (kg) x	0	1	2	3	4	5	6	7	8
Length (cm) y	23.5	25	26.5	27	28.5	31.5	34.5	36	37.5

(a) Plot these pairs of values on a scatter diagram taking 1 cm to represent 1 kg on the horizontal axis and 1 cm to represent 2 cm on the vertical axis.

(4)

(b) (i) Write down the mean value of the load ($\bar{x}$).

(1)

(ii) Write down the standard deviation of the load.

(1)

(iii) Write down the mean value of the length ($\bar{y}$).

(1)

(iv) Write down the standard deviation of the length.

(1)

(c) Plot the mean point ($\bar{x}$, $\bar{y}$) on the scatter diagram. Name it L.

(1)

It is given that the covariance S_{xy} is 12.17.

(d) (i) Write down the correlation coefficient, r, for these readings.

(1)

(ii) Comment on this result.

(2)

(e) Find the equation of the regression line of y on x.

(2)

(f) Draw the line of regression on the scatter diagram.

(2)

(g) (i) Using your diagram or otherwise, estimate the length of the spring when a load of 5.4 kg is applied.

(1)

(ii) Malcolm uses the equation to claim that a weight of 30 kg would result in a length of 62.8 cm. Comment on his claim.

(1)
(Total 18 marks)

9. Oral tests are conducted by three examiners A, B and C separately. The results of the examination are classified as Credit, Pass or Fail. A χ^2 test is applied to the data collected in order to test whether or not the examiners differ in their standard of awards.

(a) State the null hypothesis, H_0, for this data.

(b) Write down the number of degrees of freedom.

Of the 135 students who sit the exam, 30 get Credit and 45 are tested by examiner A.

(c) Calculate the expected number of students who get a Credit and are tested by examiner A.

Using a 5% level of significance, the p-value is found to be 0.0327 correct to 3 s.f.

(d) State whether H_0 should be accepted. Justify your answer.

(Total 6 marks)

10. Tom performs a chi-squared test to see if there is any association between the time to prepare for a penalty kick (short time, medium time and long time) and the outcome (scores a goal, doesn't score a goal). Tom performs this test at the 10% level.

(a) Write down the null hypothesis.

(b) Find the number of degrees of freedom for this test.

(c) The p-value for this test is 0.073. What conclusion can Tom make? Justify your answer.

(Total 6 marks)

11. At the circus a clown is swinging from an elastic rope. A student decides to investigate the motion of the clown. The results can be shown on the graph of the function $f(x) = (0.8^x)(5 \sin 100x)$, where x is the horizontal distance in metres.

(a) Sketch the graph of $f(x)$ for $0 \leq x \leq 10$ and $-3 \leq f(x) \leq 5$.

(5)

(b) Find the coordinates of the first local maximum point.

(2)

(c) Find the coordinates of one point where the curve cuts the y-axis.

(1)

Another clown is fired from a cannon. The clown passes through the points given in the table below:

Horizontal distance (x)	Vertical distance (y)
0.00341	0.0102
0.0238	0.0714
0.563	1.69
1.92	5.76
3.40	10.2

(d) Find the correlation coefficient, r, and comment on the value for r.

(3)

(e) Write down the equation of the regression line of y on x.

(2)

(f) Sketch this line on the graph of $f(x)$ in part (a).

(1)

(g) Find the coordinates of one of the points where this line cuts the curve.

(2)
(Total 16 marks)

12. A researcher consulted 500 men and women to see if the colour of the car they drove was independent of gender. The colours were red, green, blue, black and silver. A χ^2 test was conducted at the 5% significance level and the value found to be 8.73.

(a) Write down the null hypothesis.

(b) Find the number of degrees of freedom for this test.

(c) Write down the critical value for this test.

(d) Is car colour independent of gender? Give a clear reason for your answer

(Total 6 marks)

13. The following table gives the amount of fuel in a car's fuel tank, and the number of kilometres travelled after filling the tank.

