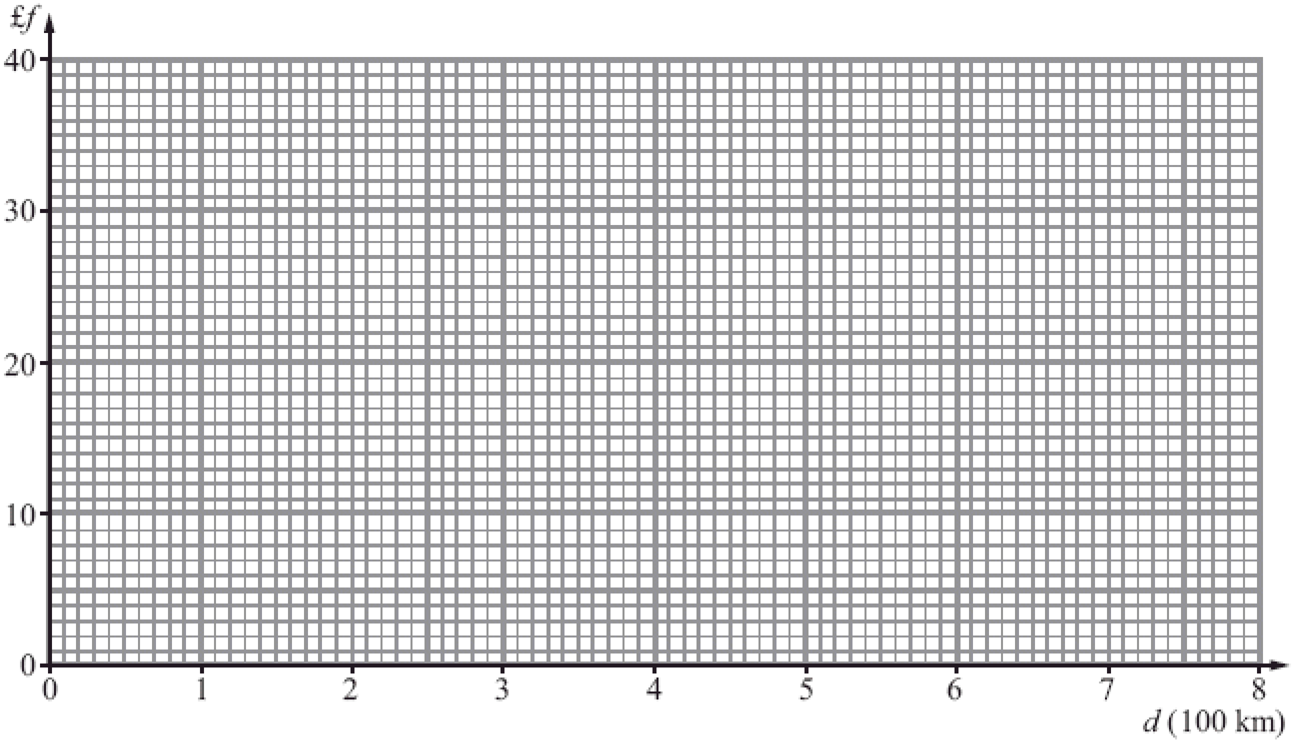
1. A travel agent sells flights to different destinations from *Beerow* airport. The distance d, measured in 100 km, of the destination from the airport and the fare £f are recorded for a random sample of 6 destinations.

Destination	*A*	*B*	*C*	*D*	*E*	*F*
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.



(2)

(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)

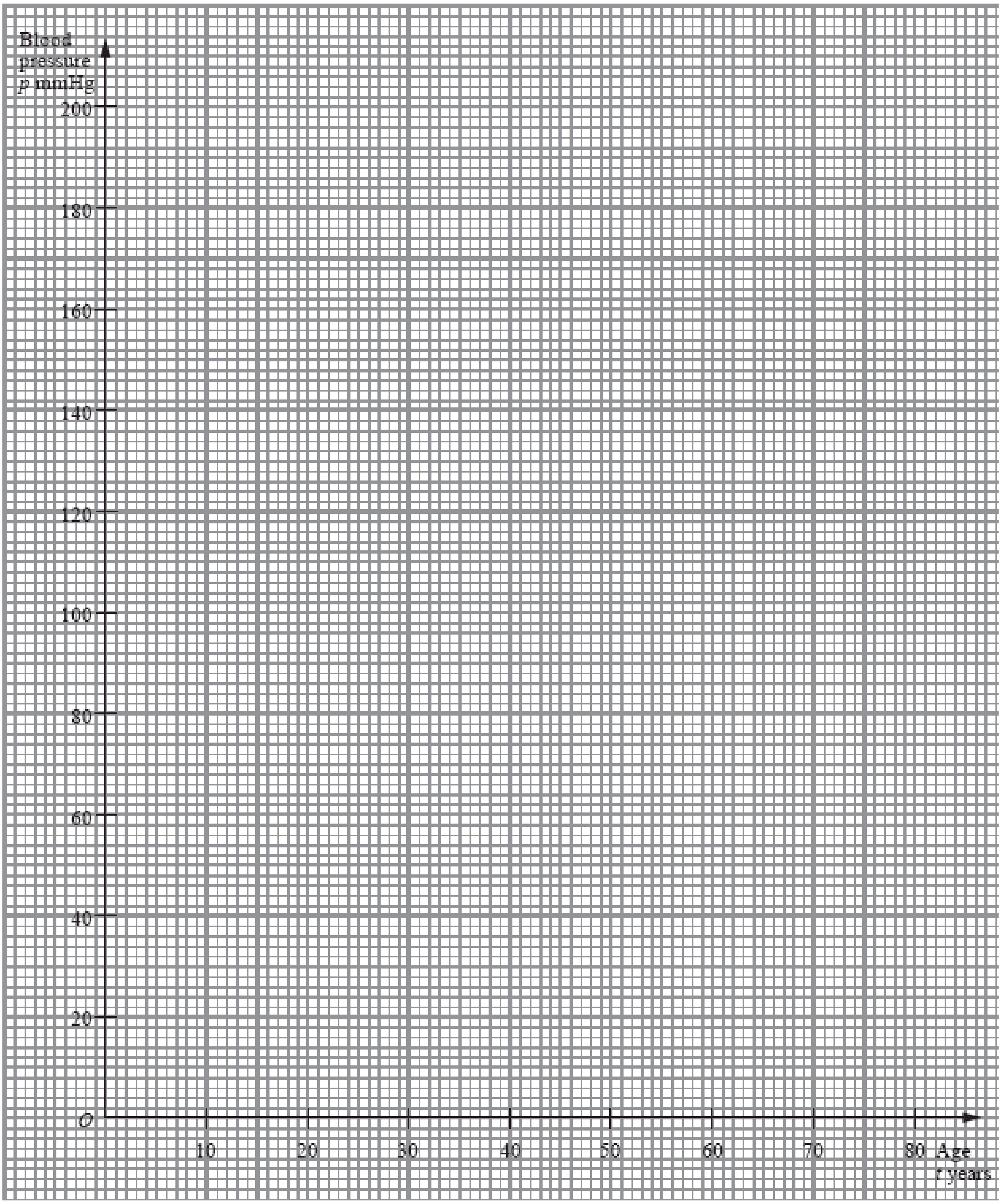
(b) Calculate the product moment correlation coefficient for these data.

