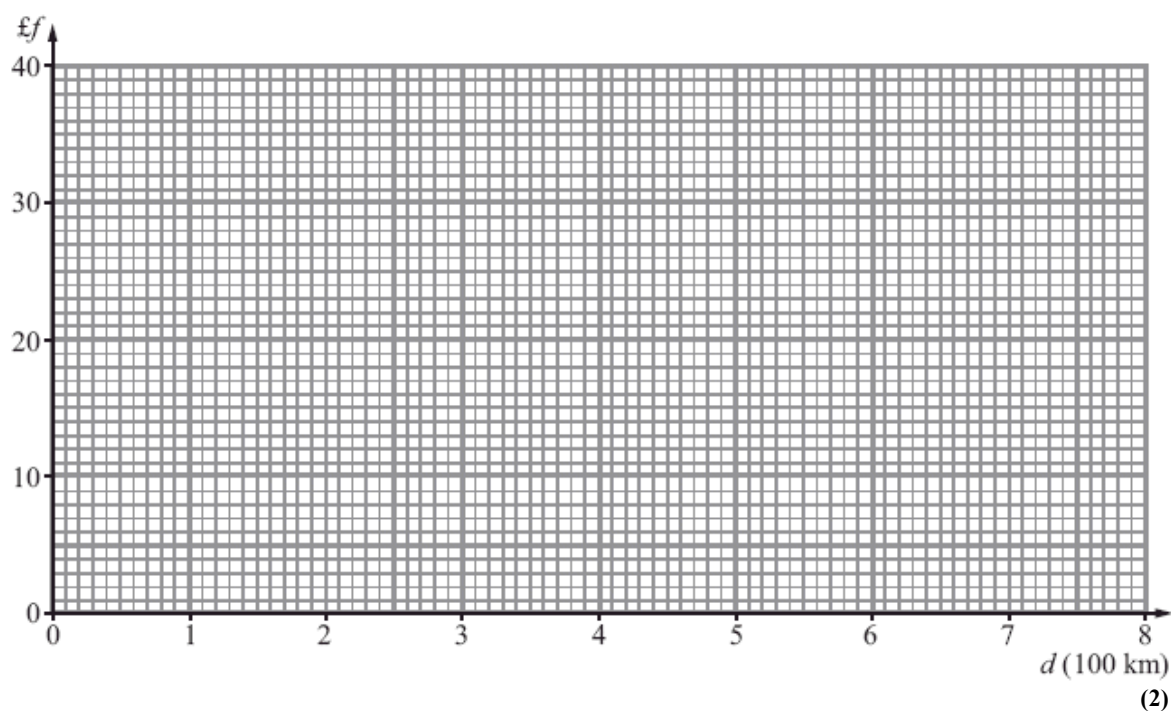


1. A travel agent sells flights to different destinations from *Beerow* airport. The distance d , measured in 100 km, of the destination from the airport and the fare $\pounds f$ are recorded for a random sample of 6 destinations.

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

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

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



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

(1)

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

(4)

- (d) Calculate the equation of the regression line of f on d giving your answer in the form $f = a + bd$. (4)

- (e) Give an interpretation of the value of b . (1)

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

- (f) Find the range of values of t for which the first travel agent is cheaper than the rival. (2)
(Total 14 marks)