Distance travelled (km)	0	220	276	500	680	850
Amount of fuel in tank (litres)	55	43	30	24	10	6

(a) On the scatter diagram below, plot the remaining points.
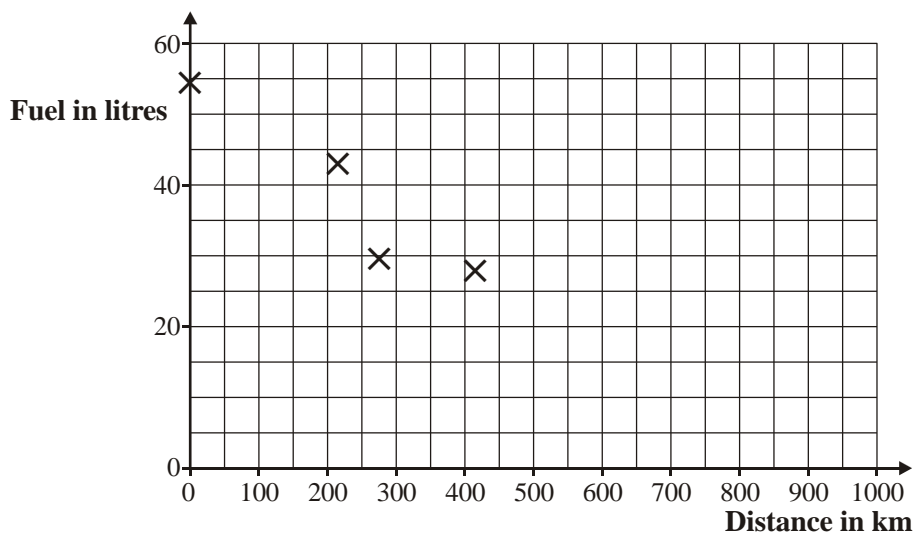
The mean distance travelled is 421 km ($\bar{x}$), and the mean amount of fuel in the tank is 28 litres ($\bar{y}$). This point is plotted on the scatter diagram.

(b) Sketch the line of best fit.

A car travelled 350 km.

(c) Use your line of best fit to estimate the amount of fuel left in the tank.

(Total 6 marks)

14. It is decided to take a random sample of 10 students to see if there is any linear relationship between height and shoe size. The results are given in the table below.

Height (cm) (x)	Shoe size (y)
175	8
160	9
180	8
155	7
178	10
159	8
166	9
185	11
189	10
173	9

(a) Write down the equation of the regression line of shoe size (y) on height (x), giving your answer in the form $y = mx + c$.

(3)

(b) Use your equation in part (a) to predict the shoe size of a student who is 162 cm in height.

(2)

(c) Write down the correlation coefficient.

(1)

(d) Describe the correlation between height and shoe size.

(2)
(Total 8 marks)

15. The eye colour and gender of 500 students are noted and the results are indicated in the table below.

	Blue	Brown	Green
Male	18	152	50
Female	40	180	60

It is believed that eye colour is related to gender in a school in Banff. It is decided to test this hypothesis by using a χ^2 test at the 5% level of significance.

(a) Write down the null hypothesis for this experiment.

(1)

(b) Show that the number of degrees of freedom is 2.

(1)

(c) Write down the χ^2 critical value for the degrees of freedom.

(1)

(d) Calculate the χ^2 test statistic for this data.

(2)

(e) Does the evidence suggest that eye colour is related to gender in this school? Give a clear reason for your answer.

(2)
(Total 7 marks)

16. In a competition the number of males and females taking part in different swimming races is given in the table of observed values below.

	Backstroke (100 m)	Freestyle (100 m)	Butterfly (100 m)	Breaststroke (100 m)	Relay (4 × 100 m)
Male	30	90	31	29	20
Female	28	63	20	37	12

The Swimming Committee decides to perform a χ^2 test at the 5% significance level in order to test if the number of entries for the various strokes is related to gender.

(a) State the null hypothesis.

(1)

(b) Write down the number of degrees of freedom.

(1)

(c) Write down the critical value of χ^2.

(1)

The expected values are given in the table below:

	Backstroke (100 m)	Freestyle (100 m)	Butterfly (100 m)	Breaststroke (100 m)	Relay (4 × 100 m)
Male	32	a	28	37	18
Female	26	68	23	b	14

(d) Calculate the values of a and b.