(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°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 $y = f - 100$.

$$\sum x = 130 \qquad \sum y = 48 \qquad \sum xy = 8880 \qquad S_{xx} = 20\ 487.5$$

(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)

1. (a) B1 B1 2

Note

1st B1 for at least 4 points correct (allow ± one 2mm square)
2nd B1 for all points correct (allow ± one 2 mm square

(b) The **points** lie reasonably close to a straight **line** (o.e.) B1 1

Note

Ignore extra points and lines
Require reference to points and line for B1.

(c) $\sum d = 27.7, \quad \sum f = 146$ (both, may be implied) B1

$S_{dd} = 152.09 - \frac{(27.7)^2}{6}$ = 24.208….. **awrt 24.2** M1 A1

$S_{fd} = 723.1 - \frac{27.7 \times 146}{6}$ = 49.06…. **awrt 49.1** A1 4

Note

M1 for a correct method seen for either – a correct expression
1st A1 for S_{dd} awrt 24.2
2nd A1 for S_{fd} awrt 49.1

(d) $b = \dfrac{S_{fd}}{S_{dd}}$ = 2.026…. **awrt <u>2.03</u>** M1 A1

$a = \dfrac{146}{6} - b \times \dfrac{27.7}{6}$ = 14.97….. so ***<u>f = 15.0 + 2.03d</u>*** M1 A1 4

<u>Note</u>

1st M1 for a correct expression for *b* – can follow through their answers from (c)
2nd M1 for a correct method to find *a* – follow through their *b* and their means
2nd A1 for *f* = ... in terms of *d* and all values awrt given expressions. Accept 15 as rounding from correct answer only.

(e) A flight costs **£2.03 (or about £2)** for every extra **100km** or about **2p** per **km**. B1ft 1

<u>Note</u>

Context of cost and distance required. Follow through their value of *b*

(f) 15.0 + 2.03*d* < 5*d* so $d > \dfrac{15.0}{(5 - 2.03)}$ = 5.00 ~ 5.05 M1

So ***t*** > 500~505 A1 2

<u>Note</u>

M1 for an attempt to find the intersection of the 2 lines. Value of *t* in range 500 to 505 seen award M1.
Value of *d* in range 5 to 5.05 award M1.
Accept *t* greater than *500* to 505 inclusive to include graphical solution for M 1A1

[14]

2. (a) $S_{pp} = 106397 - \dfrac{833^2}{7} = 7270$ M1 A1

$S_{pp} = 42948 - \dfrac{341 \times 833}{7} = 2369,$

$S_{tt} = 18181 - \dfrac{341^2}{7} = 1569.42857....$ or $\dfrac{10986}{7}$ A1 A1 4

<u>Note</u>

M1 for at least one correct expression

1st A1 for S_{pp} = 7270, 2nd A1 for S_{tp} = 2369 or 2370,
3rd A1 for S_{tt} = awrt 1570

(b) $r = \frac{2369}{\sqrt{7270 \times 1569.42857...}}$ M1 A1ft

$= 0.7013375$ **awrt (0.701)** A1 3

Note

M1 for attempt at correct formula and at least one correct value (or correct ft) M0 for $\frac{42948}{\sqrt{106397 \times 18181}}$

A1ft All values correct or correct ft. Allow for an answer of 0.7 or 0.70 Answer only: awrt 0.701 is 3/3, answer of 0.7 or 0.70 is 2/3

(c) (Pmcc shows positive correlation.)

Older patients have higher blood pressure B1 1

Note

B1 for comment in context that interprets the fact that correlation is positive, as in scheme.

Must mention age and blood pressure in words, not just "*t*" and "*p*".

(d) Points plotted correctly on graph: –1 each error or omission

(within one square of correct position) B2 2

Note

Record 1 point incorrect as B1B0 on epen. [NB overlay for (60, 135) is slightly wrong]

(e) $b = \frac{2369}{1569.42857...} = 1.509466...$ M1 A1

$a = \frac{833}{7} - b \times \frac{341}{7} = 45.467413...$ M1

$P = 45.5 + 1.51t$ A1 4

Note

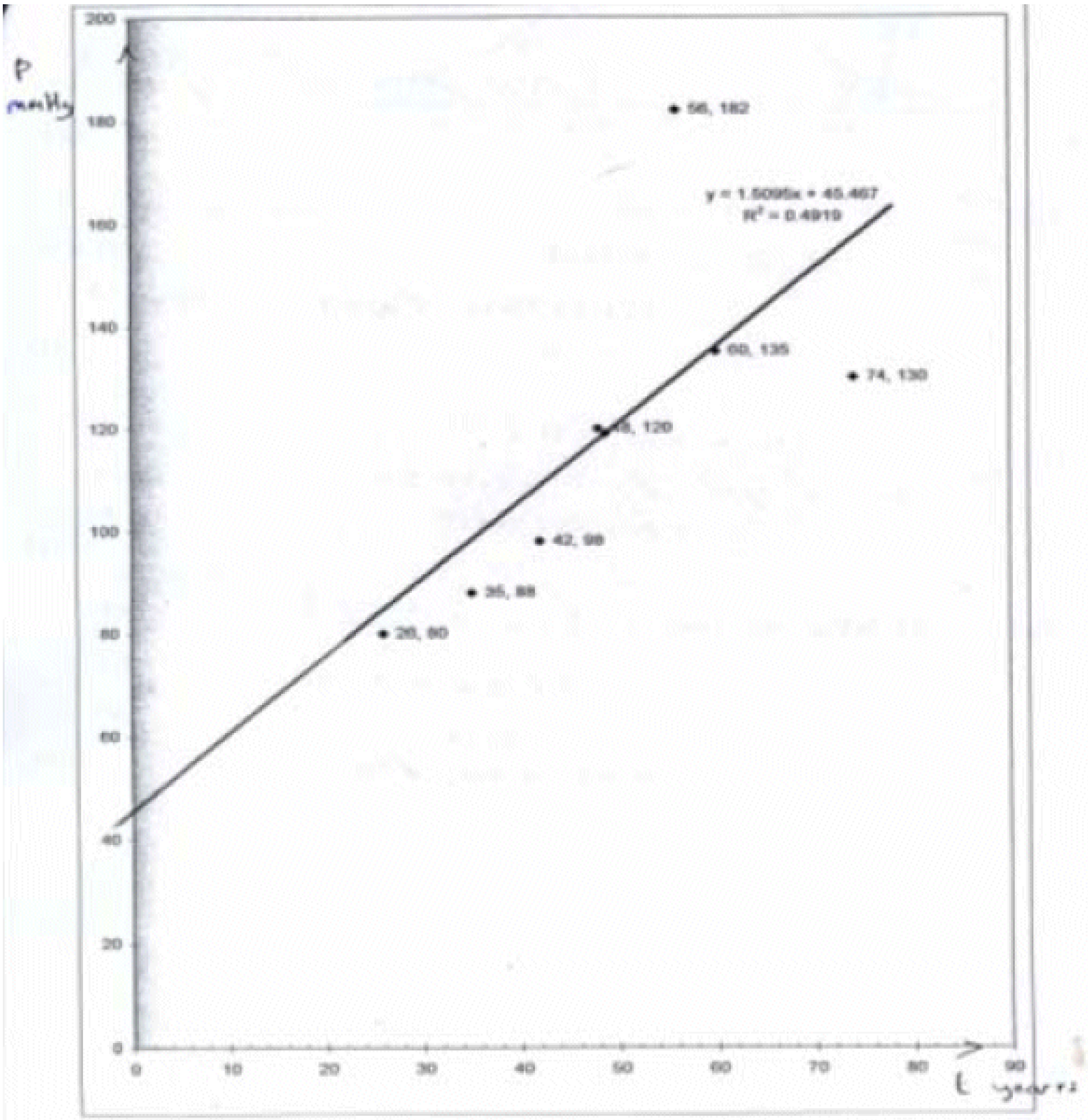
1st M1 for use of the correct formula for b, ft their values from (a)

1st A1 allow 1.5 or better

2nd M1 for use of $\bar{y} - b\bar{x}$ with their values

2nd A1 for full equation with a = awrt 45.5 and b = awrt 1.51. Must be p in terms of t, not x and y.