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

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

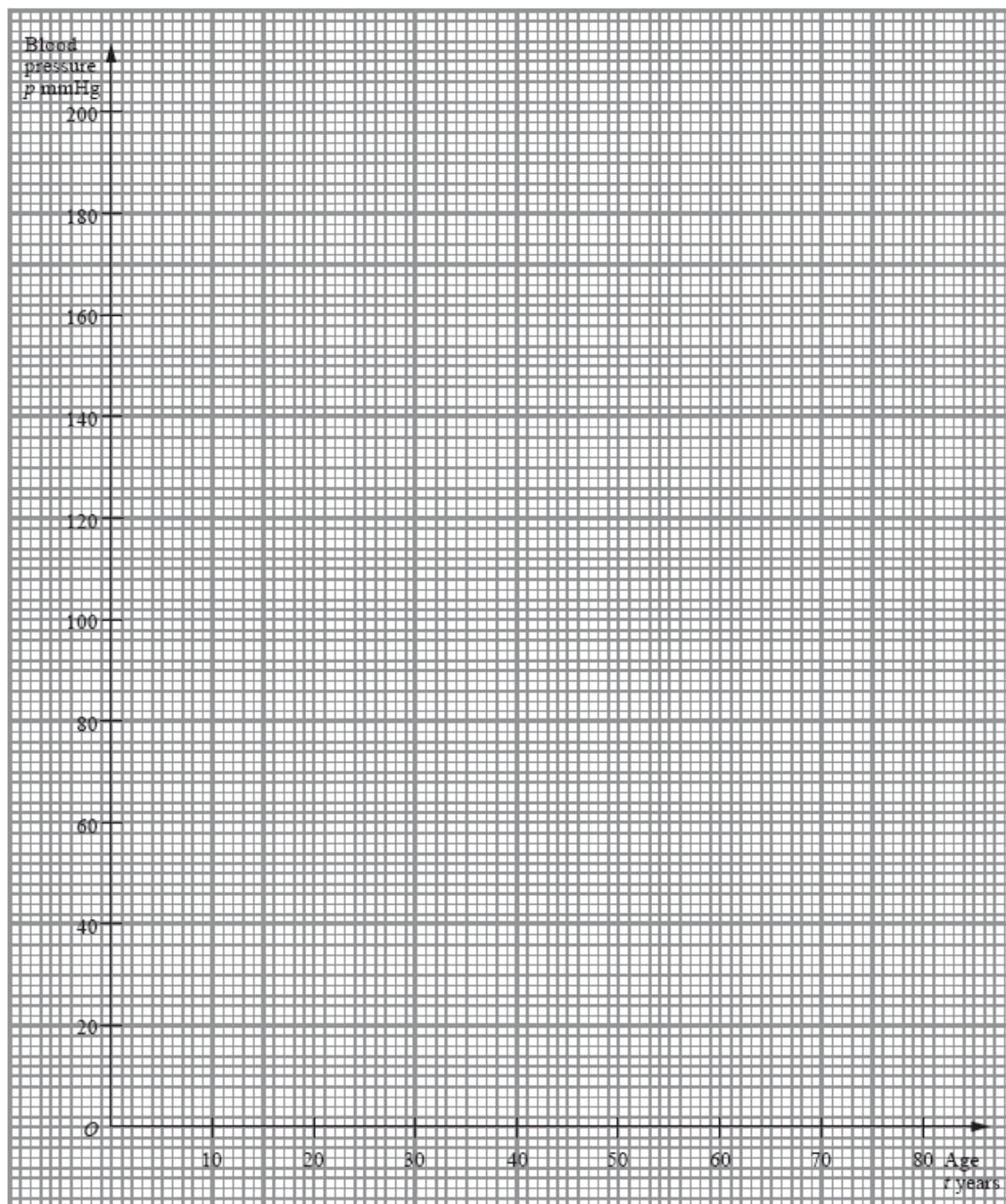
$$\left[\sum t = 341, \sum p = 833, \sum t^2 = 18181, \sum p^2 = 106397, \sum tp = 42948 \right]$$

- (a) Find S_{pp} , S_{tp} and S_{tt} for these data. (4)

- (b) Calculate the product moment correlation coefficient for these data. (3)

- (c) Interpret the correlation coefficient. (1)

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



(2)

(e) Find the equation of the regression line of p on t . (4)

(f) Plot your regression line on your scatter diagram. (2)

(g) Use your regression line to estimate the blood pressure of a 40 year old patient. (2)
(Total 18 marks)

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

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

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

(a) Find the equation of the regression line of w on l in the form $w = a + bl$. (5)

(b) Use your regression line to estimate the weight of a newborn turtle of length 60 mm. (2)

(c) Comment on the reliability of your estimate giving a reason for your answer. (2)
(Total 9 marks)

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

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

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

- (a) Calculate S_{xx} and S_{xy} . (3)
- (b) Find the equation of the least squares regression line of y on x in the form $y = a + bx$. (4)
- (c) Give an interpretation of the gradient of your regression line. (1)

Rosemary spends 3.3 hours using the revision course.

- (d) Predict her improvement in marks. (2)

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

- (e) Comment on Lee's claim. (1)
- (Total 11 marks)**

5. 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.

$$\sum t^2 = 10922.81, \sum v^2 = 42.3356, \sum tv = 677.971, \sum t = 401.3, \sum v = 25.08$$

- (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)

(Total 15 marks)

6. A metallurgist measured the length, l mm, of a copper rod at various temperatures, $t^\circ\text{C}$, and recorded the following results.

t	l
20.4	2461.12
27.3	2461.41
32.1	2461.73
39.0	2461.88
42.9	2462.03
49.7	2462.37
58.3	2462.69
67.4	2463.05

The results were then coded such that $x = t$ and $y = l - 2460.00$.