(2)

(e) Calculate the χ^2 value.

(3)

(f) State whether or not you accept your null hypothesis and give a reason for your answer.

(2)
(Total 10 marks)

17. It is thought that the breaststroke time for 200 m depends on the length of the arm of the swimmer.

Eight students swim 200 m breaststroke. Their times (y) in seconds and arm lengths (x) in cm are shown in the table below.

	1	2	3	4	5	6	7	8
Length of arm, x cm	79	74	72	70	77	73	64	69
Breaststroke, y seconds	135.1	135.7	139.3	141.0	132.8	137.0	152.9	144.0

(a) Calculate the mean and standard deviation of x and y.

(4)

(b) Given that $s_{xy} = -24.82$, calculate the correlation coefficient, r.

(2)

(c) Comment on your value for r.

(2)

(d) Calculate the equation of the regression line of y on x.

(3)

(e) Using your regression line, estimate how many seconds it will take a student with an arm length of 75 cm to swim the 200 m breaststroke.

(1)
(Total 12 marks)

18. The following table of observed results gives the number of candidates taking a Mathematics examination classified by gender and grade obtained.

<table>
<tr><td></td><td colspan="5" align="center">**Grade**</td></tr>
<tr><td></td><td></td><td>5, 6 or 7</td><td>3 or 4</td><td>1 or 2</td><td>Total</td></tr>
<tr><td></td><td>Males</td><td>5000</td><td>3400</td><td>600</td><td>9000</td></tr>
<tr><td>**Gender**</td><td>Females</td><td>6000</td><td>4000</td><td>1000</td><td>11000</td></tr>
<tr><td></td><td>Total</td><td>11000</td><td>7400</td><td>1600</td><td>20000</td></tr>
</table>

The question posed is whether gender and grade obtained are independent.

(a) **Show clearly** that the expected number of males achieving a grade of 5, 6 or 7 is 4950.

(2)

(b) A χ^2 test is set up.

(i) State the Null hypothesis.

(1)

(ii) State the number of degrees of freedom.

(1)

(iii) The calculated χ^2 value at the 5% test level is 39.957.
Write down the critical value of χ^2 at the 5% level of significance.

(1)

(iv) What can you say about gender and grade obtained?

(1)
(Total 6 marks)

19. 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.
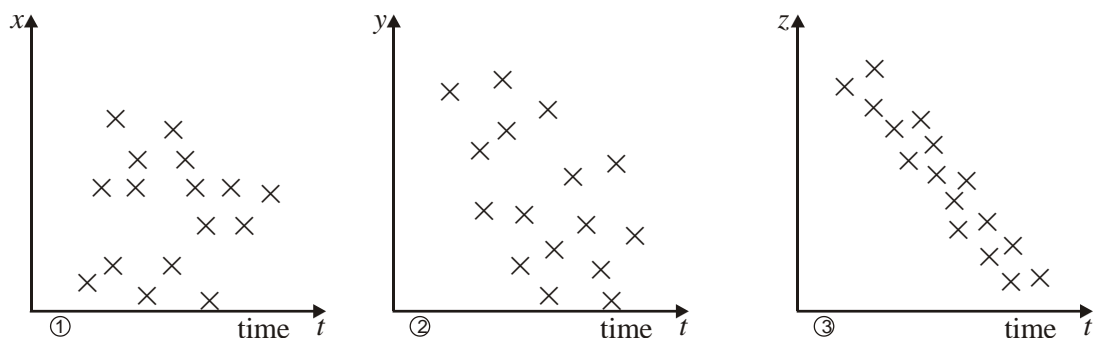
(a)   State which of the diagrams indicate that the pair of variables

(i)   is not correlated;

(1)

(ii)   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

(i)   For each sketch above state which of these five numbers is the most appropriate value for the correlation coefficient.

(3)

(ii)   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

$\sum t = 124$ $\sum w = 250$ $s_t = 6.08$ $s_w = 10.50$ $s_{tw} = 55.00$

for 20 data points.