(f)	Line drawn with correct intercept, and gradient	B1ft B1	2
	Diagram for (d) + (f)		



Note

1st B1ft	ft their intercept (within one square). You may have to extend their line.
2nd B1	for correct gradient i.e. parallel to given line (Allow 1 square out when t = 80)

	(g)	t = 40, p = 105.84... from equation or graph.	**awrt 106**	M1 A1	2

Note

M1 for clear use of their equation with t = 40 or correct value from their graph.

A1 for awrt 106. Correct answer only (2/2) otherwise look for evidence on graph to award M1

[18]

3.	(a)	$b = \dfrac{59.99}{33.381}$		M1	
		= 1.79713…..	1.8 or awrt 1.80	A1	
		a = 32.7 – 1.79713…× 51.83		M1	
		= – 60.44525…	awrt –60	A1	
		w = – 60.445251…+ 1.79713…l	l and w required and awrt 2sf	A1ft	5

Note

Special case

$b = \dfrac{59.99}{120.1}$ = 0.4995 M0A0

a = 32.7 – 0.4995 × 51.83 M1A1

w = 6.8 + 0.50l at least 2 sf required for A1

	(b)	w = – 60.445251…+ 1.79713…× 60		M1	
		= 47.3825…	In range 47.3 – 47.6 inclusive	A1	2

Note

Substitute into their answer for (a) for M1

	(c)	It is extrapolating so (may be) unreliable.		B1 B1dep	2

Note

‘Outside the range on the table’ or equivalent award first B1

[9]

4. (a) $S_{xx} = 57.22 - \frac{(21.4)^2}{10} = 11.424$ M1

A1

$S_{xy} = 313.7 - \frac{21.4 \times 96}{10} = 108.26$ A1 3

<u>Note</u>

M1 for a correct expression
1st A1 for AWRT 11.4 for S_{xx}
2nd A1 for AWRT 108 for S_{xy}
Correct answers only: One value correct scores M1 and appropriate A1, both correct M1A1A1

(b) $b = \frac{S_{xy}}{S_{xx}} = 9.4765...$ M1 A1

M1

$a = \bar{y} - b\bar{x} = 9.6 - 2.14b = (-10.679...)$ M1 4

$y = -10.7 + 9.48x$

<u>Note</u>

1st M1 for using their values in correct formula

1st A1 for AWRT 9.5

2nd M1 for correct method for a (minus sign required)

2nd A1 for equation with a and b AWRT 3 sf (e.g. $y = -10.68 + 9.48x$ is fine)
Must have a full equation with a and b correct to awrt 3 sf

(c) Every (extra) <u>hour</u> spent using the programme produces about <u>9.5 marks improvement</u> B1ft 1

<u>Note</u>

B1ft for comment conveying the idea of <u>b marks per hour</u>. Must mention value of b but can ft their value of b. No need to mention "extra" but must mention "marks" and "hour(s)" e.g. "…9.5 times per hour …" scores B0

(d) $y = -10.7 + 9.48 \times 3.3$, $= 20.6$ awrt 21 M1, A1 2

<u>Note</u>

M1 for sub $x = 3.3$ into their regression equation from the end of part (b)
A1 for awrt 21

(e)	Model may not be valid since [8h is] outside the range [0.5 – 4].		B1	1

Note

B1 for a statement that says or implies that it may not be valid because outside the range.
They do not have to mention the values concerned here namely 8 h or 0.5 – 4

[11]

5. (a) $S_{tt} = 10922.81 - \frac{401.3^2}{15} = 186.6973$ awrt 187 M1A1

$S_{vv} = 42.3356 - \frac{25.08^2}{15} = 0.40184$ awrt 0.402 A1

$S_{tv} = 677.971 - \frac{401.3 \times 25.08}{15} = 6.9974$ awrt 7.00 A1 4

M1 any one attempt at a correct use of a formula.
Award full marks for correct answers with no working.
Epen order of awarding marks as above.

(b) $r = \frac{6.9974}{\sqrt{186.6973 \times 0.40184}}$ M1A1ft

$= 0.807869$ awrt 0.808 A1 3

M1 for correct formula and attempt to use
A1ft for their values from part (a)

NB Special Case for $\frac{677.971}{\sqrt{10922.81 \times 42.3356}}$ M1A0

A1 awrt 0.808
Award 3 marks for awrt 0.808 with no working

(c) t is the explanatory variable as we B1
can control temperature but not frequency of noise or equivalent comment B1 2

Marks are independent.
Second mark requires some interpretation in context and can be statements such as ‘temperature effects / influences pitch or noise’
B1 ‘temperature is being changed’ BUT B0 for ‘temperature is changing’

(d) High value of r or r close to 1 or Strong correlation B1 1

	(e)	$b = \dfrac{6.9974}{186.6973} = 0.03748$	awrt 0.0375	M1A1	
		$a = \dfrac{25.08}{15} - b \times \dfrac{401.3}{15} = 0.6692874$	awrt 0.669	M1A1	4

M1 their values the right way up
A1 for awrt 0.0375
M1 attempt to use correct formula with their value of *b*
A1 awrt 0.669

	(f)	$t = 19$, $v = 0.6692874 + 0.03748 \times 19 = 1.381406$	awrt 1.4	B1	1

awrt 1.4

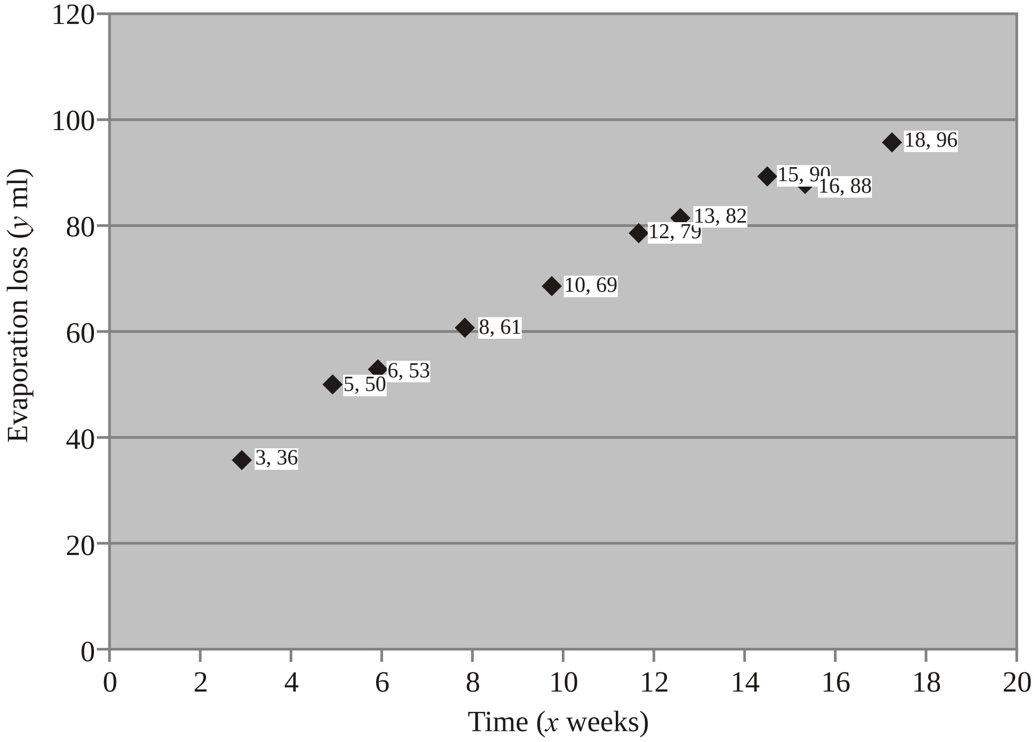
[15]