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

(You may use $\Sigma x^2 = 15965.01$ and $\Sigma xy = 757.467$)

(5)

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

(5)

- (c) Estimate the length of the rod at 40°C .

(3)

- (d) Find the equation of the regression line of l on t .

(2)

- (e) Estimate the length of the rod at 90°C .

(1)

- (f) Comment on the reliability of your estimate in part (e).

(2)

(Total 18 marks)

7. A manufacturer stores drums of chemicals. During storage, evaporation takes place. A random sample of 10 drums was taken and the time in storage, x weeks, and the evaporation loss, y ml, are shown in the table below.

x	3	5	6	8	10	12	13	15	16	18
y	36	50	53	61	69	79	82	90	88	96

- (a) On the grid below, draw a scatter diagram to represent these data.

(3)

- (b) Give a reason to support fitting a regression model of the form $y = a + bx$ to these data.

(1)

- (c) Find, to 2 decimal places, the value of a and the value of b .

(You may use $\Sigma x^2 = 1352$, $\Sigma y^2 = 53\,112$ and $\Sigma xy = 8354$.)

(7)

- (d) Give an interpretation of the value of b .

(1)

- (e) Using your model, predict the amount of evaporation that would take place after

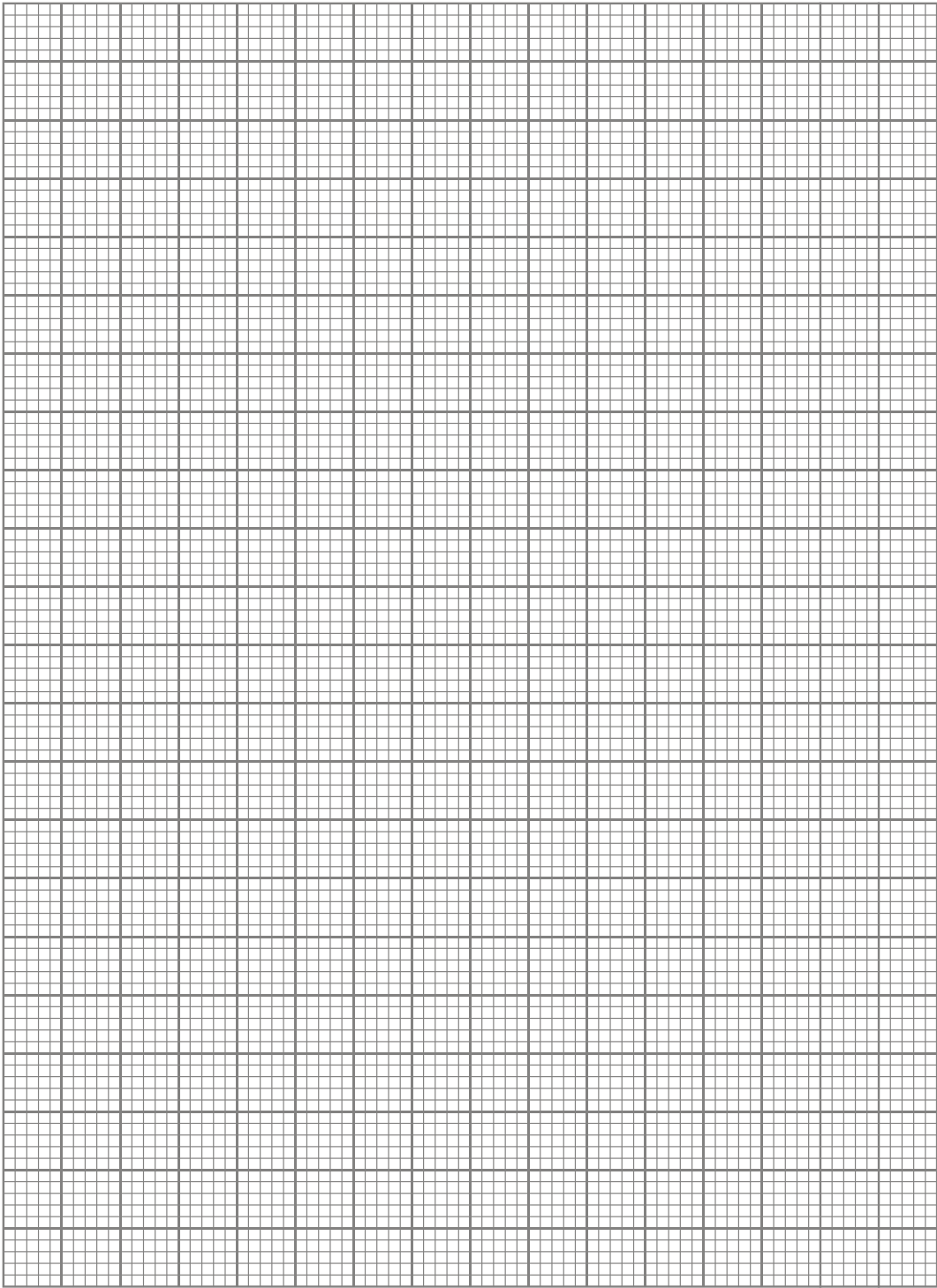
(i) 19 weeks,

(ii) 35 weeks.

(2)

(f) Comment, with a reason, on the reliability of each of your predictions.

(4)



(Total 18 marks)

8. A long distance lorry driver recorded the distance travelled, m miles, and the amount of fuel used, f litres, each day. Summarised below are data from the driver's records for a random sample of 8 days.

The data are coded such that $x = m - 250$ and $y = f - 100$.

$$\Sigma x = 130 \quad \Sigma y = 48 \quad \Sigma xy = 8880 \quad S_{xx} = 20\,487.5$$

- (a) Find the equation of the regression line of y on x in the form $y = a + bx$. (6)

- (b) Hence find the equation of the regression line of f on m . (3)

- (c) Predict the amount of fuel used on a journey of 235 miles. (1)
- (Total 10 marks)**

9. The following table shows the height x , to the nearest cm, and the weight y , to the nearest kg, of a random sample of 12 students.

x	148	164	156	172	147	184	162	155	182	165	175	152
y	39	59	56	77	44	77	65	49	80	72	70	52

- (a) On graph paper, draw a scatter diagram to represent these data.
(One sheet of graph paper to be provided) (3)

- (b) Write down, with a reason, whether the correlation coefficient between x and y is positive or negative. (2)

The data in the table can be summarised as follows.

$$\Sigma x = 1962, \quad \Sigma y = 740, \quad \Sigma y^2 = 47\,746, \quad \Sigma xy = 122\,783, \quad S_{xx} = 1745.$$