Calculate

(i)   the product moment correlation coefficient for this data;

(2)

(ii)   the equation of the regression line of w on t in the form $w = at + b$.

(5)
(Total 14 marks)

20. A bag containing 60 sweets is opened. The bag contains sweets of the following colours.

Colour	Frequency
Red	18
Orange	17
Green	10
Purple	9
Blue	6

According to the manufacturer, the various colours should have the following percentages.

Colour	Percentage
Red	35%
Orange	25%
Green	20%
Purple	15%
Blue	5%

(a) Calculate the expected number of sweets of each colour in a bag containing exactly 60 sweets.

(3)

Before you can perform the chi-squared test on this data, it is necessary to combine the data for one of the colours with that of another colour.

(b) Which colour is this and why is this necessary?

(2)

(c) Using the chi-squared test at the 5% significance level, investigate the hypothesis that the sweets in the packet may be regarded as a random sample. Remember to state the null hypothesis, the number of degrees of freedom and the critical value of chi-squared.

(7)
(Total 12 marks)

21. A shopkeeper wanted to investigate whether or not there was a correlation between the prices of food 10 years ago in 1992, with their prices today. He chose 8 everyday items and the prices are given in the table below.

	sugar	milk	eggs	rolls	tea bags	coffee	potatoes	flour
1992 price	$1.44	$0.80	$2.16	$1.80	$0.92	$3.16	$1.32	$1.12
2002 price	$2.20	$1.04	$2.64	$3.00	$1.32	$2.28	$1.92	$1.44

  (a) Calculate the mean and the standard deviation of the prices

     (i) in 1992;

     (ii) in 2002.

(4)

(b) (i) Given that $s_{xy} = 0.3104$, calculate the correlation coefficient.

     (ii) Comment on the relationship between the prices.

(4)

(c) Find the equation of the line of the best fit in the form $y = mx + c$.

(3)

(d) What would you expect to pay now for an item costing $2.60 in 1992?

(1)

(e) Which item would you omit to increase the correlation coefficient?

(2)
(Total 14 marks)

22. 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)

23. 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)

24. 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)

25. 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.
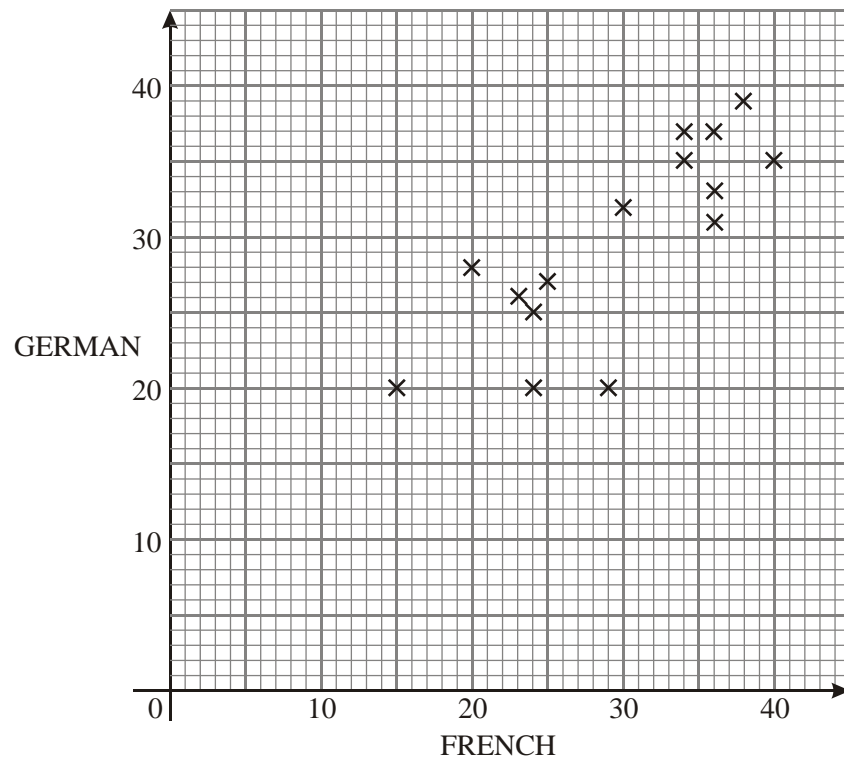
(a)