6.	(a)	$\sum x = \sum t = 337.1, \qquad \sum y = 16.28$ *Can be implied*	B1 B1	
		$S_{xy} = 757.467 - \dfrac{337.1 \times 16.28}{8} = 71.4685$ *either method, awrt 71.5*	M1 A1	
		$S_{xx} = 15965.01 - \dfrac{337.1^2}{8} = 1760.45875$ *awrt 1760*	A1	5
	(b)	$\text{b} = \dfrac{71.4685}{1760.45875} = 0.04059652$ *÷ correct way up, awrt 0.0406*	M1 A1	
		$\text{a} = \dfrac{16.28}{8} - \text{b} \times \dfrac{337.1}{8} = 0.324364$ *using correct formula, awrt 0.324*	M1 A1	
		$y = 0.324 + 0.0406x$ *3 sf or better but award for copying from above*	A1ft	5
	(c)	At $t = 40$, $x = 40$, $y = 1.948$, $l = 2461.948$ *sub x = 40, awrt 1.95, awrt 2461.95*	M1 A1 A1ft	3
	(d)	$l - 2460 = 0.324 + 0.0406t$ *LHS required*	M1	
		$l = 2460.324 + 0.0406t$ *awrt 2460.32 f.t. their 0.0406, l and t*	A1	2

	(e)	at t = 90, l = 2463.978 *awrt 2464*	B1	1
	(f)	90 °C outside range of data	B1	
		unlikely to be reliable	B1	2

[18]

7.	(a)	Sensible graph scales, labels, shape		B1, B1, B1	3



	(b)	Points lie close to a straight line		B1	1
	(c)	$S_{xy} = 8354 - \dfrac{106 \times 704}{10} = 891.6$		B1	
		$S_{xx} = 1352 - \dfrac{106^2}{10} = 228.4$		B1	
		$b = \dfrac{891.6}{228.4} = 3.903677 \ldots$	awrt 3.9	M1 A1	
		$a = \dfrac{704}{10} - b\,\dfrac{106}{10} = 29.021015 \ldots$	awrt 29	M1 A1	
			29.02, 3.90	A1ft	7

(d)	For every extra week in storage, another 3.90 ml of chemical evaporates		B1	1
(e)	(i)	103.12		
	(ii)	165.52	B1 B1	2
(f)	(i)	Close to range of x, so reasonably reliable	B1, B1	
	(ii)	Well outside range of x,	B1	
		could be unreliable since no evidence that model will continue to hold	B1	4

[18]

8. (a) $S_{xy} = 8880 - \frac{130 \times 48}{8} = (8100)$ — B1

may be implied

$S_{xx} = 20487.5$

$b = \frac{S_{xy}}{S_{xx}} = \frac{81000}{20487.5} = 0.395363\ldots$ — M1 A1

Allow use of their S_{xy} for M
awrt 0.395

$a = \frac{48}{8} - (0.395363...)\frac{130}{8} = -0.424649\ldots$ — M1 A1

allow use of their b for M
awrt –0.425

$y = -0.425 + 0.395x$ — B1ft — 6

3s.f.
Special case answer only B0 M0 B1 M0 B1 B1
(fully correct 3sf)
(≡ to B0 M0 A1 M0 A1 B1 on the epen)

(b) $f - 100 = -0.424649\ldots + 0.395\ldots(m - 250)$ — M1 a1ft

subst f – 100 & m – 250

$f = 0.735 + 0.395m$ — A1 — 3

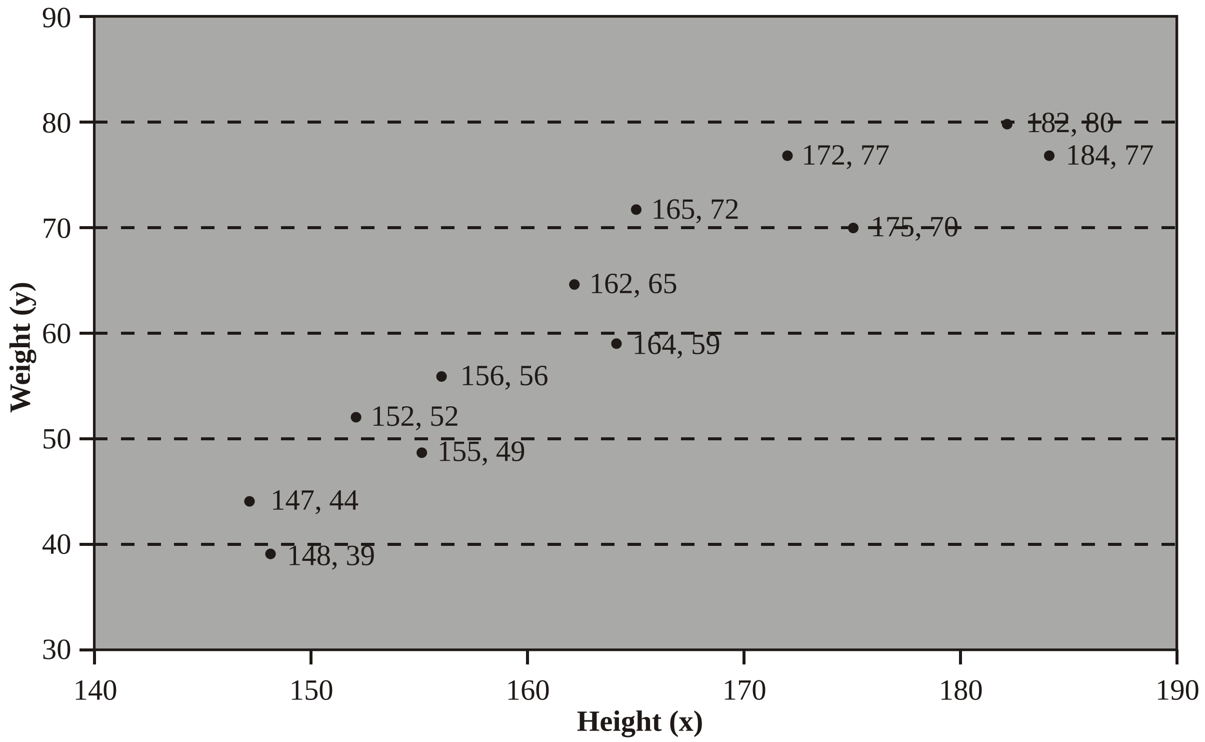
3 s.f.