- (c) Find S_{xy} . (2)

The equation of the regression line of y on x is $y = -106.331 + bx$.

- (d) Find, to 3 decimal places, the value of b . (2)

- (e) Find, to 3 significant figures, the mean \bar{y} and the standard deviation s of the weights of this sample of students. (3)

- (f) Find the values of $\bar{y} \pm 1.96s$. (2)

- (g) Comment on whether or not you think that the weights of these students could be modelled by a normal distribution. (1)
(Total 15 marks)

10. An experiment carried out by a student yielded pairs of (x, y) observations such that

$$\bar{x} = 36, \quad \bar{y} = 28.6, \quad S_{xx} = 4402, \quad S_{xy} = 3477.6$$

- (a) Calculate the equation of the regression line of y on x in the form $y = a + bx$. Give your values of a and b to 2 decimal places. (3)

- (b) Find the value of y when $x = 45$. (1)
(Total 4 marks)

11. A researcher thinks there is a link between a person's height and level of confidence. She measured the height h , to the nearest cm, of a random sample of 9 people. She also devised a test to measure the level of confidence c of each person. The data are shown in the table below.

h	179	169	187	166	162	193	161	177	168
c	569	561	579	561	540	598	542	565	573

[You may use $\sum h^2 = 272\,094$, $\sum c^2 = 2\,878\,966$, $\sum hc = 884\,484$]

- (a) Draw a scatter diagram to illustrate these data. (4)
- (b) Find exact values of S_{hc} , S_{hh} and S_{cc} . (4)
- (c) Calculate the value of the product moment correlation coefficient for these data. (3)
- (d) Give an interpretation of your correlation coefficient. (1)
- (e) Calculate the equation of the regression line of c on h in the form $c = a + bh$. (3)
- (f) Estimate the level of confidence of a person of height 180 cm. (2)
- (g) State the range of values of h for which estimates of c are reliable. (1)

(Total 18 marks)

12. An office has the heating switched on at 7.00 a.m. each morning. On a particular day, the temperature of the office, t °C, was recorded m minutes after 7.00 a.m. The results are shown in the table below.

m	0	10	20	30	40	50
t	6.0	8.9	11.8	13.5	15.3	16.1

- (a) Calculate the exact values of S_{mt} and S_{mm} . (4)
- (b) Calculate the equation of the regression line of t on m in the form $t = a + bm$. (3)
- (c) Use your equation to estimate the value of t at 7.35 a.m. (2)
- (d) State, giving a reason, whether or not you would use the regression equation in (b) to estimate the temperature
- (i) at 9.00 a.m. that day,
- (ii) at 7.15 a.m. one month later.