(b)

(c)

(d)
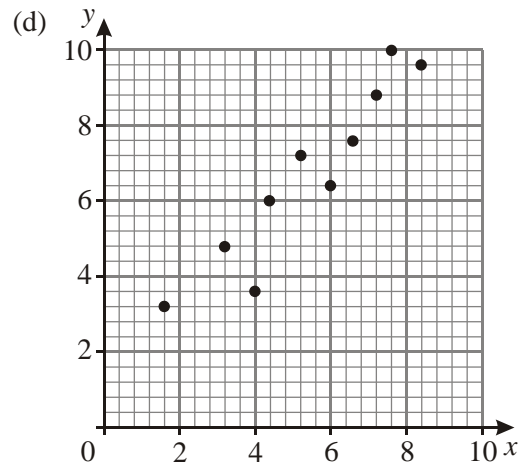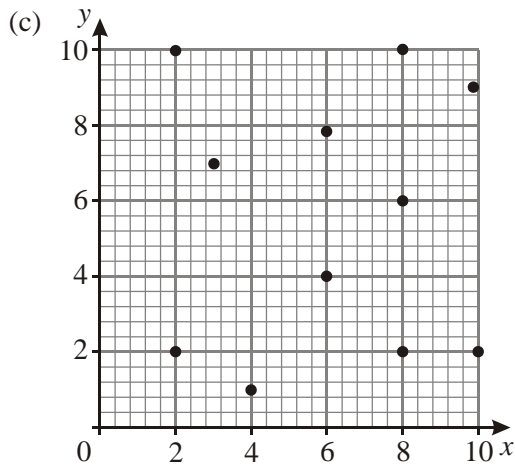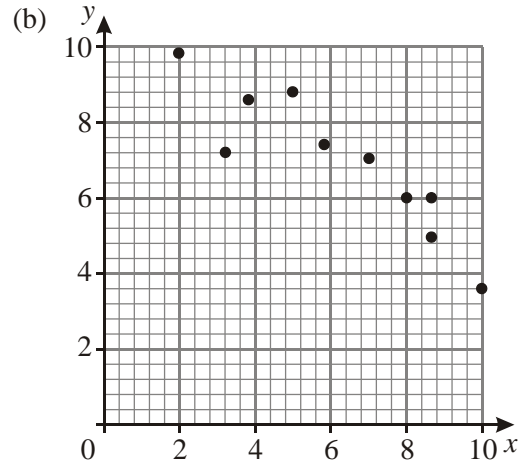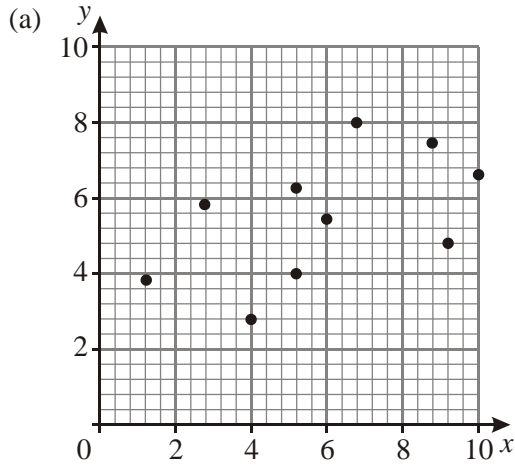
Answers:

(a) ……………………………………………

(b) ……………………………………………

(c) ……………………………………………

(d) ……………………………………………

(Total 4 marks)

26. 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.

(a) Describe the relationship between the marks scored in the two tests.

(b) On the graph mark the point M which represents the mean of the distribution.

(c) Draw a suitable line of best fit.

(d) Idris scored 32 marks on the French test. Use your graph to estimate the mark Idris scored on the German test.

(Total 4 marks)

27. 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)