(c) $m = 235 \Rightarrow f = 93.64489\ldots$ — B1 — 1

awrt 93.6/93.7

[10]

9. (a)



sensible scales	B1	
labels	B1	
shape	B1	3

(b) Positive; as x increases, y increases B1;B1g 2

context OK

(c) $S_{xy} = 122783 - \dfrac{1962 \times 740}{12} = 1793$ M1A1 2

use of formula, cao

(1793 only M1A1)

(d) $b = \dfrac{S_{xy}}{S_{xx}} = \dfrac{1793}{1745} = 1.027507\ldots$ M1A1 2

division, 1.028

(**SR** 1.028 B1 only)

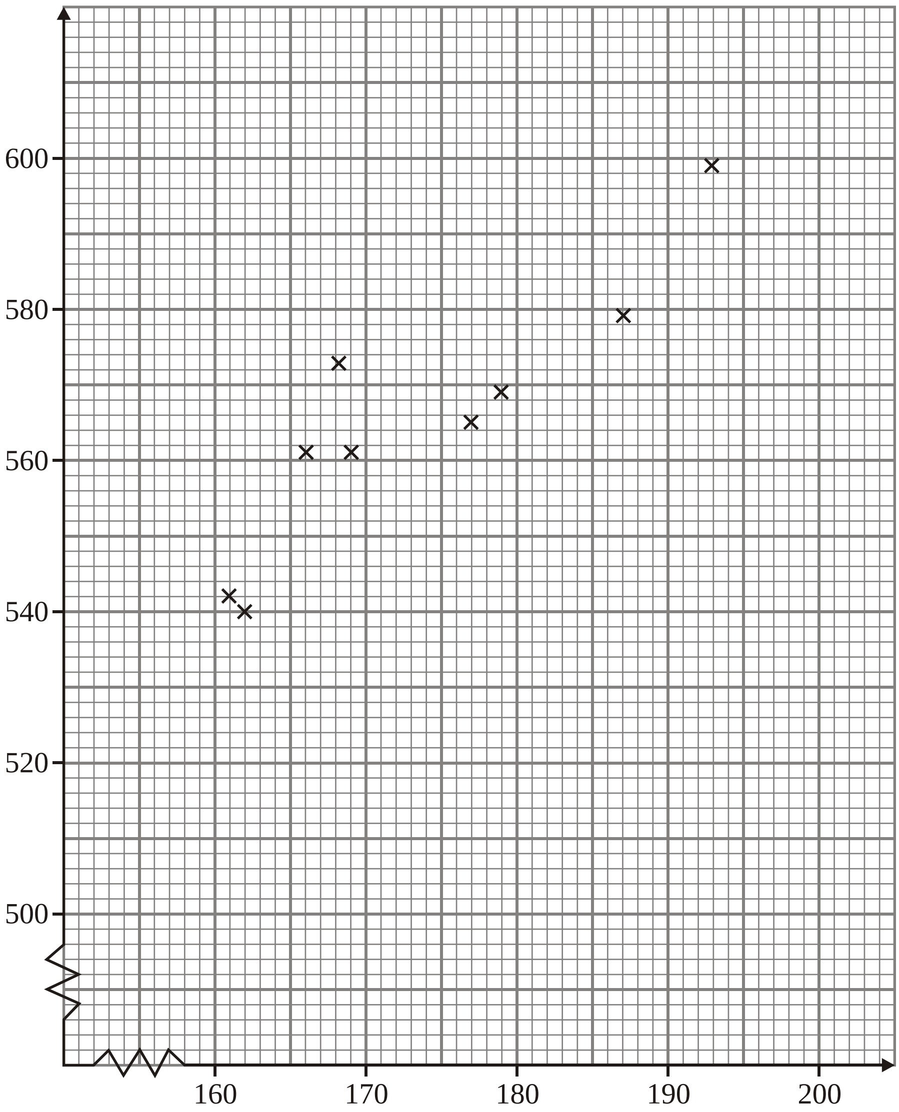
	(e)	$\bar{y} = \frac{740}{12} = 61\frac{2}{3}$	B1	
		$61\frac{2}{3}$ or $61.\dot{6}$ or 61.7		
		$s = \sqrt{\frac{47746}{12} - \left(\frac{740}{12}\right)^2} = 13.26859$	M1A1	3
		Use of formula including root, 13.3 or 13.9		
		(**SR** 13.3 or 13.9 B1 only)		
	(f)	34-36, 87-89	B1B1	2
		strict limits,3sf or better		
	(g)	All values between their 35.7 and their 87.7 so could be normal. Reason required	B1	1

[15]

10.	(a)	$b = \frac{S_{xy}}{S_{xx}} = \frac{3477.6}{4402} = 0.7900\ldots$	B1	
		awrt0.79		
		$a = \bar{y} - b\bar{x} = 28.6 - (0.7900\ldots) \times 36 = 0.159836\ldots$	B1	
		awrt 0.16		
		$y = 0.16 + 0.79x$	B1ft	3
		or equivalent		
	(b)	**OR** just answer B1 ONLY $y = 0.16 + 0.79 \times 45 = 35.71$ awrt 35.7	B1	1

[4]

11. (a)



Labels (not *x*, *y*)	B1	
Sensible scales allow axis interchange	B1	
Points	B2	4

(–1 ee)

(b) $S_{hc} = 884484 - \dfrac{1562 \times 5088}{9} = \underline{1433\tfrac{1}{3}}$ M1

correct use of S

$1433\tfrac{1}{3}$; $1433.\dot{3}$ A1

$S_{hh} = 1000\,{}^{2}\!/_{9}$; $S_{cc} = 2550$ A1; A1 4

$1000\,{}^{2}\!/_{9}$, $1000.\dot{2}$; 2550

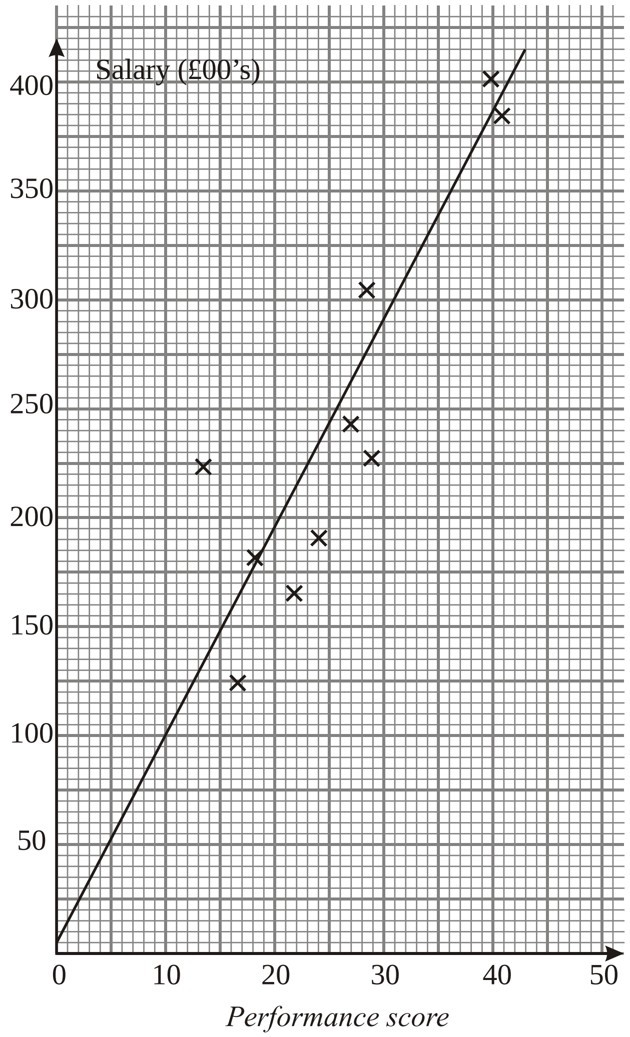
(NB: accept :- 9; i.e.:- $159\,{}^{7}\!/_{27}$; $111\,{}^{11}\!/_{81}$; $283\tfrac{1}{3}$)