(4)
(Total 13 marks)

13. A company wants to pay its employees according to their performance at work. The performance score x and the annual salary, y in £100s, for a random sample of 10 of its employees for last year were recorded. The results are shown in the table below.

x	15	40	27	39	27	15	20	30	19	24
y	216	384	234	399	226	132	175	316	187	196

[You may assume $\Sigma xy = 69\,798$, $\Sigma x^2 = 7\,266$]

- (a) Draw a scatter diagram to represent these data. (4)

- (b) Calculate exact values of S_{xy} and S_{xx} . (4)

- (c) (i) Calculate the equation of the regression line of y on x , in the form $y = a + bx$.
Give the values of a and b to 3 significant figures.

- (ii) Draw this line on your scatter diagram. (5)

- (d) Interpret the gradient of the regression line. (1)

The company decides to use this regression model to determine future salaries.

- (e) Find the proposed annual salary for an employee who has a performance score of 35. (2)
(Total 16 marks)

14. Eight students took tests in mathematics and physics. The marks for each student are given in the table below where m represents the mathematics mark and p the physics mark.

		Student							
		A	B	C	D	E	F	G	H
Mark	m	9	14	13	10	7	8	20	17
	p	11	23	21	15	19	10	31	26

A science teacher believes that students' marks in physics depend upon their mathematical ability. The teacher decides to investigate this relationship using the test marks.

- (a) Write down which is the explanatory variable in this investigation. (1)
- (b) Draw a scatter diagram to illustrate these data. (3)

(c) Showing your working, find the equation of the regression line of p on m . (8)

(d) Draw the regression line on your scatter diagram. (2)

A ninth student was absent for the physics test, but she sat the mathematics test and scored 15.

(e) Using this model, estimate the mark she would have scored in the physics test. (2)
(Total 16 marks)

15. The chief executive of Rex cars wants to investigate the relationship between the number of new car sales and the amount of money spent on advertising. She collects data from company records on the number of new car sales, c , and the cost of advertising each year, p (£000). The data are shown in the table below.

Year	Number of new car sales c	Cost of advertising (£000) p
1990	4240	120
1991	4380	126
1992	4420	132
1993	4440	134
1994	4430	137
1995	4520	144
1996	4590	148
1997	4660	150
1998	4700	153
1999	4790	158

- (a) Using the coding $x = (p - 100)$ and $y = \frac{1}{10}(c - 4000)$, draw a scatter diagram to represent these data. Explain why x is the explanatory variable. (5)
- (b) Find the equation of the least squares regression line of y on x .
[Use $\Sigma x = 402$, $\Sigma y = 517$, $\Sigma x^2 = 17\,538$ and $\Sigma xy = 22\,611$.] (7)
- (c) Deduce the equation of the least squares regression line of c on p in the form $c = a + bp$. (3)
- (d) Interpret the value of a . (2)
- (e) Predict the number of extra new cars sales for an increase of £2000 in advertising budget. Comment on the validity of your answer. (2)
- (Total 19 marks)**

16. To test the heating of tyre material, tyres are run on a test rig at chosen speeds under given conditions of load, pressure and surrounding temperature. The following table gives values of x , the test rig speed in miles per hour (mph), and the temperature, y °C, generated in the shoulder of the tyre for a particular tyre material.

x (mph)	15	20	25	30	35	40	45	50
y (°C)	53	55	63	65	78	83	91	101

- (a) Draw a scatter diagram to represent these data. (3)
- (b) Give a reason to support the fitting of a regression line of the form $y = a + bx$ through these points. (1)
- (c) Find the values of a and b . (4)
- (You may use $\Sigma x^2 = 9\,500$, $\Sigma y^2 = 45\,483$, $\Sigma xy = 20\,615$)
- (d) Give an interpretation for each of a and b . (2)

- (e) Use your line to estimate the temperature at 50 mph and explain why this estimate differs from the value given in the table.

(2)

A tyre specialist wants to estimate the temperature of this tyre material at 12 mph and 85 mph.

- (f) Explain briefly whether or not you would recommend the specialist to use this regression equation to obtain these estimates.

(4)

(Total 16 marks)