	(c)	$r = \dfrac{1433\frac{1}{3}}{\sqrt{1000\frac{2}{9} \times 2550}}$	M1	
		substitution in correct formula		
		= $\underline{0.897488....}$	A1 ft A1	3
		AWRT 0.897(accept 0.8975)		
	(d)	Taller people tend to be more confident	B1	1
		context		
	(e)	$b = \dfrac{1433.\dot{3}}{1000.\dot{2}} = \underline{1.433014.....}$	M1	
		$a = \dfrac{5088}{9} - \dfrac{1433.\dot{3}}{1000.\dot{2}} \times \dfrac{1562}{9} = \underline{316.6256...}$	M1	
		allow use of their b		
		$\therefore \underline{c = 317 + 1.43h}$ (3sf)	A1	3
	(f)	$h = 180 \Rightarrow c = 574.4$ or 574.5683….	M1	
		subt. of 180		
		574 – 575	A1	2
	(g)	$161 \le h \le 193$	B1	1
				[18]
		NB (a) No graph paper ⇒ 0/4		
12.	(a)	$\Sigma m = 150$; $\Sigma m^2 = 5500$ $\Sigma t = 71.6$; $\Sigma t^2 = 930$; $\Sigma mt = 2147$	B1	
		5500 & 2147 seen		
		$S_{mt} = 2147 - \dfrac{150 \times 71.6}{6} = \underline{357}$	M1 A1	
		Accept $\dfrac{357}{60} = 59.5$		
		$S_{mm} = 5500 - \dfrac{150^2}{6} = \underline{1750}$	A1	4
		Accept $291.\dot{6}$		
		No working shown SR: B1 B1 only		

(b)	$b = \dfrac{357}{1750} = \underline{0.204}$		M1	
	$a = \dfrac{71.6}{6} - 0.204 \times \dfrac{150}{6} = \underline{6.8\dot{3}}$		M1	
	$\therefore \underline{t = 6.83 + 0.204\ m}$			
	No working seen SR: t = 6.83 + 0.204m B1 only		A1	3
	Accept $6.8\dot{3}$, *6.83,* $6\frac{5}{6}$ *%*			
(c)	7.35 ⇒ m = 35			
	$\therefore t = 6.8\dot{3} + 0.204 \times 35 = \underline{13.97\dot{3}}$		M1 A1	2
	14.0 AWRT			
(d)	(i)	9.00 ⇒ m = 120		
		No; outside range of data (after 7.50 am)	B1; B1	
	(ii)	No; No evidence model will apply one month later	B1; B1	4

[13]

13. (a)



Scales and labels	B1	
Accept x, y points	B3	4
(–1e.e.)		

(b) $S_{xy} = 69798 - \dfrac{256 \times 2465}{10} = \underline{6694}$

256, 2465 in (b) B1

S_{xy} *or* S_{xx} M1

$S_{xx} = 7266 - \dfrac{256^2}{10} = \underline{712.4}$

6694 B1

712.4 B1

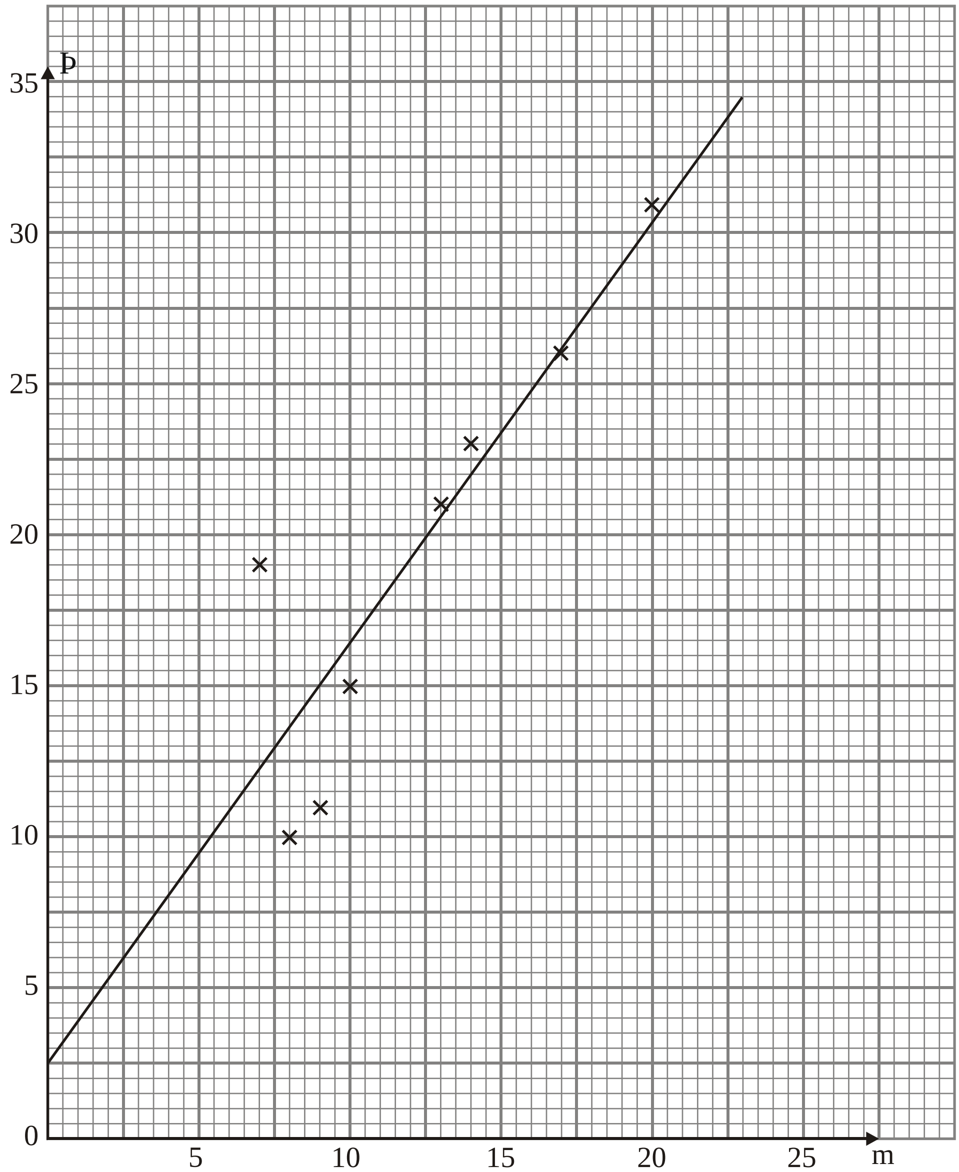
SR: No working ⇒ B0 M0 B1 B1 4

(c)	(i)	$b = \frac{6694}{712.4} = \underline{9.3964\ldots}$ *(their S_{xy} and S_{xx}) AWRT 9.40*	M1 A1	
		$a = \frac{2465}{10} - \frac{6694}{712.4} \times \frac{256}{10} = \underline{5.95199}\ldots$ *Using their values*	M1	
		$\therefore \underline{y = 5.95 + 9.40x}$ *3.s.f.*	A1 ft	
	(ii)	Line on graph By eye Not through origin. Accept broken scales	B1	5
(d)		Salary increases by <u>£940 for every 1 point</u> performance increase	B1 ft	1
(e)		$x = 35 \Rightarrow y = 334.95$ *Evidence – calculation or graph*	M1	
		Salary is £33,495 *33,000 – 34,000*	A1	2

[16]

14.	(a)	*m* is explanatory variable	B1	1

(b)



scales and labels	B1	
points	B2	
(6,7 points)	B1	3
Line	M1 A1	

(c) $\Sigma m = 98$; $\Sigma p = 156$; $\Sigma m^2 = 1348$; $\Sigma mp = 2119$

$S_{mp} = 2119 - \frac{98 \times 156}{8} = 208$ M1 A1

$S_{mm} = 1348 - \frac{98^2}{8} = 147.5$ A1

$\therefore\ b = \frac{S_{mp}}{S_{mm}} = \frac{208}{147.5} = 1.410169$ (awrt 1.41) M1 A1

$a = \frac{156}{8} - (1.410169\ldots) \times \frac{98}{8} = 2.225429$ (awrt 2.23) M1 A1

$\therefore$ p = 2.23 + 1.41m A1 ft 8

(d) Line on graph M1 A1 2

(e) $p = 2.23 + 1.41 \times 15 = 23.38$ M1 A1 2

[14]

15. (a)

x	20	26	32	34	37	44	48	50	53	58
y	24	38	42	44	43	52	59	66	70	79

B1

Change in cost of advertising influences number of new car sales B1
Graph: Scale and labels B1
Points all correct B2 5

(b) $S_{xy} = 22611 - \frac{402 \times 517}{10} = 1827.6$ M1 A1

$S_{xx} = 17538 - \frac{402^2}{10} = 1377.6$ A1

$b = \frac{S_{xy}}{S_{xx}} = \frac{1827.6}{1377.6} = 1.326655\ldots$ M1 A1

$a = \frac{517}{10} - (1.326655\ldots) \times \frac{402}{10} = -1.63153\ldots$ B1

$\therefore\ y = -1.63 + 1.33x$ B1 ft 7

(c) $\frac{c - 4000}{10} = -1.63 + 1.33(p - 100)$ M1 A1 ft

$c = 2653.7 + 13.3p$ A1 3

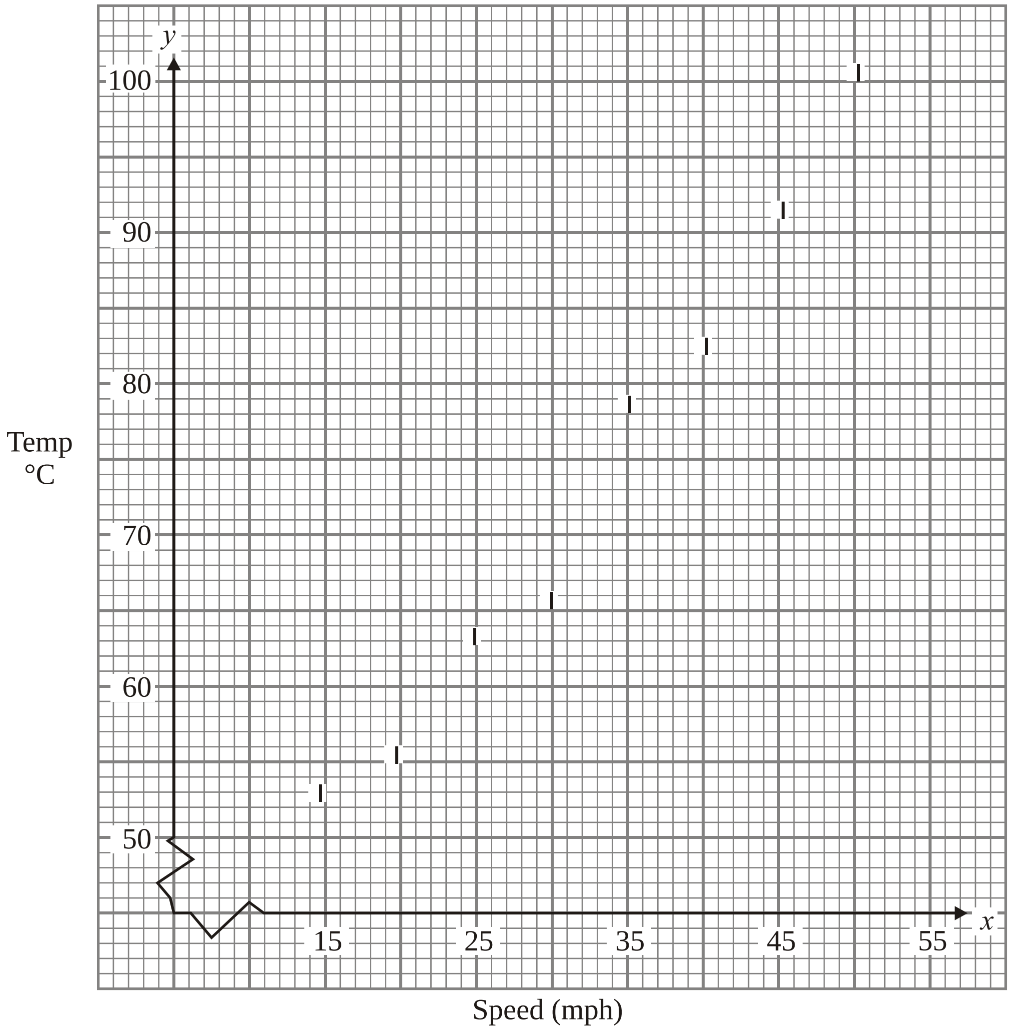
(d) No. sold if no money spent on advertising B1
$p = 0$ is well outside valid range – meaningless B1 2

(e) 2 × 13.3 = 27 extra cars sold B1

Only valid in range of data for 1990s B1 2

[19]

16. (a)



Scales & labels	B1	
Points	B2, 1, 0	3

(b) Points lie reasonably close to a straight line B1 1

(c)	$b = \dfrac{8\times 20615 - 260\times 589}{8\times 9500-(260)^2} = \dfrac{11780}{8400} = 1.40238\ldots$			
		(accept awrt 1.40)	M1 A1	
	$a = \dfrac{589}{8} - (1.40238\ldots)\left(\dfrac{260}{8}\right) = 28.0476175\ldots$			
		(accept awrt 28.0)	M1 A1	4
	$\therefore\ y = 28.0 + 1.40\,x$			
(d)	$a \Rightarrow$ surrounding air temperature when tyre is stationary		B1	
	$b \Rightarrow$ for every extra mph, temperature rises by 1.40 °C		B1	2
(e)	$y = 28.0 + 1.40 \times 50 = 98$		B1	
	Regression line is only a line of best fit and does not necessarily pass through all points		B1	2
	12 mph – reasonable to use line; 12 is just below lowest x-value		B1; B1	
	85 mph – not reasonable to use line; 85 is well outside range of values		B1; B1	4

[16]

1. The vast majority of candidates produced accurate scatter diagrams and on the rare occasion that there was a point missing it was predominantly point *D*. Explaining exactly why a linear regression model was appropriate proved to be difficult for candidates overall. Most candidates seemed to have the general idea but did not express this in the required terms and consequently very few earned this mark. Comments tended to be much more general about why linear regression is carried out and most talked about correlation being high without explaining that the points lie close to a line.

On the whole the correct formulae were used in calculations of S_{dd} and S_{fd}, with most candidates earning the method mark at the very least. The same was true in the calculations of *b* and *a* overall, although a common mistake was to calculate S_{ff} and go onto use that in the calculation of *b*. Premature approximation cost many candidates accuracy marks. Interpretations of the value of *b* were considerably varied, with relatively few candidates gaining this mark and some opted to omit this part altogether. Most candidates failed to relate their value to the context of the question and often tended to discuss *b* merely in terms of being the gradient. As a consequence, despite having the right kind of idea and correctly understanding the concept of the gradient, frequently candidates failed to gain this mark due to missing out the relevant units, mixing up the units or not quoting the actual value of *b*.

Very few candidates were able to formulate the correct equation with the correct units in part (f), and the majority found this particularly challenging, either omitting this part or resorting to evaluating the lines at the data points rather than equating and solving the equations. Often no clear strategy was apparent and a common mistake was to equate their equation to 5. There was clearly confusion over *t* and *d* and even out of those who were able to solve the required equation or inequality, not many found the value of *t* or range of *t* in km, as most tended to give their answer in terms of *d*. Occasionally the intersection point was evaluated using their graph after the lines had been plotted.

2. This was a high scoring question for most candidates. The calculations in parts (a) and (b) were answered very well with very few failing to use the formulae correctly. Part (c) received a good number of correct responses but many still failed to interpret their value and simply described the correlation as strongly positive. The scatter diagram was usually plotted correctly and most knew how to calculate the equation of the regression line although some used S_{pp} instead of S_{tt} and some gave their final equation in terms of y and *x* instead of *p* and *t* . Plotting the line in part (f) proved quite challenging for many candidates and a number with the correct equation did not have the gradient correct. Part (g) was usually well done but some chose to use their graph rather than their equation of the line and lost the final accuracy mark.

3. There were some good responses to this question, but some candidates calculated the slope as 59.99/120.1, although 3/5 marks were obtained if they went on to produce the equation as $w = 6.8 + 0.50l$ provided a minimum of 2 significant figures were used. Candidates should be able to identify the independent and dependent variables from a contextual question. The accuracy mark for the calculation of the intercept was lost if they used the rounded value of 1.8 for the slope in the calculation for the intercept. Many candidates did not believe 60mm to be quite far enough away from the data range to be called extrapolation showing that they did not go back and read the question carefully enough and consider the range of values given.

4. This proved to be a straightforward starter for most candidates who were able to tackle part (a) confidently, usually scoring full marks. Part (b) was answered well too; the correct formulae were selected and answers were usually given to 3 sf or better. Some candidates lost the final mark here for failing to give the full equation. Part (c) though was not answered well. There were plenty of comments about the gradient being positive or there being positive correlation or even skewness. Few realised that the instruction to "interpret" wanted an answer in context and comments conveying the idea that every extra hour spent on the programme yields an extra 9.5 marks were rare. Part (d) was straightforward again but some did not use their regression equation to find the estimate but rather tried to interpolate between the values of 3 and 3.5 given in the table. Part (e) had a mixed response. Many good candidates rejected Lee's comment on the basis that 8 hours was outside the range of the data and they secured the mark. Other, less successful, candidates simply calculated the value and then agreed with Lee or they rejected his claim on some other basis such as the difficulty of revising for 8 hours or 60 marks might take him above the total score on the paper.

5. This was done well by all but the weakest students with most using sufficient accuracy to score highly. Many candidates demonstrated an understanding of the use of the formulae to achieve full marks in part (a) and part (b). By far the main reason for loss of marks was premature approximation. Part (c) and part (d) were done well by good candidates. Only the more able candidates had a correct reason why t was the explanatory variable. Many called v the explanatory variable but gave a correct reason for t. The written parts were not universally done correctly, although the ability of students to deal with this topic has improved considerably in recent examinations. Rounding once again caused issues in part (e), but usually did not have an effect on part (f).

6. In part (a) calculating Σl instead of the required Σy was the most common reason for losing marks. In part (b) premature approximation was frequent and caused a loss of marks in other parts of the question. In part (c) substituting t=40 was usually attempted but some then neglected to add on the 2460. Candidates are now very well primed to say that a certain value is out of range and hence the result is not reliable.

7. Graphs were well done and candidates are finally labelling axes, but poor choice of scale for the x-axis meant some struggled to plot the graph accurately. For a standard question part (b) was disappointing with many answers referring to correlation but not to a straight line or line of best fit. Part (c) was generally well answered with the inevitable loss of the last mark through lack of accuracy by using 3.9 or not reading the question for the 2 decimal places required for the answers. A significant minority also thought that b represented the product moment correlation coefficient. Responses to part (d) usually missed the context of the question and in part (f) the proximity to the range of values of x was often omitted.

8. Candidates were well prepared for this question. The major problems arose as a result of rounding. The most surprising was rounding to 1 significant figure! This came up a great deal too frequently. It should be established now that there is a need to keep values for a and b un-rounded when 'decoding' the line but to express answers to 3 significant figures in the final stages.

9. Most candidates can plot and interpret scatter diagrams and use the formulae given in the formula book. A significant number of candidates still cannot correctly calculate the standard deviation to the required accuracy. A significant minority worked out the standard deviation of the x-values by mistake and of those who worked out the correct standard deviation, many used a premature approximation of the mean of 61.7 losing the accuracy mark

10. Most candidates were able to score well on this question. The values of both *b* and *a* were usually found accurately with most candidates giving the equation of the regression line of y on x to the required degree of accuracy. The value of y when x = 45 rarely caused any problems.

11. This question was familiar to most candidates and many of them answered it very well. This being said, too many used scales that were not sensible for the scatter diagram and far too many ignored the instruction to 'find the exact value'. The interpretation of the correlation coefficient was rarely given in terms of the context of the question and many candidates did not give the values of *a* and *b* to 3 significant figures in spite of previous advice.

12. Apart form arithmetic errors the first three parts of this question were well answered and many candidates gained most of the marks. It was good to see that many more of the regression equations were calculated with coefficients given to 3 significant figures. In the final part of the question, whilst there were many good solutions, some candidates did not state whether or not they would use the equation and others did not appreciate the context of the question.

13. Although the data in this question did not lend itself to easily chosen scales, most candidates did manage to produce a reasonable scatter diagram and eventually they were able to draw their regression line on it. Most candidates answered parts (b) and (c) well although some of them did not give their values of *a* and *b* to 3 significant figures as stated in the question. Whilst many candidates knew what was required in parts (d) and (e) they were unable to handle the units.

14. Many candidates appeared not to have sufficient time to complete this question. Not all candidates recognised the explanatory variable, leading some of them to find the wrong regression line. Apart from the use of silly scales the scatter diagram was often correctly drawn with many candidates going on to find correct values for the regression coefficients. Accuracy was much better handled in this question than in similar questions on previous papers. Too many candidates gave their final answer in terms of x and y rather than m and p.

15. Overall candidates responded well to this question. They knew how to work out the values of a and b in part (b) but their accuracy often let them down. They did not work to a sufficient degree of accuracy and a value of –1.77 was often seen instead of –1.63.Scatter diagrams were often correctly drawn but the scales used by many candidates were often not sensible. The back substitution in part (c) and the prediction in part (e) was beyond many of the candidates.

16. No Report available for this question.